

VIII. ULUSAL BİYOSTATİSTİK KONGRESİ

20-22 EYLÜL 2005

ULUDAĞ ÜNİVERSİTESİ
TIP FAKÜLTESİ
BİYOSTATİSTİK ANABİLİM DALI
VE
BİYOSTATİSTİK DERNEĞİ

<http://biyoistatistik.uludag.edu.tr>

T.C. Uludağ Üniversitesi Yayınıdır.

Ulusal Biyoistatistik Kongresi (8. : 2005: Bursa)

VIII. Ulusal Biyoistatistik Kongresi, 20-22 Eylül 2005,
[Uludağ Üniversitesi, Bursa] / [dznl.] Uludağ Üniversitesi Tıp
Fakültesi Biyoistatistik Anabilim Dalı, Biyoistatistik Derneği

1. Biyometri - - Kongreler I. E.a. II. Uludağ Üniversitesi
Tıp Fakültesi Biyoistatistik Anabilim Dalı III. Biyoistatistik
Derneği.

570.15195 Ul8s 2005
ISBN 975 - 6149 - 10 - 8

Uludağ Üniversitesi Basımevi Müdürlüğü
Tel-Faks: 442 82 50 Bursa – 2005

VIII. ULUSAL BİYOİSTATİSTİK KONGRESİ

Uludağ Üniversitesi Biyoistatistik Anabilim Dalı BURSA

DÜZENLEME KURULU

Prof. Dr. İsmet KAN (**Başkan**)
Yard. Doç. Dr. Bülent EDİZ
Dr. İlker ERCAN (**Sekreter**)
Dr. Semra AKGÖZ
Arş. Gör. Deniz SİĞİRLİ
Arş. Gör. Gökhan OCAKOĞLU
Arş. Gör. Güven ÖZKAYA
Arş. Gör. Şengül CANGÜR

BİLİMSEL DANIŞMA KURULU

Prof. Dr. Ahmet DİRİCAN
Prof. Dr. Ergun KARAAĞAOĞLU
Prof. Dr. Ersöz TÜCCAR
Prof. Dr. Fikret İKİZ
Prof. Dr. İsmet KAN
Prof. Dr. Kadir SÜMBÜLOĞLU
Prof. Dr. Kazım ÖZDAMAR
Prof. Dr. M. Yusuf ÇELİK
Prof. Dr. Mustafa ŞENOCAK
Prof. Dr. Nazan ALPARSLAN
Prof. Dr. Osman SAKA
Prof. Dr. Ömer SATICI
Prof. Dr. Refik BURGUT
Prof. Dr. Reha ALPAR
Prof. Dr. Rian DİŞÇİ
Prof. Dr. Vildan SÜMBÜLOĞLU

Prof. Dr. Yıldır ATAKURT
Prof. Dr. Yüksel BEK

ÖNSÖZ

Bilindiđi üzere, Biyoistatistik 1940'lı yıllar sonrasında sistematik medikal arařtırmalardaki artışla birlikte gelişme göstererek, ayrı bir disiplin haline gelmiştir. Öncelikle toplum büyüklüğü ve yaşam tabloları üzerinde yapılan çalışmalar sonrasında özellikle tıpta olan gelişmelere paralel olarak daha ileri boyutta analiz yöntemlerine gereksinim duyulması biyoloji, tıp ve sađlık bilimleri için istatistiđin uygulamalı bir bilim dalı olarak yer almasını sađlamıştır. Ülkemizde de bu gelişmelere paralel olarak Biyoistatistik biliminin yapılanması ve gelişme içinde olması sevindiricidir.

Uludađ Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı ve Biyoistatistik Derneđi'nin bu kongreyi düzenleyerek, Biyoistatistik bilim dalına ve dolayısıyla diđer tıp bilim dallarına yaptıkları katkıdan dolayı teşekkür ederim.

Rektör
Prof. Dr. Mustafa YURTKURAN

ÖNSÖZ

20-22- Eylül 2005 tarihleri arasında Bursa'da gerçekleştireceğimiz kongreye gönderilen bildirileri bir kitap halinde sunuyoruz. Kitabın basımı, kongre tarihinden önce olduğu için bazı bildiri ve posterlerin tam metinleri elimize ulaşamadı. Bu nedenle, yalnız özeti olan bu bildiri ve posterleri özet olarak basmak zorunda kaldık.

Kitap iki bölümden oluşmaktadır. Birinci bölümde, kongre programındaki sıralamaya göre sözlü bildiriler, ikinci bölümde ise geliş sırasına göre poster bildirileri yer almaktadır. Birinci bölümdeki 58 sözlü bildirinin 31 tanesi tam metin ve 27 tanesi özet, ikinci bölümdeki 22 posterin ise 4 tanesi tam metin ve 18 tanesi de özet olarak basılmıştır.

Kongre kitabının hazırlanması ve basılması aşamalarında; katkılarından dolayı başta sayın rektörümüz Prof.Dr. Mustafa Yurtkuran olmak üzere, bilimsel kurul üyelerine, anabilim dalımızdaki öğretim elemanlarına ve maddi destek sağlayan sponsor kuruluşlara teşekkür ederim.

Bursa 2005

Editör
Prof. Dr. İsmet KAN

İÇİNDEKİLER

Sayfa No:

SÖZLÜ BİLDİRİLER.....	1
Biyostatistik Alanında Makale, Sözlü Bildiri, Poster ve Rapor Hazırlama Yöntemlerinin Örneklerle Açıklanması (APA, ASA, Consort, ICMJE kurallarındaki son değişikliklerin incelenmesi) Kazım ÖZDAMAR.....	3
Tekrarlanan Ölçüm İçeren Denemelerde Örnek Genişliğinin Belirlenmesi: Faktörlerden Biri Tekrarlanan Ölçüm İçeren Araştırmalar Bahar TAŞDELEN, Arzu KANIK.....	5
İki Grup Karşılaştırmalarında Bağımlı Değişkenin Sürekli, İkili ve Sıralı Olduğu Durumda Örneklem Büyüklüğünün ve Gücün Hesaplanması Özgül VUPA, Gül ERGÖR.....	12
Wilcoxon İşaret Testinde Örneklem Büyüklüğü ve Güç Hesaplama Yaklaşımları S. Yavuz SANISOĞLU, Mesut AKYOL, Osman SARAÇBAŞI, İlker ETİKAN, Zeki AKKUŞ.....	28
Örneklem Genişliklerinin Küçük Olduğu Durumlarda Korelasyon Katsayısına Ait Güven Aralıkları Ersin ÖĞÜŞ, Ayşe Canan YAZICI, Fikret GÜRBÜZ.....	36
Varyans Analizi Uygulamalarında Çoklu Karşılaştırma Tekniklerinin 1. Tip Hata ve Güç Bakımından Karşılaştırılması İtler HELVACI, E. Arzu KANIK.....	37
Tanı Testlerini Değerlendirmede GEE ve Latent Class Modelleri R. BURGUT, Y. SERTDEMİR, R. TUNCER, M.U. SANDIKCI, İlker ÜNAL.....	39
Diagnostik Tanı Testlerinin Gold Standart Test Olmadan Gizli Sınıf (Latent Class) Analizi İle Değerlendirilmesi Meriç ÇOLAK, Vildan SÜMBÜLOĞLU.....	40
Altın Standart Yöntemin Bulunmadığı Çalışmalarda Tanı Testlerinin Etkinliğinin	

Saptanmasında Gizli Sınıf Analizinin Kullanılması <i>E. Arzu KANIK, Bahar TAŞDELEN, Murat ARSLAN, Mahmood MORSHEDI, Sergio OEHNINGER.....</i>	63
Tele-Radyoloji Uygulaması Etkililiğinin Gizli Sınıf (Latent Class) Analizi İle Değerlendirilmesi <i>Meriç ÇOLAK, Vildan SÜMBÜLOĞLU, Cenk SÖZEN, Korkut ERSOY.....</i>	65
İleriye Dönük Çalışmalarda Tanı Testlerinin Duyarlılık ve Seçicilikleri İçin Örneklem Büyüklüğü ve Güç Hesaplaması <i>Jale KARAKAYA, Erdem KARABULUT, Sevilay KARAHAN, Osman SARAÇBAŞI.....</i>	94
Kullback-Leibler Uzaklığı ve Lorenz Eğrisinden Elde Edilen Ölçüler Kullanılarak Tanı Testlerinin Değerlendirilmesi <i>Umut ARSLAN, Barış KAYA, Necla ÖZER, Ergun KARAĞAĞOĞLU.....</i>	97
ROC Eğrisi Altındaki Alanın Parametrik ve Parametrik Olmayan Yöntemlerle Hesaplanması <i>Sema ERDEN, Bahar TAŞDELEN</i>	103
Farklı Korelasyon Yapılarına Sahip Veri Setlerinde Bazı Regresyon Diagnostiklerinin Tartışılması <i>Ayşe Canan YAZICI, Arzu KANIK, Fikret GÜRBÜZ.....</i>	104
Regresyon Modellerine Alternatif Bir Yaklaşım: MARS <i>Gülhan OREKİCİ TEMEL, Handan ÇAMDEVİREN, A. Canan YAZICI.....</i>	105
Negatif Binom Regresyon Modeli <i>Anıl AKTAŞ, Osman SARAÇBAŞI</i>	124
Populasyon Büyüme Oranının Doğrusal Olmayan Bir Regresyon Modeli İle Tahmini <i>Gürol ZIRHLIOĞLU, Kazım KARA.....</i>	130
Mersin'de Yaşayan Çocuklarda İntihar Girişim Sıklığı İçin Poisson Regresyon Modelinin Kullanımı <i>Handan ANKARALI (ÇAMDEVİREN), Mehmet MENDEŞ, Zeki AKKUŞ, Sema ERDEN, Semra ERDOĞAN, Sıdık KESKİN, Fevziye TOROS.....</i>	138
Güç Doğum Nedeni İle Ölen Buzağılara Mevsim ve Yıl Faktörlerinin Etkisinin Poisson Regresyon Analizi Kullanarak İncelenmesi <i>I. Safa GÜRCAN, İlnur ÖZMEN, Aytaç AKÇAY, Mehmet N. ORMAN....</i>	149
İki Poisson Ortalamasının Karşılaştırılması <i>S. KARAHAN, J. KARAKAYA, O. SARAÇBAŞI.....</i>	151
Kontenjans Tablolarında Asimptotik P, Exact P ve Monte Carlo P Değerlerinin	

Karşılaştırılması	
<i>Ferhan ELMALI, Canan DEMİRÜSTÜ, Ertuğrul ÇOLAK, Kazım ÖZDAMAR</i>	154
Burun Yapısının İstatistiksel Şekil Analizi İle İncelenmesi	
<i>İlker ERCAN, İbrahim GÜNEY, Abdullah ETÖZ, Gökhan OCAKOĞLU, İsmet KAN, Mesut ÖZCAN</i>	155
Kategorik Veri Çözümlemesinde Logaritmik Doğrusal Modeller ve Bir Uygulama	
<i>Beyza DOĞANAY, Kenan KÖSE, Umud ÖZBEK, Yıldır ATAKURT</i>	156
Alt Grup Analizi	
<i>Beyza AKDAĞ, Vildan SÜMBÜLOĞLU, Kadir SÜMBÜLOĞLU</i>	158
Populasyon Genetiği Prensipieri ve Hardy-Weinberg Kuralı	
<i>Mesut AKYOL, S. Yavuz SANİSOĞLU, İlker ETİKAN, Zeki AKKUŞ, Reha ALPAR</i>	160
Vaka-Kontrol Çalışmalarında Hastalık-Gen İlişkinin İstatistiksel Analizi	
<i>Aysun ÇETİNYÜREK, Atilla Halil ELHAN, Osman Tolga KASKATI</i>	169
Genelleştirilmiş Sıralı Lojistik Regresyon Modeli	
<i>N. Anıl BARAK, Osman SARAÇBAŞI, Meltem HALİL, Servet ARIOĞUL</i>	171
Eş Grupların Analizinde Şartlı Lojistik Regresyon Modelinin Kullanımı	
<i>Sema ERDEN, Handan ÇAMDEVİREN, Semra ERDOĞAN</i>	181
Simmental X Güney Anadolu Kırmızısı Sığırlarına Ait Canlı Ağırlık Ölçümleri İçin Lojistik Büyüme Eğrisi	
<i>Cemil ÇOLAK, Mehmet N. ORMAN</i>	190
Gözleme Dayalı Çalışmalarda Propensity Skor: Bir Uygulama	
<i>Elif Çiğdem ALTUNOK, Nural BEKİROĞLU, Mehmet GENCELİ</i>	191
Koroner Arter Hastalığının Tahmininde Sınıflandırma Tekniklerinin Performanslarının Karşılaştırılması	
<i>İmran KURT, Mevlüt TÜRE, A. Turhan KÜRÜM</i>	203
Neural Networks İle Lojistik Regresyon Yöntemlerinin Eksik Ölçümlü veya Uç Değerler İçeren Veri Setleri Üzerinde Karşılaştırılması	
<i>İlker ÜNAL, Refik BURGUT, Yaşar SERTDEMİR</i>	215
Yapay Sinir Ağı İle Gruplanmış Zamanlı Verilerin Analizi	
<i>Bahar TAŞDELEN, Sema ERDEN, Hakan KALEAĞASI, Aynur ÖZGE</i>	217
Randomize Kontrollü Araştırmalarda Bilimsel Kalitenin Ölçümü İçin Yaklaşımlar: 4 Farklı Yöntemin İrdelenmesi	
<i>Mustafa ŞENOCAK, Necdet SÜT, Ömer UYSAL, Penbe ÇAĞATAY</i>	218

Hasta Memnuniyetine İlişkin Veri Setinde Boyut İndirgeme Yöntemlerinin Karşılaştırılması	
<i>Mevlüt TÜRE, İmran KURT, Zekeriya AKTÜRK</i>	231
Seçkin Türk Tıp Dergilerindeki Randomize Kontrollü Araştırmaların Yöntembilimsel Kalitesi	
<i>Mustafa ŞENOCAK, Necdet SÜT, Ömer UYSAL</i>	241
Seçkin Kanser Dergilerindeki Randomize Kontrollü Araştırmaların “Consort” Ölçütleri İle Değerlendirilmesi	
<i>Mustafa ŞENOCAK, Ömer UYSAL, Necdet SÜT, Hilal KÖKSALAN</i>	251
Klinik Denemelerde Hata Kaynakları ve Yanlılık Tipleri	
<i>Mehmet Ali SUNGUR, Emine Arzu KANIK</i>	262
Klinik Deneylede Grupların Belirlenmesinde “Minimizasyon” Metodu	
<i>Özden CALAY, Sevim PURİSA, Mustafa ŞENOCAK</i>	264
Randomizasyonda Katmanlamanın Yeri: Bir Benzetim Çalışması Dolayısı İle Değerlendirme	
<i>Mustafa ŞENOCAK, Ömer UYSAL, Necdet SÜT</i>	266
Randomize Kontrollü Klinik Denemelerde Tedavi Etkinliğinin Saptanmasında NNT'nin (Number Needed to Treat) Kullanımı: Güven Aralığı Sorunu	
<i>E. Arzu KANIK, Bahar TAŞDELEN</i>	277
Klinik Araştırmalarda “Körlük”: Kavram, Sınırlar, Uygulamalar ve Değerlendirme	
<i>Mustafa ŞENOCAK, Engin BAYKAL, Necdet SÜT, Ömer UYSAL, Ahmet DİRİCAN</i>	279
İki Renkli Microarray Deneme Düzenlerinin Planlanması	
<i>Seval KUL, E. Arzu KANIK</i>	281
Tekrarlı Ölçümlerin Eğri Altında Kalan Alan Yardımıyla Analizi	
<i>Umut ÖZBEK, Derya ÖZTUNA, Beyza DOĞANAY, S. Kenan KÖSE</i>	283
Bayesian Olasılık Kuramının Tıpta Kullanılması	
<i>Nazan BİLGEL</i>	285
İki Metropol Şehirde Halka Verilen Su Kalitesi ve Belediyelerin Su Politikaları	
<i>Ömer SATICI</i>	293
Yarışan Riskler İçin Cox Regresyon Modeli	
<i>Durdu SERTKAYA, Birce ÜNAL</i>	310
Türkiye’de Beklenen Yaşam Süresi: 1980-2001 Yılları Arasında Gerçekleşen Değişim	
<i>Nilgün MORALI, Esra TURAN</i>	311

Yaşam Sürelerindeki Farklılıkların Gini Katsayılı İle İncelenmesi Pınar ÖZDEMİR GEYİK, A. Kerem ULUDAĞ, Erdem KARABULUT, Osman SARAÇBAŞI.....	324
Vekil Sonlanım Değişkenlerinin Geçerliliğinin Belirlenmesinde Hangi Klinik Denemeler Kullanılmalı Yaşar SERTDEMİR, Refik BURGUT, İlker ÜNAL.....	333
Rasch Analizi Hakkında Kısa Bir Özet Atilla Halil ELHAN, Yıldır ATA KURT	335
Menapoz Semptomlarını Değerlendirme Ölçeği (MRS), Kupperman Menapoz Ölçeği (KMÖ) ve Nothingam Sağlık Profili (NSP) Ölçeklerinin Güvenilirlik Ve Yanıtlama Etki Büyüklüklerinin Yeni Menapozlu Hastalarda Karşılaştırılması N. BEKİROĞLU, Ayşe GÜRBÜZ, R. KONYALIOĞLU	344
Rasgele ve Sistematik Hataların Sınıf İçi ve Uyum Korelasyon Katsayıları ve Bland & Altman Yöntemi Üzerine Etkileri: Bir Simülasyon Çalışması Semra ERDOĞAN, E. Arzu KANIK.....	346
Hepatit A Virüs Enfeksiyonunun İleriye Yönelik Tahmin Edilmesinde Dört Farklı Zaman Serisi Yönteminin Karşılaştırılması Mevlüt TÜRE, İmran KURT.....	360
Tek Yönlü Varyans Çözümlemesinde Farklı Varyanslılık Problemi A. Fırat ÖZDEMİR, Serdar KURT.....	372
Dengelenmemiş Verilerde Varyans Analizi Tekniği ve Bir Uygulama Hayal BOYACIOĞLU, Saim KENDİR, Soner DUMAN.....	384
Varyans Heterojenliğinde Grup Ortalamalarının Karşılaştırılmasında Kullanılan Alternatif Yaklaşımlar Semra ERDOĞAN, Handan ANKARALI, Sema ERDEN	385
Normal Dağılım Gösteren Populasyonlardan Alınan Örneklerde Varyansların Homojenliğini Test Etmek İçin Kullanılan Bazı Yöntemlerin Karşılaştırılması Sıddık KESKİN, Ensar BAŞPINAR, Handan ANKARALI, Fikret GÜRBÜZ	401
POSTER BİLDİRİLER	413

Tıp Eğitiminde Biyoistatistik İlker ERCAN, Güven ÖZKAYA, Gökhan OCAKOĞLU, Bülent EDİZ,	
--	--

Berna YAZICI, İsmet KAN	415
Değişken Yapılarına Göre İstatistiksel Yaklaşım Zeki AKKUŞ, M. Yusuf ÇELİK	416
Yapay Sinir Ağları ve İstatistik Yöntemlerle İlişkileri Ayşe Canan YAZICI, Ersin ÖĞÜŞ, Handan ÇAMDEVİREN, Sinan CANAN	417
Türkiye'deki Özürlü Grupların Yapısının Uyum Analizi İle İncelenmesi Şengül CANGÜR, Deniz SİĞİRLİ, Bülent EDİZ, İlker ERCAN, İsmet KAN	418
Türkiye ve Avrupa Birliği'ne Üye Ülkelerin Sağlık Düzeyi Ölçütlerinin Çok Boyutlu Ölçekleme Analizi İle İncelenmesi Deniz SİĞİRLİ, Şengül CANGÜR, Bülent EDİZ, İlker ERCAN, İsmet KAN	420
Klinik Araştırmalarda, "Üstünlük", "Eşdüzeylelilik" ve "Daha Düşük Olmamak" Değerlendirmeleri İlkeleri F.N. TURAN, M. ŞENOCAK	421
Path Analizinin, Korelasyon Analizi ve Çok Değişkenli Regresyon Analizi İle Karşılaştırılması Rojan ARCAK, Hasan AKBAYIN, Tuba BİÇER	422
Döngüsel Süreçlerde Olay Kestirimi: Markov Model Yaklaşımı Necdet SÜT, Mevlüt TÜRE, Mustafa ŞENOCAK	424
18 Yaş Üzeri Sağlıklı Bireylerde Kantitatif Ultrason Ölçümlerinin İncelenmesinde Değişim Katsayısının Kullanılması Başak GÜRTEKİN, Sevda ÖZEL, Rian DiŞÇİ	434
Mevsimsel İnsidans Hızı Farklarının Anlamlılık Testi ve Bir Uygulama Ahmet DİRİCAN, Ömer UYSAL, Sevim PURİSA, Ayşegül BOSTAN, Mustafa ŞENOCAK	443
Sağlık Personelinin Empatik Eğilimlerinin ve Mesleki Doyumlarının Çok Değişkenli Kovaryans Analizi (Mancova) İle Değerlendirilmesi Semra AKGÖZ, İsmet KAN, Alis ÖZÇAKIR, Yusuf SİVRİOĞLU, Kamuran TOMBUL, Yıldız ALTINSOY	444
Dört Farklı Normallik Testinin Tip I Hata ve Güç Bakımından Değişik Dağılımlarda İncelenmesi Derya ÖZTUNA, Atilla Halil ELHAN, Ersöz TÜCCAR	446
Ülkemizde Ağrı ve Kapı İğne Zamanını, Uluslar arası Klavuzların Önerdiği Süreye Uygun mudur?	

Hatice BULUT, Gülşah SEYDAOĞLU, Fevziye TÜREMEZ.....	448
Genelleştirilmiş Tahmin Denklemleri: GTD	
Seval KUL, E. Arzu KANIK.....	450
Bir Kentte 1995-2005 Yılları Arasında Tutulan 63.831 Ölüm Sertifikası Bize Ne Anlatıyor?	
Gülşah SEYDAOĞLU, Fevziye TÜREMEZ, Nilgün ARPACI, Nazan ALPARSLAN.....	452
Faktör Analizi Modelinde Ortak Varyansları Belirlemede Genelleştirilmiş Terslerin Kullanımı	
Seval SÜZÜLMÜŞ, Sadullah SAKALLIOĞLU.....	454
Fuzel Yağından İzomal Asetatın Sürekli Üretiminde Optimum Koşulların Belirlenmesi İçin İki Yöntemin Bir Arada Kullanılması: Yanıt Yüzeyleri ve Yapay Sinir Ağları	
Ömer Utku ERZENGİN, Havva ELEMZADE, Gamze (Yalçın) ERZENGİN, Afife GÜVENÇ.....	455
Türkiye’de İl ve İlçe Merkezleri Ölüm İstatistikleriyle Mortalitenin Yaşa, Cinsiyete ve Zamana Göre Değişimi (1990-2003)	
R. Erol SEZER, Hafize SEZER.....	458
Diskriminant (Ayırma) Analizi İle El Kullanımı Tayini	
Bülent EDİZ, Şengül CANGÜR, Gökhan OCAKOĞLU, İlker ERCAN, N. Şimşek CANKUR, İhsaniye COŞKUN, İsmet KAN.....	465
İki Yanıt Değişkenli Değişik Veri Setlerinde Lojistik Regresyon ve Diskriminant Analizinin Doğru Sınıflandırma Oranları: Bir Simülasyon Çalışması	
Bülent EDİZ, Güven ÖZKAYA, Deniz SİĞİRLİ, Şengül CANGÜR, Gökhan OCAKOĞLU, İlker ERCAN, İsmet KAN.....	466
Araştırma Etik Kurullarında Biyoistatistik Uzmanının Önemi	
Elif ATICI, Sezer ERER.....	468
Over Kanserlerinde 5 Yıllık Sağ Kalım Analizi	
E. ASLANER, M.A. VARDAR, R. BURGUT, A. ALTINTAŞ.....	480

SÖZLÜ BİLDİRİLER

**BİYOİSTATİSTİK ALANINDA MAKALE, SÖZLÜ BİLDİRİ,
POSTER VE RAPOR HAZIRLAMA YÖNTEMLERİNİN
ÖRNEKLERLE AÇIKLANMASI
(APA, ASA, Consort, ICMJE kurallarındaki
son değişikliklerin incelenmesi)**

Kazım Özdamar

Eskişehir Osmangazi Üniversitesi Biyoistatistik Anabilim Dalı

Biyoistatistik bilimi bilgisayar teknolojisinin ve paket programların gelişmesine paralel olarak çok hızlı bir gelişim kaydetmektedir. Bu gelişim ile birlikte çok hızlı ve kaliteli bilimsel bilgi üretimi yapılabilmektedir. Ülkemizde de biyoistatistik konusunda yapılan yurt içi ve yurtdışı yayınlar artmış bulunmaktadır. Ülkemizde üretilen bilimsel bilgilerin gerek yurt içi ve gerekse yurtdışı platformlarda sunulmasında yazım, bildirim ve sunum tekniklerinde yapılan temel yaklaşım hatalarından dolayı gecikmeler, bloklamalar ve geri çevirmeler nedeniyle bilimsel bilgilerin bilim çevrelerine aktarılması, iletilmesi gecikmektedir, hatta mümkün olamamaktadır.

Ayrıca Tıp Dergileri Editörleri Uluslar arası Komitesi (International Committee of Medical Journals Editors, ICMJE) üye olan uluslar arası dergilerde yayınlanacak makalelerin belirli gereksinimleri karşılama ve Biyoistatistik yöntemlere uygun olarak düzenleme, veri toplama, özetleme, analiz ve yorumlama kurallarına uygun olmasını şart koşmaktadır. Bu gereksinimlerin tam olarak yerine getirilmesi için Biyoistatistiksel denetiminden geçmesini şart koşmaktadır. Bu açıdan da Biyoistatistikçinin tıbbi bilimsel yazıların hazırlanması (porejelendirme), veri toplama, özetleme, analiz, yorumlama ve yazım aşamalarında uyulması gereken koşulları çok iyi bilmesi ve bilim insanlarını yönlendirmesi gerekmektedir.

Özellikle ICMJE kurallarına göre 01.07.2005 tarihinden sonra yapılacak klinik denemelerin mutlaka kayıt altına alınması ve üye dergilerde yayınlanması için bazı gereksinimlere uyulması gereği biyoistatistikçilerin danışmanlık, denetim ve ortak çalışmalara katılma aşamalarında daha dikkatli olmaları gerekliliğini ortaya koymaktadır. Bu nedenle biyoistatistikçi, kendi özgün çalışmaları yanında diğer tıp dallarındaki çalışmalarda uyulacak araştırma projelendirme, araştırma yönetimi ve yayıma dönüştürme aşamalarında her türlü gelişmeyi izlemek, bilmek ve uygulamak zorundadır.

Bu sunumumuzda bilimsel araştırma ve yayınlarda APA, ASA, CONSORT, ICMJE kurallarında Biyoistatistik için büyük önem taşıyan değişiklik, eklemeler üzerinde örneklerle durulmuştur.

Bilim adamı, ürettiklerini bilim camiasına etkili olarak sunabilen kişidir. Bu açıdan Biyoistatistikçiler bilimsel yazı yazım ve yayım kurallarını çok iyi bilmeli, değişiklikleri izlemeli ve yeni gelişmelere kolayca adapte olabilmelidir.

Anahtar Kelimeler: APA Style, ASA yazım stili, ICMJE- gerekli koşulları, Yazım kuralları, CONSORT gereksinimleri, Araştırma Planlama ve Yayınlama kuralları.

Kaynaklar

1. APA (2005), APA Style Publication Manuel (5th Edition), www.APASStyle.org (12.07.2005).
2. ASA (2005), American Statistical Association Style Guide, www.amstat.org (12.7.2005)
3. CONSORT (2005), CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel-Group Randomized Trials, (www.consort-statement.org) (12.7.2005)
4. ICMJE (2004), Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication, (updated October 2004), www.icmje.org (12.7.2005)
5. ICMJE (2005), Clinical Trial Registration: A Statement from the International Committee Medical Journal Editors, www.icmje.org (12.7.2005)
6. Özdamar K. (2004), Modern Bilimsel Araştırma Yöntemleri, Kaan Kitabevi, Eskişehir.

TEKRARLANAN ÖLÇÜM İÇEREN DENEMELERDE ÖRNEK GENİŞLİĞİNİN BELİRLENMESİ: FAKTÖRLERDEN BİRİ TEKRARLANAN ÖLÇÜM İÇEREN ARAŞTIRMALAR

Bahar Taşdelen, Arzu Kanık

Mersin Üniversitesi Tıp Fakültesi Biyoistatistik AD, MERSİN

e-mail: bahartasdelen@mersin.edu.tr

Özet

Bu çalışmada, iki faktörlü, faktörlerden birinin tekrarlanan ölçüm içerdiği bir denemede örnek genişliğini belirlemek için kullanılan iki yaklaşım karşılaştırılmıştır. İlk yaklaşım, tekrarlanan ölçümler arasındaki korelasyon ve tekrar sayısı hakkında ön bilgi yokken kullanılmaktadır. İkinci yaklaşım ise bu ön bilgiler mevcut iken kullanılmaktadır. Farklı korelasyonlar ve tekrar sayıları için örnek genişlikleri hesaplanarak iki yaklaşım arasındaki farklılık incelenmiştir. Tekrar sayısı $m=2$, korelasyon $\rho=0.20$ ve standardize edilmiş etki büyüklüğü $\Delta=0.2$ için örnek genişliği 390 iken, $\rho=0.80$ için örnek genişliği 585'dir. Ancak m ve ρ hakkında herhangi bir bilgi yokken örnek genişliği, 650'dir. Yani, II. Yaklaşım, iki tekrarlı ve ölçümler arasında çok yüksek korelasyon bulunan denemeler için örnek genişliği I. yaklaşım ile bulunan örnek genişliğine yakın bulunmaktadır. Ayrıca $m=3$ ve $\rho=0.20$ için örnek genişliği 304 olarak bulunmuştur. Tekrarlanan ölçüm sayısı arttıkça ve korelasyon azaldıkça örnek genişliği önemli ölçüde azalmaktadır. Aynı zamanda standardize edilmiş etki büyüklüğü arttıkça örnek genişliği azalmaktadır. Sonuç olarak, tekrarlanan ölçümler arasındaki korelasyon hakkında hiçbir ön bilgi mevcut değilse, bu durumda I. yaklaşımın kullanılmasında bir sakınca yokken, tekrar sayısı ve korelasyon biliniyorsa II. yaklaşımı kullanmak daha doğru olacaktır.

Giriş

Ölçümlerin birbirini takip eden zamanlarda aynı denekler üzerinde tekrarlandığı veya araştırmada aynı deneğin birden çok ölçümünün kullanıldığı kısaca; ölçümlerin birbirine bağımlı olduğu denemelere tekrarlanan ölçümlü (repeated-measurement) denemeler adı verilir. Bu tip ölçümleri içeren ve içermeyen iki faktörün birlikte etkilerinin incelendiği klinik araştırmalar oldukça yaygın olarak kullanılmaktadır. Örneğin farklı anestezi yöntemlerin hastanın operasyon sırasındaki ve sonrasındaki klinik bulguları

üzerine etkisinin operasyon sonrasında geçen zamana göre incelendiği araştırmalar.

Bu tip araştırmalarda tekrarlanan ölçüm, genellikle bir zaman faktörü olarak çıkar karşımıza. Klasik istatistik metotlarından paired-t test, tekrarlanan ölçümün iki seviyesi (önce-sonra) bulunması durumunda tekrarlanan ölçümlü varyans analizinin (repeated-measurement ANOVA) özel bir halidir.

Tüm deneme düzenlerinde olduğu gibi tekrarlanan ölçüm içeren bu tip denemeler için gerekli örneklem büyüklüğü öncelikle testin gücü ve etki büyüklüğü ile ilgilidir. Ancak bu tip araştırmalarda tekrarlanan ölçüm sayısı ve ölçümler arası korelasyon da dikkate alındığında tekrarlanan ölçüm içeren tüm denemelerde sabit bir güce karşılık gerekli örneklem genişliği, aynı etki büyüklüğü için, bağımsız ölçüm içeren denemeler için gerekli örneklem genişliğinden her zaman daha düşük olacaktır. Bu durum tekrarlanan ölçüm içeren deneme düzenlerinin en önemli avantajlarından birisidir. Araştırma için gerekli örneklem genişliğinin saptanmasında bir önemli konu da maliyettir. Eğer hastalara uygulanan tedavi şeklinin ve izleme süresinin maliyeti yüksek ise, bu tedavinin etkisini araştırmak için aynı hasta üzerinde çok uzun süre denemesi veya çok fazla hastaya uygulanması gereksiz yere maliyeti artıracaktır. Bununla beraber tedavinin çok uzun sürmesi sonucunda, hastada yeni hastalıkların gözlenmesi ve bu hastalıkların tedavi sonuçlarını etkilemesi de mümkündür. Ancak, tekrarlanan ölçüm sayısı arttıkça gerekli örneklem genişliğinin de azalması söz konusu olabilir. Bu nedenle tekrarlanan ölçümlü denemelerde, tekrarlanacak ölçüm sayısı da oldukça önemlidir.

Bu çalışmada, iki faktörlü faktörlerden birinin tekrarlanan ölçüm içerdiği bir denemede örnek genişliğini belirlemek amacıyla kullanılan iki yaklaşımdan (I. Yaklaşım ve II. Yaklaşım) söz edilecektir. Bu yaklaşımlardan yararlanarak hesaplanan örnek genişlikleri karşılaştırmalı olarak incelenecektir.

Gereç ve Yöntem

İki faktörlü, faktörlerden birinin tekrarlanan ölçüm içerdiği bir denemede model;

$$X_{ijm} = \mu + \alpha_i + S_{j(i)} + \beta_m + \alpha\beta_{im} + e_{ijm}$$

olup burada;

X_{ijm} i. muamele grubundaki j. denekten elde edilen m. ölçüm değeri

α_i i. muamelenin etkisi ($i=1, \dots, k$)

S_{ij} i. muamele grubundaki j. deneğin rastgele etkisi ($j=1, \dots, n_i$), $N(0, \sigma_S^2)$

β_m tekrarlanan faktörün etkisi ($m=1, \dots, t$), $N(0, \sigma_\beta^2)$

$\alpha\beta_{im}$ i. muamele grubunun, tekrarlanan faktörün m. seviyesindeki etkisi

e_{ijm} tesadüften ileri gelen hata miktarıdır $N(0, \sigma^2)$.

Böyle bir denemenin kurulabilmesi için normal dağılım şartının yanı sıra, belirli ön şartların da yerine getirilmesi gerekir. Bunlardan biri; bağımlı gözlemler arasındaki korelasyon katsayılarının homojen olmasıdır. Yani bağımlı gözlemlere (tekrarlanan ölçümlere) ait kovaryans matrisinin tam simetrik olması gerekir (Gürbüz ve ark. 1999). İkinci ön şart; bağımsız faktörün seviyelerinde hesaplanan kovaryans matrislerinin homojen olmasıdır (Davis, 2002). Üçüncü şart ise küresellik (sphericity) şartıdır. Tekrarlanan ölçümler arasındaki farklara ait varyansların homojen olması anlamına gelir.

Deneme sonucunda kontrol edilecek hipotezlerden en önemlisi, muamele grupları arasındaki farkın tesadüften ileri gelip gelmediğidir.

$$H_0 : \alpha_i = \alpha_{i'}$$

$$H_1 : \alpha_i \neq \alpha_{i'}$$

Bu durumda, örnek genişliğini belirlemek amacıyla literatürde kullanılan, iki temel yaklaşım üzerinde durulacaktır.

I. yaklaşımda örnek genişliği her grupta denek sayısı eşit kabul edilerek ($n_i = n_{i'}$),

$$n \geq \frac{2\sigma^2(z_{\alpha/2} + z_\beta)^2}{d^2}$$

formülünden hesaplanır. Burada,

$$\sigma^2 = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (\hat{\mu}_{ij} - \hat{\mu}_i)^2}{\sum_{i=1}^k (n_i - 1)}$$

dır (Chow ve ark. 2003). Yani grup içi (within- group) varyanstır. d ise her zaman periyodunda ortalama değişim bakımından, iki grup arasında önemli kabul edilebilecek en küçük önemli farktır. $z_{\alpha/2}$ ve z_β ise I. ve II. Tip hata olasılıkları için standart normal dağılım kritik değerleridir. Kritik değerler bulunurken Bonferroni düzeltmesi dikkate alınır (Chow ve ark. 2003).

II. Yaklaşımda ise örnek genişliği,

$$n \geq \frac{2\sigma^2[1+(m-1)\rho](z_{\alpha/2} + z_{\beta})^2}{md^2}$$

şeklinde hesaplanır. Burada, σ^2 grup içi varyans olup, tekrarlanan ölçümler arasında doğal olarak bir korelasyon söz konusu olduğu varsayımı ile düzeltilmiştir. Her grupta ortalamalara ait varyans,

$$V(\bar{X}) = \frac{\sigma^2[1+(m-1)\rho]}{m}$$

olarak bulunur (Schouten, 1999).

Bulgular

İki yaklaşımı birbirinden ayıran belirgin özellikleri vardır. I. yaklaşım tekrarlanan ölçümler arasındaki korelasyonları ve tekrarlanan ölçüm sayısını dikkate almamaktadır. Buna karşılık II. Yaklaşım literatürde genellikle, iki muamele grubunun (kontrol ve tedavi) karşılaştırıldığı denemelerde kullanılmıştır.

Bu durumda Tablo 1 ve Tablo 2'de her iki yaklaşımla elde edilen sonuçlara dikkat edilecek olursa, tekrarlanan ölçüm sayısının dikkate alınmadığı I. yaklaşımda $\Delta = d/\sigma = 0.2$ $\alpha = 0.05$ ve $\beta = 0.05$ için her grupta örnek genişliği 650 bulunurken, II yaklaşımda tekrarlanan ölçüm sayısı ve ölçümler arasındaki korelasyon da dikkate alındığında, ancak ölçümler arasında çok yüksek korelasyon olması durumunda bu değer 585 ($m=2$ $\rho = 0.80$) olarak bulunmaktadır. Tekrarlanan ölçüm sayısı arttığı zaman gerekli örnek genişliği azalmaktadır.

Aynı zamanda standardize edilmiş etki büyüklüğü arttıkça örnek genişliği azalmaktadır (Şekil 1). Ayrıca ölçümler arasındaki korelasyonun da daha düşük olduğu durum düşünülürse, gerekli örnek genişliği daha az olacaktır (Şekil 2).

Her iki yaklaşımın da literatürde kullanılan yaklaşımlar olduğu göz önünde bulundurularak, tekrarlanan ölçümler arasındaki korelasyon hakkında hiçbir ön bilgi mevcut değilse, bu durumda I. yaklaşımın kullanılmasında bir sakınca olmadığı, ancak tekrar sayısı ve korelasyon biliniyorsa II. Yaklaşımı kullanmanın daha doğru olacağı sonucuna varılabilir. Aksi takdirde örnek genişliği gereksiz artmaktadır.

Örnek genişliğinin belirlenmesi sırasında maliyetin de önemli olduğu göz ardı edilmemelidir. Öyle ki, aynı hastadan uzun süre ölçüm alınması

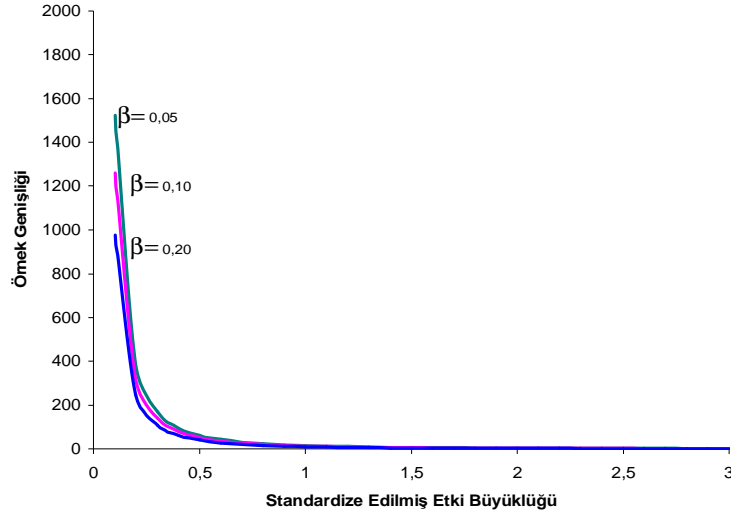
veya çok fazla hastaya uygulanması gereksiz yere maliyeti artırabilir. Bu nedenle de araştırmacı, en küçük örnek genişliği ile yetinmek zorunda kalabilir.

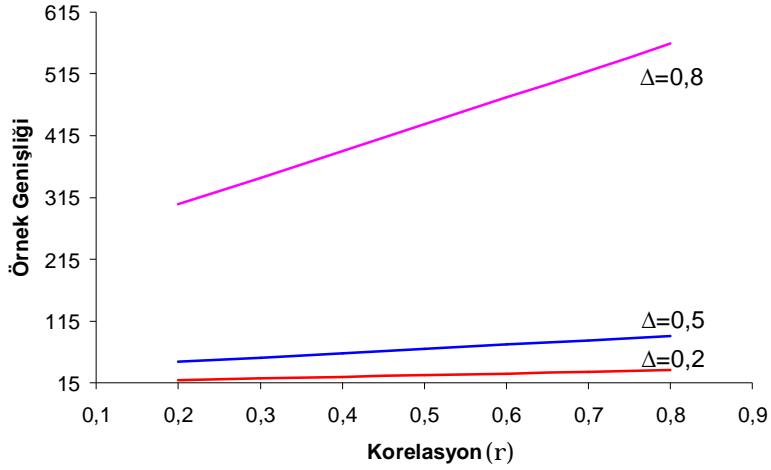
Tablo 1. Tekrarlar arası korelasyon dikkate alınmadan elde edilen örnek genişliklerine karşılık, farklı korelasyonlara sahip 2 tekrarlı denemelerde örnek genişlikleri

Δ	k=2					
	alpha 0,05			alpha 0,01		
	beta 0,05	beta 0,10	beta 0,20	beta 0,05	beta 0,10	beta 0,20
0,2	650	525	390	890	745	585
0,5	104	84	62	142	119	94
0,8	41	33	24	56	47	37
1,5	12	9	7	16	13	10
2	7	5	4	9	7	6
	k=2, m=2, $\rho=0,20$					
Δ	alpha 0,05			alpha 0,01		
	beta 0,05	beta 0,10	beta 0,20	beta 0,05	beta 0,10	beta 0,20
	0,2	390	315	236	535	447
0,5	63	51	38	86	72	57
0,8	25	20	15	34	28	22
1,5	7	6	5	10	8	7
2	4	4	3	6	5	4
	k=2, m=2, $\rho=0,50$					
Δ	alpha 0,05			alpha 0,01		
	beta 0,05	beta 0,10	beta 0,20	beta 0,05	beta 0,10	beta 0,20
	0,2	488	394	294	669	558
0,5	78	63	48	107	90	71
0,8	31	25	19	42	35	28
1,5	9	7	6	12	10	8
2	5	4	3	7	6	5
	k=2, m=2, $\rho=0,80$					
Δ	alpha 0,05			alpha 0,01		
	beta 0,05	beta 0,10	beta 0,20	beta 0,05	beta 0,10	beta 0,20
	0,2	585	473	353	802	670
0,5	94	76	57	129	108	85
0,8	37	30	23	51	42	33
1,5	11	9	7	15	12	10
2	6	5	4	9	7	6

Tablo 2. $\rho=0,20$ $\rho=0,50$ $\rho=0,80$ için $m=3$ tekrarlı denemelerde örnek genişlikleri

Δ	$k=2, m=3, \rho=0,20$					
	alpha 0,05		alpha 0,01			
	beta 0,05	beta 0,10	beta 0,20	beta 0,05	beta 0,10	beta 0,20
0,2	304	245	183	416	347	273
0,5	49	40	30	67	56	44
0,8	19	16	12	26	22	18
1,5	6	5	4	8	7	5
2	4	3	2	5	4	3
Δ	$k=2, m=3, \rho=0,50$					
	alpha 0,05		alpha 0,01			
	beta 0,05	beta 0,10	beta 0,20	beta 0,05	beta 0,10	beta 0,20
0,2	434	350	262	594	496	389
0,5	70	56	42	96	80	63
0,8	28	22	17	38	31	25
1,5	8	7	5	11	9	7
2	5	4	3	6	5	4
Δ	$k=2, m=3, \rho=0,80$					
	alpha 0,05		alpha 0,01			
	beta 0,05	beta 0,10	beta 0,20	beta 0,05	beta 0,10	beta 0,20
0,2	564	455	340	773	645	506
0,5	91	73	55	124	104	81
0,8	36	29	22	49	41	32
1,5	11	9	7	14	12	9
2	6	5	4	8	7	6

**Şekil 1.** $\alpha=0,05$ I. tip hata olasılığı ile farklı II. tip hata olasılıkları için standardize edilmiş etki büyüklüklerine karşılık örnek genişliklerinin dağılımı



Şekil 2. $\Delta=0,2$, $\Delta=0,5$ ve $\Delta=0,8$ için tekrarlanan ölçümler arasındaki farklı korelasyonlara karşılık örnek genişliğinin dağılımı

Kaynaklar

1. Gürbüz F, Ögüş E, Çamdeviren H, Kanık EA, Tekindal B. Tek Faktörlü Tekrarlanan Ölçümlü Deneme Düzeni. 4. Ulusal Biyoistatistik Kongresi. Ankara, 1999.
2. Davis CS. Statistical Methods for the Analysis of Repeated Measurements, New York, Springer, 2002.
3. Chow SC, Shao J, Wang H. Sample Size Calculations in Clinical Research. New York, Marcel Dekker, Inc. 302-307, 2003.
4. Schouten HJA. Planning Group Sizes in Clinical Trials with a Continuous Outcome and Repeated Measures. Stat. Med. 18. 255-264., 1999.
5. Ahn C, Overall JE, Tonidandel S. Sample size and power calculations in repeated measurement analysis. Computer Methods and Programs in Biomedicine 64. 121-124, 2001.

İKİ GRUP KARŞILAŞTIRMALARINDA BAĞIMLI DEĞİŞKENİN SÜREKLİ, İKİLİ ve SIRALI OLDUĞU DURUMDA ÖRNEKLEM BÜYÜKLÜĞÜNÜN ve GÜCÜN HESAPLANMASI

Özgül Vupa¹, Gül Ergör²

¹Dokuz Eylül Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü

²Dokuz Eylül Üniversitesi, Halk Sağlığı Anabilim Dalı

e-mail: ozgul.vupa@deu.edu.tr

Özet

Klinik deneyler için uygun örneklem büyüklüğünü belirlemek bu deneylere ait çalışmanın istatistiksel tasarımında en önemli adımlarından biridir. Yeterli örneklem büyüklüğünü belirlenmesi tespiti; bir çalışmanın güvenilirliğini artırmada, herhangi bir hastalığın görülme sıklığını ve o hastalığın tanı testlerinin ölçümlerinin doğruluğunu bulmada yardımcıdır.

Bu çalışmada iki grup karşılaştırmasında bağımlı değişkeninin sürekli, ikili, sıralı ya da oransal olduğunda örneklem büyüklüğünün hesaplanması ve bunun yanında da t testini kullanarak güç hesaplanması, tabloları ve şekilleri incelenmiştir.

Anahtar Kelimeler: Örneklem Büyüklüğü, Güç Analizi, İki Grup Karşılaştırılması, I. ve II. Tip Hata.

Giriş

Klinik denemeler için uygun örneklem büyüklüğünü belirlemek bu çalışmanın istatistiksel tasarımında en önemli adımlarından biridir. Yeterli örneklem büyüklüğü tespiti; bir çalışmanın güvenilirliğini artırmada, herhangi bir hastalığın görülme sıklığını ve o hastalığın tanı testlerinin ölçümlerinin doğruluğunu bulmada yardımcıdır. Ne yazık ki birçok çalışma gerekli örneklem büyüklüğü ile yapılmamasına rağmen kesin sonuçlar ile literatürde yer almaktadır. Bu durum etik olarak çok yanlıştır. Çünkü bu çalışmaların sonuçları insanlar üzerinde uygulanmaktadır.[3]

Güç analizi, etki büyüklüğü ve örneklem büyüklüğünün verildiği belirli çalışmalarda gücün düzeyini belirlemede veya isteğe bağlı olarak oluşturulmuş güç düzeyi ile önceden belirlenmiş etki büyüklüğü verildiğinde örneklem büyüklüğünü hesaplamada kullanılır. Güç analizi beş istatistiksel faktörle ilgilenir. Bunlar sırasıyla önemlilik düzeyi (α), örneklem büyüklüğü (n), örneklem varyansı (S^2), etki büyüklüğü (ES) (effect size) ve güç ($1-\beta$) şeklindedir. Verilen bu faktörlerden ilk dördü birbiri ile ilişkiliyken, sonuncusu yani testin gücü araştırmacı tarafından kolayca belirlenebilir.[6]

Yöntem

Bu çalışmada iki grup karşılaştırmasında bağımlı değişkeninin sürekli, ikili, sıralı ya da oransal olduğunda örneklem büyüklüğünün hesaplanması incelenmiştir. Ayrıca t testini kullanarak güç hesaplanması, tabloları ve şekilleri incelenmiştir. Güç analizi için gösterilen tablolar Cohen'in çalışmalarından alınmıştır.

Örneklem büyüklüğünü belirleme ve güç analizi tekniklerini öğrenme, karışık matematiksel hesaplamalar ve notasyonlar yüzünden istatistiksel paket programları tarafından daha kısa sürede yapılmaktadır.

Hipotez Testi

Verilerden sonuç çıkarabilmek için H_0 ve H_1 hipotezlerinin p değerlerinden yararlanılır. Karşılaştırılan gruplar arasında fark yoksa H_0 hipotezinden, fark varsa H_1 hipotezinden yararlanılır. Örneğin yaşam verileri ile ilgili iki deneyin etkisi karşılaştırılırken, birinci deneyden elde edilen hastalardaki yaşam oranının diğer deneydeki hastalardan elde edilen yaşam oranına eşit olma durumunu H_0 hipotezi olarak belirleriz. Bu hipotezin geçerli ve yararlı olabilmesi için veri toplama aşamasından önce H_0 hipotezi kurulmalıdır. Matematiksel olarak H_0 hipotezi farklı yollarla da ifade edilebilir. Örneğin yaşam verileri ile ilgili yeni deney etkileri araştırılırken H_0 hipotezi, yaşayan hastaların oranının birbirine eşit olduğu $P_1 = P_2$ şeklinde ifade edilirken odds oranının veya rölatif risk'lerinin 1'e eşit olduğu $OR = RR = 1$ şeklinde de ifade edilebilir. İkisinde de H_0 'ın ifade etmek istediği anlam aynıdır. Diğer yandan H_1 hipotezi, farklılığın büyüklüğünü tanımlamalıdır. Güç analizi ve örneklem büyüklüğünü belirlemede bu farklılık etki büyüklüğü adını alır.

Klinik araştırmalarda deneysel verilerden elde edilen sonuçlar kurulan hipoteze bağlı olarak kabul ya da reddedilir. Bu hipotez kabul ya da reddedildiğinde her zaman doğru sonuca varıldığı ya da varılan kararın doğru olduğu söylenemez. Burada dikkate alınması gereken iki tip hata

vardır. Bu hatalara I. tip ve II. tip hatalar adı verilir. Doğru olan H_0 hipotezinin yanlışlıkla reddedilmesi I. tip hatayı, yanlış olan H_0 hipotezinin reddedilememesi ise II. tip hatayı ortaya çıkarır. Ancak bu iki hata hiçbir zaman birlikte ortaya çıkmaz. Hipotezleri reddetme ya da etmeme ile ilgili sadece yanlış kararlar (hatalar) yoktur. Doğru kararlar da vardır. Bunların hepsi aşağıdaki şekilde de sıralanabilir.

Doğru olan H_0 hipotezini reddetme	: I. tip hata, α
Doğru olan H_0 hipotezini reddedememe	: Doğru karar, $1 - \alpha$
Yanlış olan H_0 hipotezini reddedememe	: II. tip hata, β
Yanlış olan H_0 hipotezini reddetme	: Doğru karar, Testin gücü, $1 - \beta$

I. tip hataya bağlanan olasılığa önemlilik düzeyi adı verilir. Araştırmacı I. tip hatayı azaltmak için α 'yı küçük seçerken II. tip hata üzerinde böyle bir denetim hakkı yoktur. Ancak her zaman β , α 'dan daha büyüktür. Bir çok tasarım ve planlama analizi yöntemlerinde çalışmanın gücü örneklem büyüklüğünden, etki büyüklüğünden ve α 'dan hesaplanır.

Etki Büyüklüğü

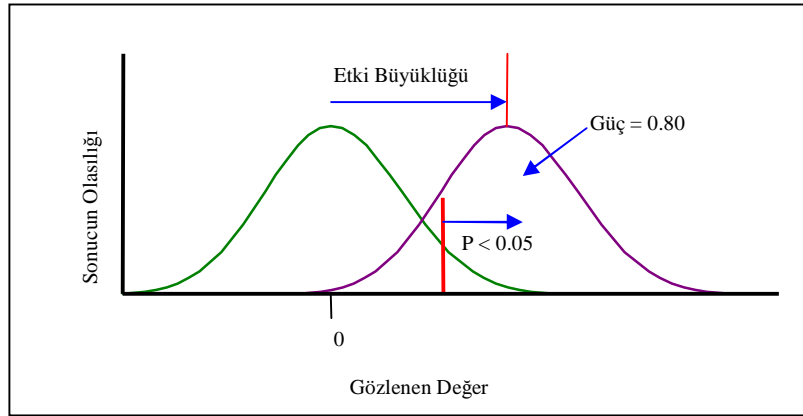
Güç analizi yöntemlerine ve örneklem büyüklüğü hesabına geçmeden önce ilk olarak istatistiksel analiz için etki büyüklüğünün ne anlama geldiğini tanımlanmalıdır. Etki büyüklüğü, örneklem ortalamaları arasındaki farkın büyüklüğünün ölçüsü veya bağımsız değişkenin etkisinin belirtme ölçüsü olarak da tanımlanabilir. İstatistiksel bir ölçüt olarak bu değer sınırı ölçüm birimlerine bağlıdır. Güç analizinde standartlaştırılmış bir yöntem türetmek için t ve F'deki birim serbestliğinde de olduğu gibi verilen tüm birimlere uygulanabilecek bir ölçüte ihtiyaç duyarız. Etki büyüklüğü örneklem ortalamaları arasındaki farkın, bu farkın kendi standart sapmasına bölme işlemi ile standartlaştırılmıştır. Tabii ki burada varyansların birbirine eşitliği de unutulmamalıdır. Etki büyüklüğü, uygun tabloları kullanma koşuluyla örneklem büyüklüğü ile güç arasındaki ilişkiyle bulunur. Güç analizine geçmeden önce etki büyüklüğü elde edilmelidir. Peki bu nasıl olacaktır?

Bunun için araştırmacı veri toplamadan önce gerekli örneklem büyüklüğünün belirlemelidir. Bu değer hesaplanabilmesi içinde aynı değişken(ler)in incelendiği önceki araştırmalardan, pilot araştırma verilerinden veya klinik deneylerden yararlanılmalıdır. Bu seçeneklerden hiçbirini gerekli kestirimi yapabilmek için kullanamıyorsak Cohen küçük, orta ve büyük etki büyüklüğü için uygun değerleri çalışmalarında elde etmiştir.

Buna göre küçük etki büyüklüğü, iyi bir deneyin olmadığı veya deneyin ya da çalışmanın anlaşılmadığı durumlarda, orta etki büyüklüğü rahatlıkla fark edilen ama önemsenmeyen durumlarda ve büyük etki büyüklüğü ise farkın fazla olduğu durumlarda kullanılır.

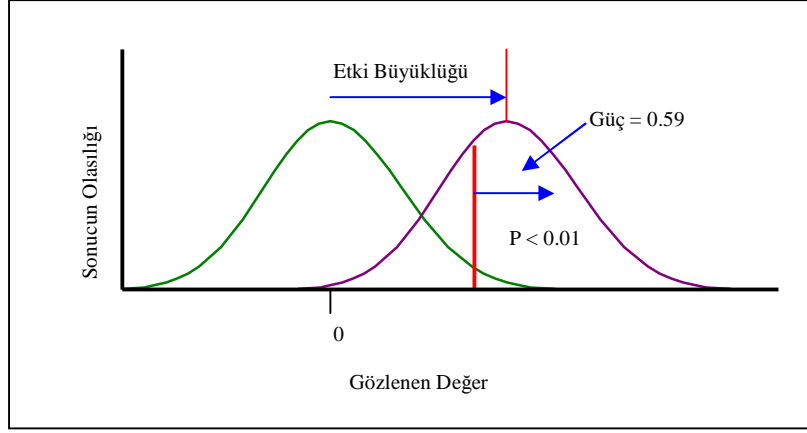
Gücün ve Diğer Faktörlerin Birbiri İle Olan Etkileşimlerinin Belirlenmesi

Güç analizinin grafiksel gösterimi Şekil 1'deki gibi ifade edilir. Sol taraftaki eğri H_0 hipotezi altında beklenen sonuçların dağılımını gösterirken, sağ taraftaki eğri ise H_1 hipotezi altında beklenen sonuçların dağılımını gösterir. Bu iki eğrinin tepe noktalarının yerlerinin farkı H_1 hipotezi tarafından tanımlanmış etki büyüklüğüdür. Çalışmanın gücü, sağ taraftaki eğride dikey çizgisinin sağında kalan bölgedir. Gücün 0.80'e eşit olması bu bölgenin altında kalan alanın tüm bölgenin % 80'nini oluşturduğu anlamına gelmektedir.

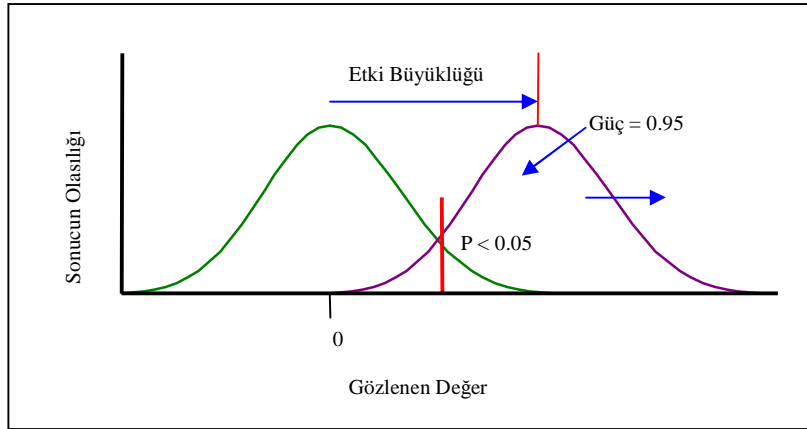


Şekil 1. Gücün Grafiksel Gösterimi

α 'nın büyümesi ya da küçülmesi gücün değişiminde rol oynar. Eğer α değeri küçülürse sağ taraftaki eğrinin altında kalan dik çizgi yani istatistiksel önemlilik büyüyecek ve sağa doğru kayacaktır. Dolayısıyla güç de azalacaktır. Eğer α ve örneklem büyüklüğünün sabit olduğu varsayımı altında, H_1 hipotezi tarafından tanımlanmış etki büyüklüğü artarsa sağ taraftaki eğri sağa doğru kayacaktır. Bu da gücün değerinin büyümesine sebep olacaktır. Bu değişimlerin gösterimi sırasıyla Şekil 2 ve 3'de ifade edilir.

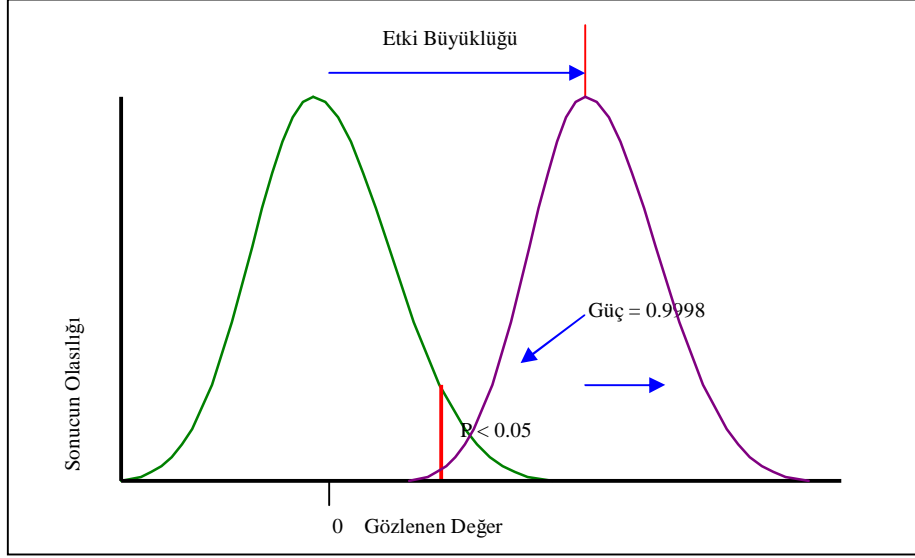


Şekil 2. Önemlilik Düzeyinin Küçülmesiyle Gücün Grafikselleştirilmesi



Şekil 3. Etki Büyüklüğünün Artmasıyla Gücün Grafikselleştirilmesi

Klinik araştırmalarda genellikle önemlilik düzeyi ve etki büyüklüğü önceden belirlenir. Örneklem büyüklüğünün artması gücü de artırır. Çünkü örneklem büyüklüğü arttıkça H_0 ve H_1 hipotezleri altındaki beklenen sonuçları gösteren eğriler birbirine yaklaşacaktır. H_0 hipotezi altındaki sonuçlar için daralan eğri yüzünden istatistiksel önemlilik kriteri sola doğru kayacaktır. Bu değişimlerin gösterimi Şekil 4'te ifade edilir.



Şekil 4. Örneklem Büyüklüğünün Artmasıyla Gücün Grafikselsel Gösterimi

Örneklem Büyüklüğünün Hesaplanması

Çalışmalarda genellikle büyük örneklem kullanmakla küçük güven aralıkları ve p değerleri elde edilir. Ama burada önemli olan sorun ne kadarlık bir örneklem büyüklüğünün gerekli olduğudur. Araştırmacı ideal bir çalışma için kullanılan testin gücünün yüksek olmasını ister. İncelenen iki grubun örneklem büyüklükleri birbirine eşitse o zaman o verilerden elde edilen gücün sayısal ölçümü de büyüktür. Ama araştırmalarda bu eşitliği her zaman sağlamak olanaklı değildir. Genellikle eşit olmayan örneklem büyüklükleri ile çalışılır. Bütün istatistiksel analizlerde olduğu gibi eşleştirilmiş veya eşleştirilmemiş örneklemelerde, olgu-kontrol çalışmalarında, logaritmik ve tekrarlı ölçümlerde kullanılan örneklem büyüklüğü hesaplamaları farklıdır. Ayrıca bağımlı değişkenin sürekli, ikili veya kategorik olduğu durumlarda da bu hesaplamalar farklılık gösterir. Sonuç olarak her biri için formüller ayrı olsa da bazı ortak karakteristik özellikleri de vardır.

İki grubun karşılaştırılmasında, bağımlı değişkenin sürekli ve normal dağıldığı durumda istatistiksel test olarak iki örneklem t testi kullanılır. Her şeyden önce örneklem büyüklüğünü hesaplayabilmek için kitlenin standart sapması bilinmelidir. Örneklem büyüklüğüne ait formül aşağıdaki gibi verilir.

$$n = \frac{2(Z_{1-\alpha/2} + Z_{1-\beta})^2}{d^2} + \frac{Z_{1-\alpha/2}^2}{4}$$

Burada n, gerekli olan örneklem büyüklüğünü ifade ederken d yani standartlaştırılmış fark $\frac{ES}{\sigma}$ 'e eşittir. Bu denklemdeki son terim düzeltme faktörüdür.

Ekte verilen Tablo 1'de, çift yanlı 0.05 önemlilik düzeyindeki farklı güç düzeylerine ait standartlaştırılmış farkın (d) farklı değerleri için gerekli örneklem büyüklüğü verilmektedir.

Lehr[5] yaptığı çalışmaya göre örneklem büyüklüğünün hesaplanmasında

$$n = \frac{16}{d^2}$$

formülünü elde etmiştir. Bu formülde d'nin küçük değerleri, α 'nın 0.05'e ve $1-\beta$ 'ninde 0.80'e eşit olması esas alınmıştır. Aksi takdirde elde edilen örneklem büyüklüğü kitlenin büyüklüğüne eşit olacaktır.

Bağımlı değişkeninin iki kategoriden oluştuğu durumda önceki çalışmalara ait oranların farklarından yararlanılır. Bu durumda etki büyüklüğü

$$ES = p_A - p_B$$

şeklinde yazılır. Burada p_A ve p_B sırasıyla iki deneme grubundaki oranlar olarak ifade edilir. İki ikili değişken arasındaki ilişki için kullanılacak istatistiksel test Pearson χ^2 testidir. Bu durumda örneklem büyüklüğünün hesabı aşağıdaki formül ile hesaplanır.

$$n = \frac{\left[Z_{1-\alpha/2} \sqrt{2p(1-p)} + Z_{1-\beta} \sqrt{p_A(1-p_A) + p_B(1-p_B)} \right]^2}{ES^2}$$

Burada $\bar{p} = (p_A + p_B)/2$ şeklindedir.

Bu formülün basitleştirilmiş halini kullanmak örneklem büyüklüğünün hesaplamasını yapmada kolaylık sağlar. Bu basitleştirilmiş formülde aşağıdaki gibi yazılabilir.

$$n = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 [p_A(1-p_A) + p_B(1-p_B)]}{ES^2}$$

$0 \leq p_A \leq 0.45$, $0.05 \leq p_B \leq 1.00$, $\alpha = 0.05$ ve $1 - \beta = 0.80$ olduğu zaman örneklem büyüklüğünün sayısal değeri ekte Tablo 2'de verilmiştir.[1]

Olgu-Kontrol gibi gözlemsel çalışmalarda iki grup arasında karşılaştırılma yaparken oranlar arası farktan çok odds oranından veya rölatif riskten yararlanır. Eğer risk faktörüne maruz kalan vakaların oranı p_A ve aynı risk faktörüne maruz kalan kontrollerin oranı ise p_B ise o zaman verilen risk faktöründe durumun etki büyüklüğü yani odds oranı aşağıdaki şekilde ifade edilebilir.[7]

$$OR = \frac{p_A(1-p_B)}{p_B(1-p_A)}$$

Bu ifade ile örneklem büyüklüğü için gerekli olan formül ise aşağıdaki gibi olur.

$$n = \frac{2(Z_{1-\alpha/2} + Z_{1-\beta})^2}{\log(OR)^2 \bar{p}(1-\bar{p})}$$

Aynı şekilde Lehr[5] yaptığı çalışmaya göre örneklem büyüklüğünün hesaplanmasında

$$n = \frac{16\bar{p}(1-\bar{p})}{(p_A + p_B)^2}$$

formülünü kullanır.

Eğer ilgilenen sonuç değişkeni likert ölçeğindeki gibi sıralı ya da oransal ölçülüyse o zaman kullanılacak olan istatistiksel test Mann-Whitney U testidir. Bu tip ölçeklendirmenin örneklem büyüklüğü hesaplamasında odds oranı kullanımı daha kolay olacağından ilk olarak etki büyüklüğünün

bilinmesi gereklidir. Sonuç değişkeninin k kategoriye sahip olduğunu düşünürsek, A grubundaki beklenen oranlar $p_{A1}, p_{A2}, \dots, p_{Ak}$ şeklinde ifade edilirken B grubu için de $p_{B1}, p_{B2}, \dots, p_{Bk}$ gibi benzer notasyonla ifade edilir. Bu olasılıkların toplamının 1'e eşit olduğu $p_{A1} + p_{A2} + \dots + p_{Ak} = 1$ ve $p_{B1} + p_{B2} + \dots + p_{Bk} = 1$ eşitlikleri ile gösterilebilir. $c_{A1}, c_{A2}, \dots, c_{Ak}$ notasyonları ile gösterilen toplamsal olasılıklarda c'ler $c_{A1} = p_{A1}$, $c_{A2} = p_{A1} + p_{A2}$ ve benzer eşitliklerle gösterilir. Kategori sayısına bağlı olarak ilk kategori için odds oranı

$$OR = \frac{c_{A1}/(1 - c_{A1})}{c_{B1}/(1 - c_{B1})}$$

şeklinde ifade edilirken bu oranlar k-1'e kadar devam eder. Sonuç olarak sıralı ya da oransal ölçülü bağımlı değişken için örneklem büyüklüğünün hesabı aşağıdaki gibidir.

$$n = \frac{6(Z_{1-\alpha/2} + Z_{1-\beta})^2 / (\log OR)^2}{\left[1 - \sum_{i=1}^k p_i^{-3}\right]}$$

Burada OR, i. kategoride olan hastaların i ve i'den daha az olanları içeren odds oranı iken k, kategori sayısı ve \bar{p}_i ise i. kategoride beklenen ortalama oran olarak ifade edilir. Ayrıca burada bütün kategori çiftlerinin odds oranının birbirine eşit olduğu varsayımı da vardır. Bu varsayım Mann-Whitney U testinin kullanılmasını sağlar. Sıralı kategorik verileri için çift yanlı 0.05 önemlilik düzeyinde, farklı güç ve odds oranı için $6(Z_{1-\alpha/2} + Z_{1-\beta})^2 / (\log OR)^2$ değerleri ekteki Tablo 3'de verilmiştir. Eğer kategorilerin sayısı çok fazla ise oranların tahmini de zor olacaktır. Ayrıca kategori sayısı 5'den küçük olduğu durumda düzeltme faktörü kullanılır.

Örneklem büyüklüğü tespitinden sonra güç analizine geçmek daha anlamlıdır. t testi için güç hesaplaması iki grup karşılaştırılmasında kullanılırlar. Literatürde bu formüllerden çok oluşturulmuş tablolardan yararlanılır.

t Testi İçin Güç Analizi

t testinde kullanılan güç analizi gruplandırılmış ve gruplandırılmamış veri setleri için kullanılabilir. Etki büyüklüğü ölçütü d ile gösterilir. Eşit varyanslı gruplandırılmamış veriler için etki büyüklüğünün hesaplanması aşağıdaki gibidir.

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S}$$

Burada, \bar{X}_1 ve \bar{X}_2 grup ortalamaları iken, S ise ortak standart sapmadır.

Varyans homojenliği olmadığında yani eşit olmayan varyans durumunda d 'nin hesaplanmasında sadece standart sapma farklı bir şekilde hesaplanır. Bu hesaplama aşağıdaki gibi ifade edilir.

$$S = \sqrt{\frac{S_1^2 + S_2^2}{2}}$$

Tekrarlı ölçüm tasarımlarından elde edilen verilerde d 'nin hesaplanması eşleştirilmiş skorlara dayanır. İlk olarak ortak standart sapma ile iki test durumu için ortalamaları kullanarak d' hesaplanır. Daha sonra bu d' değeri ile d hesaplanır. Bu hesaplamaların gösterimi sırasıyla aşağıdaki gibi verilir.

$$d' = \frac{\bar{X}_1 - \bar{X}_2}{S}$$

$$d = \frac{d'}{\sqrt{1-r}}$$

Burada r , gruplandırılmış veriler için korelasyon katsayısıdır. r 'nin hesabının olmadığı durumda d'

$$d' = \frac{\bar{d}}{S_d}$$

şeklinde hesaplanır. Burada \bar{d} farklı skorların ortalaması iken, S_d farklı skorların standart sapmasıdır. Güç tablolarında kullanılan d 'nin değeri aşağıdaki gibidir.

$$d = d' \sqrt{2}$$

d direkt olarak hesaplanamadığında aşağıda verilen etki büyüklüğü ölçütleri kullanılır.

Küçük Etki Büyüklüğü İçin	d = 0.20
Orta Etki Büyüklüğü İçin	d = 0.50
Büyük Etki Büyüklüğü İçin	d = 0.80

Örneğin bir klinik deneyde kolesterol düzeyini düşüren iki ilaç karşılaştırılmak isteniyor. A ilacının önceki çalışmalarda kolesterolü 20 mg/dl düşürdüğü biliniyor. B ilacının bundan daha farklı sonuç vermesi bekleniyor. İki ilacın farkının orta etki büyüklüğünde olduğu varsayılırsa %80 güç ve %5 I. tip hatayla kaç kişilik gruplar üzerinde çalışılması gerektiğinin hesabı aşağıdaki şekillerde yapılabilir.

Portney ve Watkins'in çalışmasında yer alan Cohen'den aldıkları Tablo 6'a göre $\alpha = 0.05$, $1 - \beta = 0.80$ ve orta etki büyüklüğü yani $d = 0.5$ olduğunda t testi için kullanılan örneklem büyüklüğü $n = 64$ olarak bulunur.

Campbell ve arkadaşlarının yayınladıkları makaleye göre

$$n = \frac{2(Z_{1-\alpha/2} + Z_{1-\beta})^2}{d^2} + \frac{Z^2}{4} \text{ formülü ile}$$

$$n = \frac{2(Z_{0.975} + Z_{0.80})^2}{(0.5)^2} + \frac{(Z_{0.975})^2}{4} = \frac{2(1.96 + 0.84)^2}{(0.5)^2} + \frac{(1.96)^2}{4} = 63.68 = 64$$

olacak şekilde bulunur.

Ayrıca yine aynı makalede Godfrey ve Lehr'in yaklaşımı da verilmiştir. Godfrey ve Cohen'in yaklaşımı aynıken Lehr'in

$$n = \frac{16}{d^2} = \frac{16}{(0.5)^2} = 64 \text{ olacak şekilde ifade edilir.}$$

Görüldüğü gibi bütün formüllerde ve farklı tablolarda sonuç aynı çıkmaktadır.

Ekte t testi için kullanılan güç tabloları ve örneklem büyüklüğü tablosu hakkında bilgi vardır.

Sonuç

Güç analizini yaparken önemlilik düzeyinden, etki büyüklüğünden, örneklem büyüklüğü ve örneklem varyansından yararlanır. Etki büyüklüğü için örneklem büyüklüğü, önemlilik düzeyi ve güç bilinmelidir. Eğer kaynaklar kısıtlıysa çalışmanın gücünü artırmak için etki büyüklüğü artırılır. Fakat bu pratikte mümkün olmaz. Herseyden önce örneklem büyüklüğü hesaplamaları için çalışma çok iyi planlanmalıdır. Farklı tasarımlarda farklı örneklem büyüklüğü hesabı kullanılır. Bu farklı tasarımlar iki örneklem karşılaştırılmasında bağımlı değişkeninin sürekli, ikili ve sıralı olduğu durumlar ile yaşam eğrilerinin karşılaştırılması ve bioeşdeğerlilik çalışmalarında kendini gösterir. t testinde kullanılan güç analizi gruplandırılmış ve gruplandırılmamış veri setleri için kullanılır. Testin gücü ve örneklem büyüklüğü hesaplamaları bu çalışmadaki formüllerle de hesaplanabildiği gibi genellikle paket programları ve daha önceki çalışmalarda oluşturulmuş tablolardan yararlanılarak yapılır.

Ekler

(1., 2. ve 3. Tablolar Campbell'in çalışmasından alınırken, t Testi için kullanılan Tablolar Cohen'in çalışmasından alınmıştır)

Tablo 1'de, çift yanlı 0.05 önemlilik düzeyindeki farklı güç seviyelerine ait standartlaştırılmış farkın (d) farklı değerleri için gerekli örneklem büyüklüğü verilmiştir.

d	Güç				
	99	95	90	80	50
0.10	3676	2600	2103	1571	770
0.20	920	651	527	394	194
0.30	410	290	235	176	87
0.40	231	164	133	100	49
0.50	148	105	86	64	32
0.60	104	74	60	45	23
0.70	76	54	44	33	17
0.80	59	42	34	26	13
0.90	47	34	27	21	11
1.00	38	27	22	17	9
1.10	32	23	19	14	8
1.20	27	20	16	12	7
1.30	23	17	14	11	6
1.40	20	15	12	9	5
1.50	18	13	11	8	5

Tablo 2'de $0 \leq p_A \leq 0.45$, $0.05 \leq p_B \leq 1.00$, $\alpha = 0.05$ ve $1 - \beta = 0.80$ olduğu zaman örneklem büyüklüğünün sayısal değeri verilmiştir.

p_A	P_B																			
	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1.0
0.00	152	74	48	35	27	22	18	15	13	11	10	8	7	6	6	5	4	4	3	2
0.05		435	141	76	49	36	27	22	18	15	12	11	9	8	7	6	5	4	4	3
0.10			686	199	100	62	43	32	25	20	16	14	11	10	8	7	6	5	4	4
0.15				906	250	121	73	49	36	27	22	17	14	12	10	8	7	6	5	4
0.20					1094	294	138	82	54	39	29	23	18	15	12	10	8	7	6	5
0.25						1251	329	152	89	58	41	31	24	19	15	12	10	8	7	6
0.30							1377	356	163	93	61	43	31	24	19	15	12	10	8	6
0.35								1471	376	170	96	62	43	31	24	18	14	11	9	7
0.40									1534	388	173	97	62	42	31	23	17	14	11	8
0.45										1565	392	173	96	61	41	29	22	16	12	10

Tablo 3'de sıralı kategorik verileri için çift yanlı 0.05 önemlilik düzeyinde, farklı güç ve odds oranı için $6(Z_{1-\alpha/2} + Z_{1-\beta})^2 / (\log OR)^2$ değerleri verilmiştir.

Odds Oranı	Güç				
	99	95	90	80	50
0.75	1331.97	942.09	761.77	569.03	278.50
1.25	2213.86	1565.85	1266.13	945.78	562.89
1.50	670.52	474.25	383.48	286.45	140.20
1.75	352.00	248.96	201.31	150.38	73.60
2.00	229.44	162.60	131.22	98.02	47.97
3.00	91.33	64.60	52.23	39.02	19.10
4.00	57.36	40.97	32.80	24.50	11.99
5.00	42.56	30.10	24.34	18.18	8.90
10.00	20.79	14.71	11.89	8.88	4.35

t Testi İçin Kullanılan Güç Tabloları

Tablo 4'de $\alpha = 0.05$ (tek yönlü) önemlilik düzeyinde verilmiş örneklem büyüklüğü için t testinin güç değerleri verilmiştir.

n	d										
	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	1.00	1.20	1.40
8	7	10	13	19	25	31	38	46	61	74	85
9	7	11	15	20	27	34	41	50	66	79	88
M	M	M	M	M	M	M	M	M	M	M	M
30	10	19	31	46	61	74	85	92	99		
M	M	M	M	M	M	M	M	M	M	M	M
60	13	29	50	70	86	95	98				
M	M	M	M	M	M	M	M	M	M	M	M
900	68										
1000	72										

Tablo 5'de $\alpha = 0.05$ (çift yönlü) önemlilik düzeyinde verilmiş örneklem büyüklüğü için t testinin güç değerleri verilmiştir.

n	d										
	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	1.00	1.20	1.40
8	5	7	9	11	15	20	25	31	46	60	73
9	5	7	9	12	16	22	28	35	51	65	79
M	M	M	M	M	M	M	M	M	M	M	M
30	7	17	21	33	47	63	76	86	97		
M	M	M	M	M	M	M	M	M	M	M	M
60	8	19	37	58	77	90	97	99			
M	M	M	M	M	M	M	M	M	M	M	M
900	56	99									
1000	61	99									

İki örneklem için verilen örneklem büyüklükleri birbirine eşit değil ise bu tabloları kullanabilmek için bu iki örneklem büyüklüğünün harmonik ortalamasını hesaplamak gereklidir. ($n' = (2n_1n_2 / (n_1 + n_2))$)

t Testi İçin Kullanılan Örneklem Büyüklüğü Tablosu

Tablo 6'da gücün verilen seviyesinde t testi için gerekli örneklem büyüklüğü verilmektedir. ($\alpha_1 = 0.025$ $\alpha_2 = 0.05$)

Power	d										
	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	1.00	1.20	1.40
0.70	1235	310	138	78	50	35	26	20	13	10	7
0.75	1389	348	155	88	57	40	29	23	15	11	8
0.80	1571	393	175	99	64	45	33	26	17	12	9
0.85	1797	450	201	113	73	51	28	29	19	14	10
0.90	2102	526	234	132	85	59	44	34	22	16	12
0.95	2600	651	290	163	105	73	54	42	27	19	14
0.99	3675	920	409	231	148	103	76	58	38	27	20

d'nin tam değerinin tabloda olmadığı durumda örneklem büyüklüğünün hesabı aşağıdaki gibidir.

$$n = \frac{n_{0.10}}{100d^2} + c$$

Burada $n_{0.10}$, $d = 0.10$ için verilen örneklem büyüklüğüdür. c ise önemlilik düzeyinin seviyesine bağlı sabit bir sayıdır. Cohen'in kitabına göre c 'nin değeri aşağıdaki gibi alınır.

$\alpha_1 = 0.05$	$(\alpha_2 = 0.10)$	$c = 0.7$
$\alpha_1 = 0.025$	$(\alpha_2 = 0.05)$	$c = 1.0$
$\alpha_1 = 0.01$	$(\alpha_2 = 0.02)$	$c = 1.5$
$\alpha_1 = 0.005$	$(\alpha_2 = 0.01)$	$c = 1.5$

Kaynaklar

1. Campbell M.J., Julious S.S. and Altman D.G., Estimating Sample Sizes for Binary, Ordered Categorical and Continuous Outcomes in Two Group Comparisons, BMJ, 1995, V: 311, p: 1145-1148.

2. Cohen J : Statistical Power Analysis for the Behavioral Sciences, ed 2, New York, Academic Press, 1977.
3. Kleinbaum D., Kupper F. And Morgenstern H., Epidemiologic Research / Principles and Quantitative Methods, New York, Van Nostrand Reinhold, 1982.
4. Lachin J.M., Introduction to Sample Size Determination and Power Analysis in Clinical Trials, 1981, V: 2, p:93-113.
5. Lehr R., Sixteen s squared over d squared: a relation for crude sample size estimates, Statistics in Medicine, 1992, V: 11, p: 1099-1102.
6. Portney L.G. & Watkins M., Foundation of Clinical Research Applications to Practice, 1993.
7. Rosner B., Fundamentals of Biostatistics, ed 5, USA, Duxbury, 2000.
8. Whitley E. & Ball J., Sample Size Calculations, Biomed Central Ltd, 2002, V: 6, p: 335-341.

WILCOXON İŞARET TESTİNDE ÖRNEKLEM BÜYÜKLÜĞÜ VE GÜÇ HESAPLAMA YAKLAŞIMLARI

S.Yavuz Sanisoğlu¹, Mesut Akyol², Osman Saraçbaşı³, İlker Etikan⁴,
Zeki Akkuş⁵

GATA Biyoistatistik BD¹

GATA²

Hacettepe Üniversitesi Biyoistatistik ABD³

GOP Üniversitesi Biyoistatistik ABD⁴

Dicle Üniversitesi Biyoistatistik ABD⁵

e-mail: biyoistatistik@gata.edu.tr

Özet

Bilimsel çalışmalarda çoğunlukla Tip I hata olarak adlandırılan hata dikkate alınmakta, buna karşılık Tip II hataya yer verilmemektedir. Son zamanlarda özellikle örneklem büyüklüğünün belirlenmesi ya da güç parametresinin hesaplanmasında yardımcı olmak üzere geliştirilen paket programların yaygınlaşması bu alandaki yetersizliği kısmen de olsa azaltmıştır. Bununla birlikte özellikle parametrik olmayan testler için bu parametrelerin hesaplanmasına yönelik kısımların olmaması hala eksikliğini göstermektedir.

Bu çalışmamızda en sık kullanılan parametrik olmayan testlerden birisi olan Wilcoxon işaret testi için örneklem büyüklüğünün ve testin gücünün hesaplanmasına yönelik yaklaşımlar incelenmeye çalışılmıştır. Bu bağlamda, bu konuda yapılmış en kapsamlı çalışma olan Noether formülü yanında diğer yaklaşımlar da incelenmiştir.

Sonuç olarak; Wilcoxon işaret testi için örneklem büyüklüğü ve güç hesabı için, diğer çalışma düzenlerinde olduğu gibi araştırma konusu hakkında pozitif ve negatif ortalama sıra numaraları, etki genişliği gibi bazı ön bilgilerin gerekli olduğu, bu ön bilgilerin sağlanması durumunda arzu edilen güce sahip örneklem büyüklüğünün hesaplanabileceği ortaya konmuştur.

Anahtar Kelimeler: Wilcoxon işaret testi, örneklem büyüklüğü, güç.

Giriş

Hipotez testlerinde test sonucunda yokluk hipotezi ya kabul ya da reddedilir. Yokluk hipotezinin kabul ya da reddedilmesi, hipotezin gerçek durumuna göre tip I ya da tip II hataya düşülmesine neden olur. Evrenin tümü yerine örneklem üzerinde yapılacak çalışmalarda bu hataların ortaya çıkması kaçınılmazdır. Ancak yeterli sayıda deneğin çalışmaya alınması ile hata oranları belli bir düzeye indirgenebilir.

Bilimsel bir çalışmanın planlanmasında en sık karşılaşılan ve cevaplanması da bir o kadar zor olan soru çalışmaya kaç deneğin alınacağı sorusudur (1). İyi planlanmış bir çalışma için örneklem büyüklüğünün hesaplanması büyük önem taşır (2). Gereğinden daha az örneklem üzerinde çalışmak, elde edilecek sonuçların güvenilirliklerini olumsuz yönde etkilerken, gereğinden fazla deneğin kullanılması da kaynakların boş yere kullanılmasına neden olur (3). Bu nedenle bilimsel çalışmanın planlanması aşamasında örneklem büyüklüğünün hesaplanması büyük önem taşır. Örneklem büyüklüğünün hesaplanmasında çeşitli parametreler yanında güç kavramı büyük önem taşır. Örneklem büyüklüğü yanında hesaplanması gereken bir diğer parametre olan güç, gerçekte yanlış olan hipotezin reddedilme olasılığıdır (4). Genel olarak güç formülü;

$$P[Z \leq z_{\alpha} \cdot (\text{Std.Hata})] + P[Z > z_{\alpha} \cdot (\text{Std.Hata})]$$

şeklinde tanımlanır (4).

Gereç ve Yöntem

Örneklem büyüklüğü üç temel nedenden dolayı önem taşır. Bunlar; Gereğinden daha az sayıda ya da gereğinden fazla sayıda örneklem büyüklüğünün, kullanılacak sonuçların elde edilememesi ve bu nedenle kaynakların gereksiz yere harcanması ya da ileri derecede anlamlılığın fazla bir önem taşımadığı durumlarda kaynakların boş yere harcanması gibi ekonomik ve aynı zamanda etik, yeterli örneklem büyüklüğüne sahip olmayan negatif sonuçlu bir denemenin, klinik olarak önemli (istatistiksel olarak önemsiz) bir etkiyi dikkate almaması ya da incelenen tedavinin yararı olmadığı sonucunun elde edilmesi gibi bilimsel gerekçelerdir (5).

Klinik çalışmalarda örneklem büyüklüğünün hesaplanmasında gerek duyulan parametreler çalışmanın deseni, önemlilik düzeyi, güç, en küçük önemli klinik farklılık ve kullanılacak hipotez testi ile ilişkilidir (6).

Her bir istatistiksel analiz yöntemi ve çalışma düzeni (eşleşmemiş çiftler, olgu-kontrol, tekrarlı ölçümlü kesitsel çalışmalar gibi) istenilen güçle

örneklem büyüklüğü arasında değişik formüller bulunur. Bütün bu formüllerin hemen hemen hepsi aynı değişkenlere gerek duyarlar. Bütün bunlarda ortak olan özellik, ilgi konusu özelliğin normal dağılıma uygun olup olmadığına göre matematiksel düzenlemeler yapmalarıdır (z istatistiği). Bu yüzden genellikle formülün payında $(Z_{\alpha/2} + Z_{\beta})^2$ terimi yer alır. Örneğin sürekli değişkenler için tekrarlı ölçümlerde örneklem büyüklüğü;

$$N = \frac{2}{\delta^2} \left(\sigma_b^2 + \frac{\sigma^2}{k} \right) (Z_{\alpha/2} + Z_{\beta})^2$$

dir. Genel yapı ise

$$N = \text{Sabit} \cdot \frac{(Z_{\alpha/2} + Z_{\beta})^2}{(\text{Etki büyüklüğü})^2}$$

şeklinde (3).

Wilcoxon işaret testi (WIT) için pozitif ve negatif farkların ortanca değere göre dağılımlarının binom dağılımı gösterdikleri göz önüne alınırsa binom dağılımı için

$$n = p \cdot q \cdot \left(\frac{Z}{E} \right)^2$$

şeklindeki örneklem büyüklüğü bağıntısından yararlanılabilir. Ancak hem bu bağıntı, hem de tek örneklem için parametrik test örneklem büyüklüğü bağıntısının E ile gösterilen etki büyüklüğü parametresine bağlı olmasından ötürü pek kullanışlı değildirler. Bunların yerine aralarında çok az farklılık olan iki değişik yöntem kullanılmaktadır.

Wilcoxon işaret testi için örneklem büyüklüğünün hesaplanmasında kullanılan iki temel yaklaşımdan birincisi WIT'in iki gözlem arasındaki farklar üzerinden işlem yapmasından ötürü tek örneklem için örneklem büyüklüğü bağıntılarının kullanılmasıdır. Bu yaklaşım için iki değişik yöntem kullanılabilir. Bunlardan birincisi farkların normal dağıldığı durumda tek örneklem için parametrik örneklem büyüklüğü

$$n = \frac{(t_{n-1, \alpha/2} + t_{n-1, \beta})^2}{\delta^2}$$

bağıntısının kullanılması ve buradan elde edilen sonucun Pitman asimptotik bağıl etkinlik katsayısı (Asymptotic Relative Efficiency, ARE) katsayısı ile çarpılmasıdır. Bağıntıda δ , etki genişliğini göstermekte olup Pitman ARE katsayısı 0.95 dir (6).

Wilcoxon işaret testinde farkların çarpık dağılım göstermesi durumunda ise "İşaret testi için örneklem büyüklüğü" (7) ya da Noether (8) tarafından önerilen bağıntının kullanılması ise başvurulan diğer yaklaşımlardır.

İşaret Testi örneklem büyüklüğü yaklaşımı:

X_1, X_2, \dots, X_N ortancası η olan bir populasyondan çekilmiş örnekleme gösterebilir. $\eta = \eta_0$ olup olmadığı test edilmek istensin. Basitleştirmek açısından $\eta_0 = 0$ kabul edilsin.

$$S = \text{Pozitif gözlem sayısı} \quad \mu(S) = N \cdot p \quad p = P(X > 0)$$

$$\mu_0 = N/2 \quad \sigma^2(S) = N \cdot p \cdot (1-p) \quad \text{ve} \quad \sigma_0^2(S) = N/4$$

dür. Buradan $Q(S) = 4 \cdot N \cdot (p - 1/2)^2$ ve $N = N(S)$ örneklem büyüklüğü;

$$N(S) = \frac{(Z_\alpha + Z_\beta)^2}{4 \cdot (p - \frac{1}{2})^2} \quad (1)$$

dir. (1) nolu bağıntı gerçek olasılık p nin yani teorik medyanı aşan bir gözlemin teorik olasılık $1/2$ den ne kadar saptığına bağlıdır.

Bu sonuç araştırmacı ile bir başka soruyu baş başa bırakır. "Alternatif olarak hangi p değeri seçilmelidir?". Bu değer, daha önce yapılan çalışmalarda ya da yapılacak bir pilot çalışma ile belirlenebilir. Bunlardan başka araştırmacı

$$r = \frac{p}{q} = \frac{+ \text{ gözlem sayısı}}{- \text{ gözlem sayısı}}$$

şeklinde bir odds oranı da belirleyebilir. Bununla ilişkili p değeri; $\frac{r}{1+r}$ dir.

Bu değerler örneklem büyüklüğü formülünde yerlerine konulup N değeri hesaplanabilir.

$$\rho = \frac{\sigma(s)}{\sigma_0(s)} = 2\sqrt{p \cdot (1-p)}$$

Alternatif hipotez altında $\rho < 1$ olup $N(S)$ değişmezdir. Değişmezlik derecesi $N(S)$ cinsinden $N(S) / N(p)$ ile belirlenir. Burada

$$N(\rho) = \frac{(Z_\alpha + \rho \cdot Z_\beta)^2}{4 \cdot (\rho - \frac{1}{2})^2}$$

$\sigma(s)$ gerçek değerine bağlı örneklem büyüklüğüne normal yaklaşımdır. $\beta = \alpha$ olduğunda

$$\frac{N(S)}{N(\rho)} = \frac{4}{[1 + 2\sqrt{p \cdot (1-p)}]^2}$$

olur (7).

Noether yöntemi

Noether tarafından WIT için ileri sürülen örneklem büyüklüğünün hesaplaması ise şu şekildedir;

$$W = \sum \text{Rank} | X | \quad W = \# [\text{Pozitif } (X_i + X_j)] \quad 1 \leq i \leq j \leq N$$

$$\mu(W) = N \cdot p + \frac{N \cdot (N-1) \cdot p'}{2} \quad p = P(X > 0) \text{ ve } p' = P(X + X' > 0)$$

X ve X' bağımsız iki gözlemdir. H_0 altında $p = p' = 1/2$ olup

$$\sigma_0^2(W) = \frac{N(N+1) \cdot (2N+1)}{24}$$

ve

$$Q(W) = \frac{[N \cdot (p - \frac{1}{2}) + \frac{1}{2} N \cdot (N-1) \cdot (p' - \frac{1}{2})]^2}{\frac{N \cdot (N+1) \cdot (2N+1)}{24}} = 3 \cdot N \cdot (p' - \frac{1}{2})^2$$

ve buradan da gerekli örneklem büyüklüğü

$$N(W) = \frac{(Z_{\alpha} + Z_{\beta})^2}{3.(p' - \frac{1}{2})^2} \quad (2)$$

olur. Uygun p nin seçimi p' nün seçimi ile aynıdır. Verilen bir Odds oranı için

$$r' = \frac{P(X + X' > 0)}{P(X + X' < 0)} \quad \text{ve} \quad p' = \frac{r'}{(1 + r')}$$

dür (8).

Yukarıda verilen (1)ve (2) nolu bağıntılarından önceden belirlenen p , p' ve α yanılma düzeyi değerleri için sırasıyla işaret testi ve WIT için

$$Z_{\beta} = 2.(p - \frac{1}{2}).\sqrt{N} - Z_{\alpha}$$

ve

$$Z_{\beta} = (p' - \frac{1}{2}).\sqrt{3.N} - Z_{\alpha}$$

şeklinde Z_{β} değerleri ve bu değerlerden de işaret testi ve WIT için güç değerleri bulunabilir.

İşaret testi ve Noether yöntemlerine ilişkin (1) ve (2) nolu bağıntılar kullanılarak $\alpha=0.95$, $\beta = 0.10$ ve $p=\{0.10,0.11,\dots,0.20\}$ için elde edilen örneklem büyüklükleri Tablo I'de gösterilmiştir.

Tablo I. İşaret testi (İTY) ve Noether (NY) yaklaşımlarına göre hesaplanan örneklem büyüklükleri

Yöntem	p değerleri										
	0.10	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
İTY	11	11	12	12	13	14	15	15	16	17	18
NY	12	13	13	14	15	16	17	18	20	21	23

Tablo I'de de görüldüğü gibi aynı girdi parametreleri için Noether yaklaşımına göre bulunan örneklem büyüklüğü değerleri işaret testi yaklaşımına göre bulunan değerlerden daha yüksektir.

Tartışma ve Sonuç

Wilcoxon işaret testi, bakış açısı olarak işaret testine benzemekle beraber işaret testine göre daha hassas bir testtir. Büyük örneklem için hemen hemen t testi kadar hassastır. Dağılımı bilinmeyen küçük örneklem için ise t testine göre daha hassastır. H_0 hipotezi olarak (x_i, x_j) çiftleri arasındaki farkların ortanca değerine göre simetrikliğini kabul eder (9). WIT için örneklem büyüklüğü ve güç hesaplamalarında bu yüzden ortancaya göre fark değerlerin pozitiflik ya da negatiflik durumları göz önüne alınmalıdır. Bu parametrelerin hesaplanmasında işaret testi için hesaplama ya da Noether tarafından önerilen ve pozitif farkların oranını kullanan iki temel yöntem kullanılmaktadır. Bu çalışmada bu iki yöntem üzerinde durulmuştur.

Önceden planlanmamış, örneklem büyüklüğü hesaplanmamış, gücü hesaplanmamış bir çalışmanın bilimselliğinden söz etmek etik bir davranış olmayacaktır (10). Bu yüzden klinik çalışmalarda sıkça kullanılan WIT için örneklem büyüklüğü ve güç değerinin hesaplanması büyük önem taşımaktadır.

Kaynaklar

1. Armitage P, Berry G, Matthews J N S. Statistical Methods in Medical Research, 4 th edition, USA: Blackwell Science Ltd; 2002, 137-139
2. Hinkle D E, Wiersma W, Jurs S G. Applied Statistics for the Behavioral Sciences, 4 th edition, USA: Houghton Mifflin Company; 1998, 311-313
3. Lewis R J. Power and Sample Size Determination: Concepts and Software Tools, Bildiri, 2000 Annual Meeting of the Society for Academic Emergency Medicine, USA
4. Zar J H, Biostatistical Analysis, 3rd edition, USA: Prentice Hall; 1996
5. Thabane L. Sample Size Determination in Clinical Trials, MJA 2002, 177(5): 256-257.
6. www.astrostatistics.psu.edu/su05/arnold/npara060805.pdf
26.06.2005:12:06
7. Jones B, Kenward M G. Design and Analysis of Cross-Over Trials, 2 nd edition, UK: Taylor&Francis CRC Press; 2003
8. Noether GE. Sample Size Determination for Some Common Nonparametric Tests. J American Statistical Association 1987, 82(398): 645-647

9. The Wilcoxon Matched-Pairs Signed Ranks Test, http://www.fon.hum.uva.nl/Service/Statistics/Signed_rank_Test.html, 02.05.2005,15:54
10. Sümbülođlu V, Sümbülođlu K. Klinik ve Saha Arařtırmalarında Örnek-
leme Yöntemleri ve Örneklem Büyüklüğü, Ankara: 2005, 100

ÖRNEK GENİŞLİKLERİNİN KÜÇÜK OLDUĞU DURUMLARDA KORELASYON KATSAYISINA AİT GÜVEN ARALIKLARI

Ersin Ögüş¹, Ayşe Canan Yazıcı¹, Fikret Gürbüz²

*Başkent Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı¹
Ankara Üniversitesi Ziraat Fakültesi, Zootekni Bölümü, Biyometri-Genetik
Anabilim Dalı²*

e-mail: eogus@baskent.edu.tr

Özet

Korelasyon katsayısına ait hipotez kontrolü yapılırken, örnek genişliğinin küçük ve populasyona ait korelasyon katsayısının büyük olduğu durumlarda Fisher'in Z transformasyonunun yanlış kararlar verilmesine neden olabileceği bilinmektedir. Bu durumu vurgulamak ve böyle durumlarda kullanılan diğer yöntemlere alternatif bir yaklaşım geliştirmek amacıyla 5 farklı örnek genişliği ve her bir örnek genişliği için populasyona ait korelasyon katsayısının 5 farklı değeri dikkate alınarak normal dağılım gösteren örnekler üreten bir simülasyon programı yazılmıştır. Bu program ile bu örneklerin her birinden 1 000 000 defa üretilerek korelasyon katsayıları hesaplanmış ve sonrasında % 95 olasılıktaki güven aralıkları oluşturulmuştur. Elde edilen değerler Fisher'in Z transformasyonu ve düzeltilmiş Fisher Z transformasyonu kullanılarak elde edilen sonuçlar ile karşılaştırılmıştır. Çok küçük örnek genişliklerinde çalışmamızdan elde edilen güven aralıklarının daha dar olduğu görülmüştür.

Anahtar kelimeler: Güven aralıkları, korelasyon katsayısı, önem kontrolü.

VARYANS ANALİZİ UYGULAMALARINDA ÇOKLU KARŞILAŞTIRMA TEKNİKLERİNİN 1. TİP HATA VE GÜÇ BAKIMINDAN KARŞILAŞTIRILMASI

İlter Helvacı, E. Arzu Kanık

Mersin Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, MERSİN

e-mail: ilterhelvacı@mersin.edu.tr

Özet

Bu çalışmada dengeli (balanced) ve dengeli olmayan varyans analizinden sonra uygulanan çeşitli çoklu karşılaştırma metotları 1.tip hata ve güçleri bakımından incelenmiştir. Çoklu karşılaştırma metotları arasında farklılık yaratan prensipler, deney öncesi (a priori) ve deney sonrası (a posteriori, post-hoc) planlama, bağımlılık (non-ortagonal), bağımsızlık (ortagonal) kavramları, hatalı keşif oranları ve çoklu karşılaştırma metotlarının ayırt edici özellikleri belirtilmiştir.

Varyans analizindeki hata oranları ailesel hata oranı, deneysel hata oranı (experimentwise error rate) ve karşılaştırma başı hata oranı (per comparison error rate) olarak gruplandırılabilir.

Son yıllarda microarray çalışmalarında hipotez testleri sonucu karşımıza çıkan bir çoklu karşılaştırma hatası da False Discovery Rate'dir. Buna ait düzeltmede en yaygın kullanılan metot ise Benjamini & Hochberg metodudur.

Uygulamada sıklıkla kullanılan çoklu karşılaştırma yöntemleri, gruptaki gözlem sayılarının dengesizliğine karşı 1. ve 2. tip hataları ve güçleri bakımından 1000 tekrarlı bir simülasyon çalışmasıyla karşılaştırılmıştır. Bu simülasyon çalışmasında ortalamalar arasında fark olduğunda ve olmadığına grupta gözlem sayıları eşit olduğunda ve olmadığına 3 grup için (2,5,10) ve 5 grup için (2,5,10,30,50) olduğu kombinasyonlara varyans analizi uygulanmadan ve varyans analizi uygulandıktan sonra fark bulunan denemelere ayrı ayrı çoklu karşılaştırmalar uygulanmıştır. Çoklu karşılaştırmalar sonucunda $\alpha=0.05$ düzeyinde istatistik olarak fark bulunan karşılaştırma sayılarından yararlanılarak fwe ve pwe oranları hesaplanmıştır.

Yapılan çalışmaların sonucunda elde edilen bulgular, ANOVA ve çoklu karşılaştırma testlerinin tüm sonuçları, testlerin güçleri ve yaptıkları hata oranları ile birlikte ayrıntılı bir şekilde sunan tablolar halinde verilmiş ve

bu verileri görsel olarak daha ayrıntılı ifade etmek için her tabloya ait grafik çizelgesi düzenlenmiştir.

Bu deneysel araştırmanın sonuçları yaygın kullanılan pek çok post hoc testin ortalamalar arasında gerçekte farklılık olduğu durumlarda ve ANOVA sonrası kullanıldığında oldukça yüksek güçlere sahip olduklarını ancak ANOVA yapmaksızın kullanıldıklarında özellikle LSD ve DUNCAN testlerinin kabul edilemez derecede 1. tip hataya sahip oldukları göstermektedir.

Çoklu karşılaştırma sonuçlarını incelediğimizde ise, Tukey HSD testine ait simülasyon sonuçları, ANOVA'da olduğu gibi $n=2$ 'de % 5'in altında 1. tip hata oranı ile sonuçlanmıştır. Tukey testi ideal 1. tip hata oranını $n=5$ 'de yakalamıştır. Olası tüm ikili karşılaştırmaları test etmede genel bir teknik sunan SCHEFFE testi fazlasıyla koruyucu olduğu için çok fazla sayıda kontrast olmadıkça önerilmemektedir. SCHEFFE testinin güç yününden düşük, FWE'yi korumada ise % 5 seviyesinin de altında oldukça koruyucu değerlere sahip olduğu gözlemlenmektedir. Bonferroni testi ise koruyucu bir testtir. Bunun anlamı FWE tam olarak alfaya eşit değildir ve çoğu zaman alfadan küçüktür. Bonferroni testi FWE'yi sıkı bir şekilde kontrol ettiğinden fazlasıyla koruyucu bir testtir ve çoğu zaman anlamlı farklılıkları tespit edemez.

Simülasyon sonuçlarına göre yapılan çalışmanın koşullarına paralel olarak kullanılması hem ANOVA'nın belirttiği farklılıkları en iyi tespit etmesi hem de diğer güçlü testlere göre FWE'yi istenen seviyede kontrol etmesi bakımından en uygun testin REGW F testi olduğu saptanmıştır.

Anahtar kelimeler: dengeli, dengeli olmayan varyans analizi, deney öncesi ve deney sonrası bağımlı, bağımsız karşılaştırmalar, hatalı keşif oranı, çoklu karşılaştırma metotları, 1. tip ve 2. tip hata, simülasyon çalışması.

TANI TESTLERİNİ DEĞERLENDİRMEDE GEE VE LATENT CLASS MODELLERİ

R. Burgut¹, Y. Sertdemir¹, R. Tuncer², M.U.Sandıkçı², İlker Ünal¹

*Çukurova Üniversitesi, Tıp Fakültesi, Biyoistatistik ABD, Adana¹
Çukurova Üniversitesi, Tıp Fakültesi, "İç Hastalıkları ABD, Adana²*

Özet

Duyarlı ve özgün tanı testlerinin bulunmayışı "Helicobacter Pylori (HP)" enfeksiyonunun doğal gidişinin anlaşılmasını ve prevalansının belirlenmesini önlemektedir. Bu durum kesin tanının konulamaması durumunda ("Gold standard test in olmayışı") daha da karmaşık hale gelmektedir. Klinikte; HP tanısı koymada, yeni test standard test yerine kullanılabilir mi? Değilse standard test'e destekleyici test olarak kullanılabilir mi? sorusu ile karşılaşmaktayız. Birinci soru için, yeni bir tanı testinin geçerliliğini değerlendirmede değişik yaklaşımlar olduğunu bilmekteyiz. Bunlardan bazıları, hatalar içeren ("imperfect") standart'ın veya "Discrepant resolution" veya "Composite reference Standard" yöntemlerinin kullanımı gibi yaklaşımlardır.

Kesin tanı konulmasını sağlayan (Gold Standard) yöntemlerin bulunmadığı durumlarda, hastaların klinik karakterlerinin kovariate olarak dahil edildiği standard testlerinin yerini alabilecek yeni bir testin geçerliliğini değerlendirmede son yıllarda önerilen yöntem "**Latent Class Analysis**" yaklaşımıdır. Bir başka modelleme yaklaşımı da, yeni testin standart test'i destekleyiciliği değerini değerlendirmede hasta özelliklerinin kovaryet olarak dahil edildiği marjinal regresyon modelinde Generalized Estimation Equation (GEE) yaklaşımıdır.

Bu çalışmada, HP tanısı koymada doğrudan gayta örneklerini kullanan "Direct enzyme immunoassay (HpSA)" yeni bir testin geçerliliği önce Latent Class yaklaşımı ile değerlendirilecektir. Sonra da yaş, cinsiyet ve HP tedavisinin sonuçlarının kovariate olarak alındığı GEE yaklaşımı ile diğer testleri (histolojiye dayalı test, Rapid Urea test, PCR test, IgG ve IgM) desteklemede önemi değerlendirilecektir. Her iki modelleme yaklaşımında karşılaşılan sorunlar tartışılacaktır.

Anahtar Kelimeler: Tanı testleri, Latent Class Analizi, GEE.

DIAGNOSTİK TANI TESTLERİNİN GOLD STANDART TEST OLMADAN GİZLİ SINIF (LATENT CLASS) ANALİZİ İLE DEĞERLENDİRİLMESİ

Meriç Çolak¹, Vildan Sümbüloğlu²

Sağlık Kurumları İşletmeciliği Bölümü, Başkent Üniversitesi, ANKARA¹
Biyoistatistik Ana Bilim Dalı, Tıp Fakültesi, Kara Elmas Üniversitesi,
ZONGULDAK²

e-mail: meric@baskent.edu.tr

Özet

Bu çalışmada, ilgili alt grupların belirlenmesi amacıyla yönelik, çok değişkenli, bir istatistiksel yöntem olan gizli sınıf analiz yöntemi ile, gold standart testin olmadığı veya uygulamasının mümkün olmadığı ve referans testin kesin olmadığı durumlarda tanı testlerinin hassaslık parametrelerinin değerlendirilmesi incelenmiştir. Radyolojide meme kanseri tanı testi olarak kullanılan üç tanı testinin ve birleştirilmiş tanı testlerinin hassaslık parametreleri ve prevalans değeri gold standart test sonuçları olmaksızın, Gizli Sınıf (Latent Class) Analizi ile elde edilmiştir. Elde edilen sonuçlar ayrıca mevcut histopatoloji sonuçları ile de değerlendirilmiştir.

Gizli Sınıf Analizi (GSA), doğrudan gözlenemeyen (gizli) bir teorik kavramın, gizli değişkenin mükemmel belirleyicileri olmayan gözlenen değişkenler ile arasındaki kovaryasyonu açıklamada uygulanabilir bir yöntemdir. Amaç, gözlenen yanıtlar arasındaki ilişkiyi açıklayacak birbirinden tamamen ayrık ve bağımsız sınıfları tanımlamaktır. Kategorik verilere uygulanan küme analizi olarak düşünülebilir. GSA ayrıca faktör analizi gibi veri indirgeme amacı ile de kullanılabilir. Bir diğer uygulaması ise gold standart tanı testinin olmaması durumunda diagnostik tanı testlerinin değerlendirilmesidir. Gizli sınıf modeli bir veya daha fazla kategorik, gizli, gözlenemeyen değişkenlerin ölçülmesi için oluşturulmuş iki veya daha fazla incelenilen olayın belirleyicileri olarak kabul edilen kategorik gözlenen değişkenler arasındaki ilişkiyi tanımlar. Gazi Üniversitesi Hastanesi Radyoloji Ana Bilim Dalında gerçekleştirilen çalışmanın verileri kullanılmıştır. Radyolojide meme lezyonlarında rutin tanı testi olarak kullanılan

mammography ve gray scale sonography kombinasyonu (MG-US), Power Doppler (PD) ve Contrast-Enhanced Power Doppler Sonography (CE-PD) testleri ile gerçekleştirilen meme lezyonu değerlendirmeleri kullanılarak, gizli sınıf analizi ile tanı testlerinin hassaslık parametreleri ve prevalans değeri en çok olabilirlik kestirimlerini veren E-M algoritması ile elde edilmiştir. Malin ve benin lezyonlar tek bir gizli değişkenin iki ayrı düzeyi ele alınarak iki gizli sınıflı model oluşturulmuş ve gizli sınıflar tanımlanmıştır. Her bir teste ve birleştirilmiş tanı testlerine ait duyarlılık ve seçicilik parametreleri ile prevalans değeri standart hata değerleri ile birlikte elde edilmiştir. Daha sonra, tanı testlerinden ve birleştirilmiş tanı testlerinden elde edilen olası yanıt düzeneklerine göre kestirilen sınıf üyelik olasılıkları ve sınıf atamaları verilerek, optimum sınıfa gözlemlerin ataması gerçekleştirilmiştir. Gizli sınıf analizi ile elde edilen tanı testlerine ait hassaslık parametre değerleri histopatoloji sonuçlarına göre elde edilen değerlerle karşılaştırılmış ve doğru sınıflandırma yüzdesi elde edilmiştir. Gizli sınıf analizi ile oluşturulan model parametrelerinin istatistiksel önem kontrolü gerçekleştirilmiş, analizlerde Latent GOLD (3.0) gizli sınıf analizi paket programı ile SPSS (11.5) paket programı kullanılmıştır.

Anahtar Kelimeler: Gizli Sınıf Analizi, Diagnostik Tanı Testleri, Duyarlılık, Seçicilik, Prevelans, Latent GOLD

1. Giriş

Sınıflama problemi tüm bilim dallarında ortak bir problemdir. Sürekli değişkenlerin ve özellikle normal dağılım gösteren değişkenlerin var olduğu durumlarla ilgili yöntemlerle daha çok ilgilenilirken, sürekli olmayan veriler ile kesikli ve sürekli verilerin bulunduğu karma veri setlerinin sınıflama problemleri ile pratikte sıkça karşılaşılmasına rağmen daha az ilgilenilmiştir. Çalışmanın amacı, ilgili alt grupların belirlenmesi amacıyla yönelik, çok değişkenli bir istatistiksel yöntem olan gizli sınıf analizi'nin (GSA) tıpta "gold standart" testin olmadığı veya uygulanmasının mümkün olmadığı ve referans testin kesin olmadığı durumlarda tanı testlerinin hassaslık parametrelerinin değerlendirilmesindeki kullanımının ortaya konulmasıdır. Çalışmada GSA yaklaşımı Latent GOLD paket programı ile incelenmiştir. Gold standart olarak kabul edilmiş bir standart testin olmaması durumunda tanı testlerinin hassaslık parametrelerinin ve hata oranlarının tahminlenmesi oldukça basit bir süreçtir. Fakat gold standart testin olmadığı veya uygulanmasının mümkün olmadığı, pratik olmadığı durumlarda, yeni bir tanı testinin performansı hata payı olan bir başka referans teste göre değerlendirilmektedir. Referans testin hata oranlarının dikkate alınmaması, yeni tanı testinin parametrelerinin hatalı olarak tahminlenmesine neden olmaktadır. GSA, gold standart testin olmadığı veya uygulanmasının olanaklı

olmadığı durumlarda alternatif bir istatistiksel yaklaşımdır. Gözlenemeyen hastalık durumu gizli sınıf olarak ele alınır ve referans testlerden elde edilen gözlenen değerler ile istatistiksel model oluşturulur. Her bir gözlem üzerinde en az üç tanı testinin sonucunun bulunması gerekir. En çok olabilirlik tahminleme yöntemi kullanılarak hastalık prevalans değeri ile duyarlılık, seçicilik parametre değerleri standart hata değerleri ile her bir test için elde edilir. Ayrıca GSA diagnostik tanılardaki alt kategorilerin veya hastalıkların alt tiplerinin belirlenebilmesi amacı ile pek çok sağlık uygulamalarında kullanılabilir bir yöntemdir.

Gizli değişkenler doğrudan gözlenemeyen fakat diğer gözlenen değişkenler tarafından açıklanabilen özellikler olarak tanımlanabilir. Tüm gizli değişkenlerin bir kavrama karşılık gelmesinden dolayı gizli değişkenler varsayımsal değişkenlerdir. Gizli değişkenler için varolan önsel dağılımlar, koşullu olasılıklar ve gizli değişkenlerin var olduğu durumlardaki gözlenen değişkenlere ilişkin koşullu olasılıklar ile birlikte ele alındığında, gizli değişkenler için sonsal olasılık dağılımları belirlenmektedir. Gizli ve gözlenen değişkenin metrik veya kategorik olmasına göre 4 tip gizli değişken modeli ortaya konulmuştur. Bunlar; sürekli ölçümlerden elde edilen sürekli gizli değişkenlerin tanımlanmasına yönelik faktör analizi, kesikli değişkenlerden sürekli gizli değişkenlerin tanımlanmasına yönelik gizli içerik modelleri veya kategorik faktör analizi, sürekli gözlenen değişkenlerden kesikli gizli değişkenlerin elde edilmesine olanak tanıyan gizli profil analizi ve kesikli gözlenen değişkenlerden kesikli gizli değişkenlerin tanımlandığı gizli sınıf analizidir. Gizli sınıf modelleri, geleneksel yaklaşımlarda bulunan ve çoğu zaman ihmal edilen; doğrusal ilişki, normal dağılım ve homojenlik varsayımları ile sınırlanmamıştır. Bu nedenle, varsayımların sağlanamamasından kaynaklanan hatalara daha az bağımlıdır. Ayrıca gizli değişken modelleri, karma ölçekli değişkenlerin bulunduğu analizlerde de uygulanabilmektedir.

1.1. Gizli Sınıf Analizi

GSA, iki veya daha fazla niteliksel değişken arasındaki ilişkiyi, değişkenlerin çapraz sınıflaması ile açıklayarak çok boyutlu, niteliksel gizli değişkenleri tanımlayan, ilgili alt grupların belirlenmesi amacıyla yönelik, çok boyutlu boyut indirgeme amaçlı bir istatistiksel yöntemdir. GSA'nın amacı, gözlenen değişkenlerden, ölçümlerden gizli değişken olarak adlandırılan tek bir ölçüt elde etmektir. GSA, nicel veriler için oluşturulmuş faktör analizinin nitel veriler için düzenlenmiş biçimidir. Ayrıca, GSA, gözlenen değişkenler açısından gözlemlenen yapılar arasında birbiri ile çok benzer durumların bir arada sınıflandırılması amacı ile de kullanılabilir. Bu yönü ile GSA, kategorik veriler için küme analizi olarak da düşünülebilir. GSA aynı zamanda her bir gizli sınıf içerisindeki koşullu olasılıkların incelenmesi ile gizli yapıların karakterize edilmesinde de kolaylık sağlar. GSA'de model parametreleri,

koşullu yanıt olasılıklarının en çok olabilirlik kestirimleridir. Koşullu olasılıklar ise, her bir gizli sınıf için o sınıfın bir üyesinin açıklayıcı değişkene öngörülen değeri vermesi olasılığıdır. Her bir sınıf içerisinde toplamları biri veren kestirilen üyelik oranları da GSA model parametreleridir. En çok olabilirlik kestirimi, adımsal olarak en olası gözlenen sonuçları verecek model parametrelerini hesaplar. Ayrıca en çok olabilirlik kestirim yöntemi ile üyelerin gizli sınıflara ataması yapılmaktadır. GSA' de temel varsayım, lokal bağımsızlıktır. Sınıf üyelikleri üzerinde koşullu olarak madde yanıtlarının istatistiksel olarak bağımsız olmaları anlamına karşılık gelir. Bu varsayım ile, gizli sınıf içerisindeki bireylerin maddelere verecekleri yanıtların beklenen değerlerinin eşdeğer olacağı kabul edilmiştir.

Dört değişkenli gizli sınıf modeli şu şekilde tanımlanabilir; A, B, C, D gözlenen değişkenler olup düzeyleri sırası ile I, J, K, L ile verilmektedir. ($i=1,2,\dots,I$; $j=1,2,\dots,J$; $k=1,2,\dots,K$; $l=1,2,\dots,L$) ve A, B, C, D değişkenleri arasındaki ilişkiyi açıklayan gizli değişken X olup, T düzeyinde verilmektedir. Temel gizli sınıf modeli lokal bağımsızlık varsayımına göre gizli sınıf olasılıklarının ve şartlı olasılıkların çarpımı olarak verilebilir (Hagenaars, A.Jacques., McCutcheon, L. Allan, 2002).

$$\pi_{ijkt}^{ABCDX} = \pi_t^X \pi_{it}^{\bar{A}X} \pi_{jt}^{\bar{B}X} \pi_{kt}^{\bar{C}X} \pi_{lt}^{\bar{D}X} \quad \sum_{t=1}^T \pi_t^X = 1$$

π_{ijkt}^{ABCDX} , herhangi bir gözlemin A değişkeninin i düzeyinde, B değişkeninin j düzeyinde, C değişkeninin k düzeyinde ve D değişkeninin l düzeyinde ve X gizli değişkeninin t düzeyinde birlikte gözlenme olasılığı, π_t^X , herhangi bir gözlemin X gizli değişkeninin t düzeyinde olma olasılığı. $\pi_{it}^{\bar{A}X}$, herhangi bir gözlemin X gizli değişkeninin t düzeyinde olduğu bir durumda, A değişkeninin i düzeyinde yer alma şartlı olasılığını tanımlar. Benzer tanımlamalar aşağıda verilen kısıtlamalar golda $\pi_{jt}^{\bar{B}X}$, $\pi_{jt}^{\bar{C}X}$ ve $\pi_{jt}^{\bar{D}X}$ içinde geçerlidir;

$$\sum_{i=1}^I \pi_{it}^{\bar{A}X} = 1, \quad \sum_{j=1}^J \pi_{jt}^{\bar{B}X} = 1, \quad \sum_{k=1}^K \pi_{kt}^{\bar{C}X} = 1, \quad \sum_{l=1}^L \pi_{lt}^{\bar{D}X} = 1$$

$\pi_t^X, \pi_{it}^{\bar{A}X}, \pi_{jt}^{\bar{B}X}, \pi_{kt}^{\bar{C}X}, \pi_{lt}^{\bar{D}X}$ olasılık değerleri en büyük olabilirlik tahminleme ile tahminlenebilir. Daha sonra bu tahminler kullanılarak, tahmini sıklık değeri F_{ijkl}^{ABCD} aşağıda verilen eşitlik ile hesaplanabilir;

$$F_{ijk}^{ABCDX} = \sum_t (F_{ijkl}^{ABCDX}) = \sum_t (N_x \pi_{ijkl}^{ABCDX}) = \sum_t (N_x \pi_t^X \pi_{it}^{\bar{A}X} \pi_{jt}^{\bar{B}X} \pi_{kt}^{\bar{C}X} \pi_{lt}^{\bar{D}X})$$

ve daha sonra, tahmini sıklıklar gözlenen sıklıklarla loglineer modellerde olduğu gibi araştırılabilir. Pratikte model uyumunu test eden pek çok yöntem bulunmaktadır. Bunlardan bazıları Chi-Square testi, enformasyon kriteri ve model uyum indisidir.

1.2. Tanı Testlerinin Performanslarının Gold Standart Test Olmadığı Durumda Değerlendirilmesi

Tanı testleri ve klinisyenler tarafından vakaların hastalık durumlarına göre sınıflandırılmasında klinik bilgi kullanılır. En basit durumda, sınıflandırma hastalık var ve hastalık yok şeklinde iki kategorilidir. Verilen bir tanı testinin performansını değerlendirmek için, gerçek hastalık durumuna koşullu olarak yöntemin hassaslığı ile ilgili olasılıksal kestirimlerinin hesaplanması gerekmektedir. Tanı testlerinin hassaslık parametreleri duyarlılık ve seçicilik parametreleridir. Bu parametrelerin tamamlayıcıları yanlış negatif ve yanlış pozitif oranları ise tanı testlerine ait yanlış sınıflandırma, hata parametreleridir. Ayrıca tanı testlerinin değerlendirilmesinde klinisyenler tarafından tahmini pozitif ve tahmini negatif değerleri de kullanılmaktadır (Rynard ve Walter, 1997). Tıbbi araştırmalarda kullanılan tanı testleri sağlık bakım hizmetlerinde de önemli rol oynamaktadır. Bu nedenle yeni tanı testlerinin gelişimi hızla artmaktadır. Yeni tanı testlerinin performans değerlendirmeleri en ideal durumda "gold standart" olarak kabul görmüş bir test ile karşılaştırılarak yapılmaktadır. Fakat çoğu zaman "gold standart" test bulunmamakta ve mükemmel olmayan başka bir referans test kullanılmaktadır. "Gold standart" testin bulunduğu durumda, yeni tanı testinin değerlendirmesine yönelik pek çok istatistiksel yöntem bulunmasına karşın, bu yöntemler "gold standart" test bulunmadığı veya uygulanmasının mümkün olmadığı durumlarda yetersiz kalmaktadır (Alanzo ve Pepe, 1999). Gözlemciler arası uyumsuzluk ve görelî hata, birden fazla gözlemci (görüşmeci, uygulayıcı, tanı koyucu veya tanı yöntemi) olması durumunda ortak bir konudur. Eğer kesin bir standart test varsa, hatalar tanımlanabilir ve kolaylıkla kestirilebilir.

Tıbbi arařtırmaların bazı alandaki uygulamalarında hastalığın varlığı ile ilgili pathognomic belirleyiciler olarak adlandırılan semptomlar ve belirtiler bulunmaktadır. Bu belirleyicilerin varlığı ile yokluğu hastalığın durumunu belirlemektedir. Bazı durumlarda ise hastalık tanısı klinik belirleyicilerin kombinasyonu veya kan testlerinin kombinasyonu ile konulmaktadır. Fakat bu tanı yöntemleri tıpta oldukça az rastlanan basit tanı yöntemleridir. En güvenilir tanı yöntemi ise “gold standart” olarak adlandırılan duyarlılığı ve seçiciliği % 100 olan tanı testleridir. Bununla birlikte tıbbi arařtırmaların birçok alandaki uygulamalarında, “gold standart” tanı testi bulunmamakta veya pahalı olmaları, tehlikeli olmaları veya etik olmamaları nedeni ile uygulanması mümkün olmamaktadır. Bu durumda, tanı testlerinin veya algoritmalarının performansının deęerlendirmesi problem yaratmakta ve řüpheli sonuçlar vermektedir. Mükemmel olmayan referans test problemine Walter ve Irwig (1988) tarafından ortaya konulan istatistiksel çözüm, gizli sınıf analizidir. Gizli sınıf modelleri, karşılařtırmada kullanacak tam anlamı ile hassas “gold standart” testin olmadığı durumda, tanı testlerinin performansının deęerlendirmesinde kullanılabilir. Gizli sınıf modeli ile hastalığın varlığı gizli deęişken olarak tanımlanarak, “gold standart” testin bulunmadığı veya uygulanmadığı durumlarda tanı testlerinin deęerlendirmesi yapılabilmekte ve duyarlılık, seçicilik, tahmini pozitif deęer, tahmini negatif deęer ile tanımlanan hassaslık parametreleri ile yanlış pozitif oranı ve yanlış negatif oranı ile tanımlanan hata parametreleri ve standart hataları verilebilmektedir. GSA'nin temel varsayımı, gözlenemeyen bir hastalık durumunun olduğu ve gözlenen tanı testleri ile bu hastalık durumunun istatistiksel model ile ortaya konulmasıdır. Gizli sınıf modellerinin diagnostik uyum çalışmalarındaki önemi günden güne hızla artmaktadır. GSA, tarama testlerinin her bir kişide uygulandığı ve negatif test sonucu elde edilen kişilere takip tanı testlerinin uygulanmasının mümkün olmadığı durumlarda kullanışlıdır.

Gizli sınıf modelleri tanı koyucuların, gözlemcilerin ve tanı testlerinin hassaslıklarını ve deęişen hastalık kategorilerine göre prevalans deęerini gözlenemeyen parametre olarak elde etmeyi amaçlar. Gözlenen veriye dayanarak bu parametreler kestirilir. Kısıtlanmamış GSA ile parametrelerin kestirimi için her vakaya ait en az üç gözlemin bulunması gerekmektedir. Her vaka sadece bir gizli sınıfa atanabilmektedir. Olabilirlik, yedi parametre cinsinden yazılabilmektedir (üç tane duyarlılık deęeri, üç tane seçicilik deęeri ile hastalığa ait prevalans deęeri). Toplam örneklem büyüklüğü sabitlendiğinde yedi serbestlik derecesi bulunmaktadır, dolayısıyla model tanımlıdır. Bu problem, binomial dağılımların karmasının karma dağılımının elde edilmesi ile eşdeğerdir (Rynard ve Walter, 1997). Bu modeller gerçek hastalık durumuna koşullu olarak testler arasında genellikle bağımsız hataları varsayar. En çok olabilirlik kestirim yöntemi ile hastalık prevalans deęeri ve her bir tanı testine ait hassaslık parametreleri GSA ile verilmektedir (Alonzo ve Pepe, 1999).

Hastalık, gerçek fakat direkt olarak gözlenemeyen durum olarak tanımlanmaktadır. Direkt olarak gözlenemeyen hastalık durumu gizli sınıf olarak ele alınmaktadır. Özellikle psikiyatrik araştırmalarda bu durum oldukça geçerlidir. Günlük yaşantımızda pek çok durum gözlenemeyen gizli sınıfa örnek verilebilir. Kişileri davranışlarına göre dürüst - dürüst olmayan, duygusal durumlarına göre mutlu olanlar - olmayanlar, zihinsel özelliklerine göre zeki olanlar - olmayanlar olarak sınıflamak gibi. Kişilere ait bu özellikleri direkt olarak ölçmek mümkün değildir. "Gold standart" testlerin yokluğu yıllardır psikiyatrik araştırmalarda bir problem olmuştur. Tanı konulabilen fakat konulan bu tanı laboratuvar testleri ile veya patolojik testler ile doğrulanamamaktadır. Gizli sınıflar direkt olarak gözlenemediği için hassaslık ve nedensellik olasılıkları da direkt olarak elde edilememektedir. Sadece davranışlar arasındaki ilişkiyi ortaya koyan olasılıklar elde edilecektir. Gizli karakteristikleri ortaya koyan bazı belirleyiciler ile gözlenemeyen gizli değişkeni tanımlamak mümkün olabilmektedir. GSA'nin amacı, gözlenen olasılıklar ile gözlenemeyen olasılıkların kestirilmesidir. Gizli sınıf modelleri hastalığın gerçek belirleyicileri değildir, sadece tanı testlerinin işlemsel karakteristiklerinin anlaşılmasında kolaylık sağlamaktadır. Evrendeki vakaların, gözlemlerin, deneklerin iki veya daha fazla gizli sınıfa ait olmaları ifadesinde, bu vakalara ait sınıf üyeliklerinin doğrudan elde edilemediği anlamı yer almaktadır. Gizli sınıf bir alt küme veya "genotip" olarak düşünülebilir. Bir vakanın öngörülen düzeye atanması olasılığının, vakanın içinde yer aldığı gizli sınıfa bağlı olduğu varsayılır. Örneğin, evren iki gizli sınıftan oluşabilir; normal sağlıklı bireylerin oluşturduğu hastalık-negatif ve vakaların oluşturduğu hastalık-pozitif grubu. Bir vakaya, bir gözlemci tarafından veya bir tanı testi sonucunda pozitif tanı konulma olasılığı vakanın hangi gizli sınıfa ait olmasına bağlı olacaktır. GSA ile her bir gizli sınıfa yer alan vakaların oranı ve her bir gizli sınıfta yer alan vakaların gözlemciler veya tanı testleri ile öngörülen düzeye ait olma olasılıkları belirlenmektedir.

Tanı testlerinin iki kategorili, binary, olduğu durum incelenecektir. Fakat sonuçlar ikiden fazla kategori durumuna genellenebilir. N tane vaka k tane tanı testi ile değerlendirilmektedir. Her bir değerlendirme iki kategorilidir. Tüm gözlemler değerlendirildiğinde, her bir test tarafından gerçekleştirilen pozitif sınıflamaların oranı, p_j ($j=1, \dots, k$) elde edilir. Temel gizli sınıf modeli gereği bu oranlar iki parametre seti ile belirlenir; vakaların ait olduğu c tane birbirinden ayrık gizli sınıf prevalans değerleri ve her bir gizli sınıfta vakalara pozitif veya negatif tanı konulması koşullu sınıflandırma olasılık değerleri gizli sınıfların sayısı genellikle 2 ile 5 arasındadır. Farklı gizli sınıf sayısına sahip modeller karşılaştırılır ve gözlenen veriye en iyi uyan model seçilir. Hastalığın gerçek prevalans değeri π ile gösterilmektedir. Gizli sınıf prevalans değerleri $\pi_1, \pi_2, \pi_3, \dots, \pi_c$ ile gösterilir. π_s ($s=1, 2, \dots, c$) rasgele örnekleme alınan bir vakanın s'inci gizli sınıfa ait olma olasılığı olarak tanımlanmaktadır. Koşullu sınıflandırma olasılıkları $\pi_{0|1}, \pi_{0|2}, \dots, \pi_{0|c}$ ve

$\pi_{1|1}, \pi_{1|2}, \dots, \pi_{1|c}$ ile gösterilmektedir. $\pi_{0|s}$ ve $\pi_{1|s}$ verilen bir gizli sınıf s'nin üyesinin negatif veya pozitif sınıflandırılma olasılık değeridir. $\pi_{1|s_j}$ ($s= 1, 2, \dots, c$; $j=1, 2, \dots, k$), vakanın gizli sınıf s'ye ait olduğu verildiğinde, j'inci tanı testi ile pozitif sınıflandırılma koşullu olasılığı olarak tanımlanmaktadır. k tane tanı testi ile toplam 2^k tane farklı pozitif ve negatif sınıflandırma mümkündür. U_i vektörü i'inci olası yanıt düzenini tanımlarsa vektörün j'inci elemanı U_{ij} , j'inci tanı testinin sınıflandırmasına karşılık gelmektedir. $\pi_{i,s}$, s'inci gizli sınıfa ait ve u_i sınıflandırma düzenini alan bir gözlemin birleşik olasılığı

$$\pi_{i,s} = \pi_s \prod_{j=1}^k \pi_{1|s_j}^{u_{ij}} (1 - \pi_{1|s_j})^{1-u_{ij}}$$

ile verilmektedir. u_{ij} ve $(1-u_{ij})$ üstleri birleşik olasılık ifadesinde, j'inci sınıflamanın pozitif veya negatif olmasına göre $\pi_{1|s_j}$ veya $(1-\pi_{1|s_j})$ ifadelerinin yer almasına neden olur (Uebersax ve Grove, 1990).

Walter ve Irwig (1988) tarafından verilen tanımlama şu şekildedir: Gerçek hastalık prevalansı θ ile j'inci testin yanlış negatif oranı (gerçekte hasta olan vakaların j'inci test ile sağlam olarak sınıflandırılma olasılığı) β_j ile, yanlış pozitif oranı (j'inci testin gerçekte sağlam olan vakaları hasta olarak sınıflandırma olasılığı) α_j ile verilmektedir. j'inci testin duyarlılık ve seçicilik parametreleri $(1-\beta_j)$ ve $(1-\alpha_j)$ ile verilmektedir. Her vakanın k tane binary test ile sınıflandırılması sonucu, 2^k tane olası sınıflandırma kombinasyonundan oluşan yanıt düzeni bulunmaktadır. Bu olası sınıflandırma kombinasyonu yanıt kategorileri olarak adlandırılmaktadır. $k=3$ olma durumunda, $2^3=8$ olası yanıt düzeni bulunmaktadır. i'inci testin sınıflandırması rasgele değişken X ile (eğer sınıflandırma negatif ise, $X=0$, eğer sınıflandırma pozitif ise, $X=1$), gerçek hastalık durumu Y rasgele değişkeni ile (eğer hastalık yok ise, $Y=0$, eğer hastalık var ise $Y=1$) verilmektedir.

Kestirim için uygun istatistiksel serbestlik derecesi (2^k-1) , yanıt kategorilerinin sayısı tarafından belirlenir. $k=3$ durumunda, vakaların gerçek hastalık durumuna koşullu olarak, belirli bir yanıt kategorisinde yer alma olasılığı hesaplanabilir. $k>3$ durumuna ise kolaylıkla genellenebilir. Üç test tarafından da vakanın negatif olarak sınıflandırılma koşulsuz olasılığı, iki koşullu olasılığın toplamı ile elde edilecektir;

$$\begin{aligned} P(X_1 = 0, X_2 = 0, X_3 = 0) &= P(Y = 1)P(X_1 = 0, X_2 = 0, X_3 = 0 | Y = 1) \\ &+ P(Y = 0)P(X_1 = 0, X_2 = 0, X_3 = 0 | Y = 0) \\ &= \theta\beta_1\beta_2\beta_3 + (1-\theta)(1-\alpha_1)(1-\alpha_2)(1-\alpha_3) \end{aligned}$$

benzer şekilde 8 tane yanıt düzenine ait koşulsuz olasılık değeri aşağıdaki formülle bulunabilir.

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \theta \prod_{i=1}^3 \beta_i^{1-x_i} (1-\beta_i)^{x_i} + (1-\theta) \prod_{i=1}^3 \alpha_i^{x_i} (1-\alpha_i)^{1-x_i}$$

Genellikle k tane tanı testi ve n tane gözlem için; $(n(x))$ 'in, $\sum_x n(x) = n$ olduğu $x=(x_1, x_2, \dots, x_k)$ vektörü şeklinde tanımlandığı, tüm veriye ait olabilirlik fonksiyonu birleşik olasılık fonksiyonuna orantılıdır (Rynard ve Walter, 1997).

$$L\alpha \prod_x \left[\theta \prod_{i=1}^k \beta_i^{1-x_i} (1-\beta_i)^{x_i} + (1-\theta) \prod_{i=1}^k \alpha_i^{x_i} (1-\alpha_i)^{1-x_i} \right]^{n(x)}$$

Parametrelerin kestiriminde, $\ln(L)$ fonksiyonunun maksimizasyonu işlemler açısından daha uygundur. $\ln(L)$ fonksiyonunun maksimizasyonu ile GSM' nin parametrelerine ait en çok olabilirlik kestirim değerleri elde edilecektir. Log olabilirlik fonksiyonu

$$\ln(L) = \sum_{s=1}^S \sum_x n_s(x) \ln \left[\theta \prod_{i=1}^k \beta_i^{1-x_i} (1-\beta_i)^{x_i} + (1-\theta) \prod_{i=1}^k \alpha_i^{x_i} (1-\alpha_i)^{1-x_i} \right]$$

ile verilmektedir.

1.3. Kestirim, Karşılaştırma ve Model Tanımlılığı

Her bir tanı testi sınıflandırmasını alan vakaların gözlenen oranlarının elde edildiği durumda rasgele örnekleme alınan bir vakanın s' inci gizli sınıfa ait olma olasılığı olarak tanımlan π_s ($s=1, 2, \dots, c$) prevalans değeri ile, vakanın gizli sınıf s' ye ait olduğu verildiğinde j' inci tanı testi ile

pozitif sınıflandırılma koşullu olasılığı olarak tanımlanan $\pi_{1|s_j}$ ($s=1, 2, \dots, c$; $j=1, 2, \dots, k$) koşullu olasılık parametrelerinin kestiriminde E-M algoritması kullanılacaktır. Bu kestirim işlemi GSA bilgisayar programları ile gerçekleştirilmektedir.

Parametre sayısı fazla olduğunda E-M algoritması yavaşlamaktadır. Bu nedenle, böyle durumlarda E-M algoritmasının Newton-Raphson algoritması ile birleştirilmesi önerilmektedir. Algoritmaların birleştirilmesi durumunda E-M algoritması ile kısmen daha az sayıda adımlar gerçekleştirilir. Daha sonra, elde edilen sonuç parametreleri Newton-Raphson algoritmasında başlangıç değerleri olarak kullanılarak en çok olabilirlik parametre kestirimlerinin daha hızlı bir şekilde elde edilmesi sağlanır (Uebersax ve Grove, 1990). Oluşturulan c tane gizli sınıf modelinin uyumu, gözlenen ve beklenen değerlerin Pearson ki-kare veya olabilirlik ki-kare istatistik değeri ile karşılaştırılması ile verilmektedir. Olabilirlik ki-kare testinin avantajı olarak, aynı veri seti ile oluşturulacak alternatif modellerin karşılaştırılması verilebilir. İki model arasındaki farklılığın istatistiksel önemliliği, iki modele ait olabilirlik ki-kare istatistik değerlerinin çıkarılması ile elde edilebilir. Fakat bu durum, sadece gizli sınıf sayılarında farklılık gösteren iki model için, bir modelin parametrelerinin diğer modelin parametrelerinin bir alt kümesi olması ile tanımlanan iki modelin iç içe geçmiş olması durumunu gerektirmektedir. Model uyumundaki bir diğer önemli konu örneklem büyüklüğüdür. Büyük örneklemli çalışmalarda, gözlenen ve beklenen değerler arasındaki ufak farklılıklar istatistiksel olarak önemli bulunan ki-kare değeri verebilmektedir. Bu nedenle model uyumu, örneklem büyüklüğüne daha az bağımlı olan model uyum indisleri ile değerlendirilmelidir (Uebersax ve Grove, 1990).

Lazarsfeld ve Henry (1968), Goodman (1974) ve Haberman (1979) tarafından model tanımlılığı konusu ayrıntıları ile incelenmiştir. Sonsuz sayıda çözüme neden olacak modelin tanımsız durumu, modeldeki bilinmeyen parametre sayısından daha az eşitliklerin olması durumunda söz konusu olacaktır. Gizli sınıf modelinin tanımlı olabilmesi için kestirilecek parametre sayısının gözlenen veriye ait serbestlik derecesinden küçük olması gerekir. Kestirim için gerekli olan $(c-1)$ tane prevalans değeri bulunmaktadır. Fakat, $c \times k$ tane koşullu sınıflandırma olasılık değerleri (her bir tanı testi kombinasyonu ve gizli sınıf için bir tane olmak üzere) veya toplam $c(k+1)-1$ tane parametre bulunmaktadır. Tek bir çözüm için $c(k+1) \leq 2^k$ durumu sağlanmalıdır. Model tanımlılığı için bu eşitlik sadece gerekli bir durumdur. İki gizli sınıf durumunda modelin tanımlı olabilmesi için en az her bir vakanın üç gözlemci veya tanı testi ile değerlendirilmesi gerekmektedir. Üç gizli sınıflı modelde formül dört değerlendirme öngörse de, en az beş değerlendirme gerekmektedir. Dört sınıflı modelde beş değerlendirme gerekmektedir. Model tanımlılığı için genel bir yöntem olarak, model

parametrelerine karşı koşulsuz yanıt kategori olasılıklarının türevlerinden oluşan Hessian matrisinin rank değerinin hesaplanması veya log olabilirlik fonksiyonunun model parametrelerine göre ikinci dereceden türevlerinin alınması ile oluşan matrisin rank değerinin hesaplanması verilmişti. Eğer bu işlemler sonucunda model tanımsız bulunursa, kestirilecek parametre sayısı azaltılmalıdır. Parametrelerin azaltılması ise, gizli sınıf sayısının azaltılması veya bir ya da daha fazla parametreye kısıtlar getirilerek yapılabilmektedir.

1.4. Tanı Testlerinin Hassaslık Parametrelerinin Kestirimi

Uyum verilerinden tanı testlerinin hassaslık parametrelerine ait çıkarımların elde edilebilmesi, gizli sınıfların yapısını ilgilendiren varsayımlara dayanır. Genellikle, gizli sınıfların hastalığa ait semptomların farklı düzeylerine veya hastalığın varlığı ile yokluğuna karşılık geldiği kabul edilerek hassaslık parametreleri ile ilgili çıkarımlar yapılabilir. Fakat bazı durumlarda, gizli sınıflar hastalık durumunu değil de, hastalığa ait faktörleri yansıtabilir. Böyle durumlarda, gizli sınıf modeli ile elde edilen model parametreleri sadece hassaslık parametrelerine ait üst sınır değerlerinin elde edilmesinde kullanılabilir. Eğer gizli sınıflar pozitif veya negatif hastalık durumu ile tanımlanırsa direkt olarak duyarlılık, seçicilik ve diğer ilgili parametreler kestirilebilir.

Duyarlılık parametresi, pozitif gizli sınıf üzerinden toplam alınarak,

$$(\sum \pi_s \pi_{1|s_i}) / \sum \pi_s$$

Seçicilik parametresi negatif sınıf üzerinden toplam alınarak,

$$[\sum \pi_s (1 - \pi_{1|s_i})] / \sum \pi_s$$

hesaplanır. Ayrıca, tanı testi ile pozitif sınıflandırılan bir vakanın gerçekte de pozitif olma olasılığı (pozitif tahmini değer), pozitif gizli sınıf üzerinde $(\pi_s \pi_{1|s_i})$ değerinin tüm sınıflar üzerinden alınan toplam değere bölünmesi ile elde edilebilir. Benzer şekilde, tanı testi ile negatif sınıflandırılan bir vakanın gerçekte de negatif olma olasılığı (negatif tahmini değer), $\pi_{1|s_i}$ yerine $1 - \pi_{1|s_i}$ değerinin kullanılması ve negatif sınıflar üzerinden toplam alınması ile elde edilebilir.

GSA' nin uyum verilerinin analizindeki en önemli özelliği, çoklu kararların entegrasyonunda ve yorumlanmasında açık bir Bayes yaklaşım sağlamasıdır. k tane tanı testi sonucu elde edildiğinde, herhangi bir gözlemin sınıflandırma düzeneğine göre gizli sınıf s' ye ait olma olasılığı $\pi_{i,s}/\pi_i$ ile verilmektedir.

Tüm verilen bu kavramlar, N tane vakanın rasgele seçilen k tane gözlemci tarafından değerlendirildiği değişen gözlemcili panel çalışmalarına ve her bir vakanın farklı sayıda sınıflandırıldığı çalışmalara da uygulanabilir. Bu uygulamalar daha çok sınıflamaların boşa çıktığı veya sadece bazı vakaların birden fazla sınıflandırılmasının var olduğu durumlarda gözükmektedir (Uebersax ve Grove, 1990).

2. Gereç ve Yöntem

Aralık 1999 ile Haziran 2001 yılları arasında Gazi Üniversitesi Hastanesi Radyoloji Ana Bilim Dalında Doç. Dr. Ayşegül Özdemir ve arkadaşları tarafından yürütülen çalışmanın verileri izin alınarak kullanılmıştır. Meme lezyonlarında rutin tanı testi olarak kullanılan mammography ve gray scale sonography kombinasyonu (MG-US), Power Doppler (PD) ve ilk iki teste ek olarak yapılan toplardamar içerisinden kontrast madde verilerek ses dalgalarının titreşimini inceleyen Contrast-Enhanced Power Doppler Sonography (CE-PD) testleri ile gerçekleştirilen meme lezyonu değerlendirmeleri kullanılmıştır. 69 meme lezyonunun ilk olarak MG-US değerlendirmeleri yapılmış, daha sonra power doppler görüntülemeleri toplardamar içerisinden kontrast madde "Levovist" verilmeden önce (PD) ve verildikten sonra (CE-PD) elde edilmiştir. 69 lezyondan 47'sinin sonuçları GSA ile değerlendirilmiştir. 22 lezyonun MG-US sonucu tam olarak belirlenemediği için analize dahil edilmemiştir. MG-US ile PD değerlendirmeleri farklı radyologlar tarafından gerçekleştirilmiştir. Daha sonra 47 lezyona ait histopatoloji sonuçları elde edilmiştir. Her üç tanı testine ait değerlendirmeler, hastalık (+) tanı "1" ile kodlanıp kötü huylu (malin) lezyonları, hastalık (-) tanı "0" ile kodlanıp iyi huylu (benin) lezyonları tanımlamada kullanılmıştır. Veriler SPSS (11.5) paket programına girilerek, her bir olası yanıt düzenine ait frekans değerleri elde edilmiş ve GSA' de ham veri olarak kullanılmıştır. Analizler, Windows uyumlu Latent GOLD (3.0) Gizli Sınıf Analiz yazılım programının, Gizli Sınıf Küme Analizi Modülü ile gerçekleştirilmiştir. Lezyonların değerlendirmesi iyi huylu (benin) ve kötü huylu (malin) olarak değerlendirileceğinden, iki gizli sınıflı model oluşturulmuştur. Hastalık (+) ve hastalık (-) durumları tek bir gizli değişkenin iki ayrı düzeyi olarak ele alınarak, gizli sınıf küme analizi ile gizli sınıflar tanımlanmıştır. Bu durum iki düzeyli tek bir faktörlü gizli sınıf faktör analizi

olarak da düşünülebilir. Kestirilecek parametre sayısı 7'dir (Her bir teste ait duyarlılık ve seçicilik parametreleri ile prevalans değeri). Hastalık (+) ve hastalık (-) grupları belirlenmiş ve her bir grupta, üç tanı testine ait hassaslık ve hata parametreleri standart hata değerleri ile elde edilmiştir. Daha sonra, her bir tanı testi ile meme lezyonlarının değerlendirilmesine ait, olası yanıt düzenlerine göre optimum sınıfa gözlemlerin ataması gerçekleştirilmiştir. Her bir lezyona ait histopatoloji sonuçlarının bulunması nedeni ile tanı testlerine ait hassaslık parametre değerleri ve histopatoloji sonuçlarına göre prevalans değerleri verilmiştir. GSA ile elde edilen sonuçların doğru sınıflandırma yüzdesi elde edilmiş ve hassaslık parametrelerinin karşılaştırılması yapılmıştır. Daha sonra, tanı testlerinin birleştirilmesi ile test performanslarındaki değişim incelenerek birleştirilmiş testlere ait hassaslık parametreleri elde edilmiştir. Testlerin birleştirilmesinde, en az bir test sonucunun pozitif olması durumunda pozitif, diğer durumlarda negatif olarak sınıflandırma ölçütü ilgili uzmanlar tarafından belirlenmiştir.

3. Bulgular

Her üç tanı testi ile elde edilen gözlenen yanıt düzenleri ve frekansları Tablo 3.1'de verilmektedir.

Tablo 3.1. MG-US, PD ve CE-PD Tanı Testleri Değerlendirmeleri

MG-US	PD	CE-PD	FREKANS
Negatif	Negatif	Negatif	6
Pozitif	Negatif	Negatif	5
Negatif	Pozitif	Negatif	4
Pozitif	Pozitif	Negatif	4
Negatif	Negatif	Pozitif	0
Pozitif	Negatif	Pozitif	1
Negatif	Pozitif	Pozitif	6
Pozitif	Pozitif	Pozitif	21

Oluşturulan modele ait model uygunluğu test istatistik ve ölçüt değerleri Tablo 3.2'de verilmiştir. Model uyumu istatistik değeri L^2 1.0673, model uyumu gerçekleştirildikten sonra, değişkenler arasında varolan açıklanamayan ilişkinin ölçüsüdür. Bu değer ne kadar küçük olursa model uyumu o kadar iyidir. İyi bir uyum L^2 değerinin serbestlik derecesi değerinden çok

fazla farklılaşmadığı durumlarda elde edilir. AIC, BIC ve CAIC değerleri de, L^2 değerinin yanında değerlendirilmesi gerekli indisler olup bu değerlerin küçük olduğu modeller diğer modellere göre tercih edilir. X-kare ve Cressie-Read istatistikleri L^2 'ye alternatif değerler olup geniş örneklemelerde eğer model geçerli ise L^2 ile benzer p değeri verirler. Eğer benzer p değerleri vermezlerse, bulunan L^2 değeri geçerli olmayabilir. Tabloda ayrıca L^2 'nin ve X-kare değerlerinin uygun olmadığı durumlarda kullanılacak log olabilirlik istatistikleri de verilmektedir. Olasılık değeri en yüksek sınıfa atama yapıldığında oluşacak hatalı sınıflandırma yüzdesi tabloda sınıflandırma hata değeri % 5.42 olarak verilmiştir. İndirgeme hataları, Entropy R^2 ve standart R^2 değerleri ise modelin sınıf üyeliklerini ne derecede doğru kestirildiğini gösteren değerlerdir ve bire yakın olmaları istenen durumdur. Tablo 3.3' de GSA sonucu elde edilen gizli sınıf olasılıkları ve kısmi koşullu olasılıklar verilmektedir.

Tablo 3.2. İki Gizli Sınıflı Modele Ait Latent GOLD Sonuçları

2-Sınıflı Model		Ki-Kare İstatistikleri	
Gözlem Sayısı	47	Serbestlik Derecesi	0
Parametre Sayısı	7	L-Kare (L^2)	1.0673
Rastgele Başlangıç	2240313	X-Kare	0.5765
En İyi Başlangıç Noktası	214895	Cressie-Read	0.6749
Sınıflandırma İstatistikleri		BIC (LL'e dayanan)	1.0673
Sınıflandırma Hataları	0.0542	AIC (LL'e dayanan)	1.0673
İndirgeme Hataları	0.8565	CAIC (LL'e dayanan)	1.0673
Entropy R-Kare	0.783	Log-olabilirlik İstatistikleri	
Standart R Kare	0.8263	Log-olabilirlik (LL)	-76.917
Sınıflandırma log-olabilirlik	-83.678	Log-önsel	-3.7576
AWE	242.258	Log-sonsal	-80.6746
		BIC (LL'e dayanan)	180.785
		AIC (LL'e dayanan)	167.834
		CAIC (LL'e dayanan)	187.785

Tablo 3.3. Latent GOLD Gizli Sınıf Olasılıkları ve Koşullu Olasılıkları Profile Sonuçları (Duyarlılık, Seçicilik ve Hata Olasılık Değerleri)

	1. Sınıf Hastalık (+) Grup	Std. Hata	2. Sınıf Hastalık (-) Grup	Std. Hata
Sınıf Büyüklüğü	0.62	0.1042	0.38	0.1042
MG-US				
Negatif	0.2162	0.0797	0.5431	0.1332
Pozitif	0.7838	0.0797	0.4569	0.1332
PD				
Negatif	0.0303	0.0412	0.6223	0.1587
Pozitif	0.9697	0.0412	0.3777	0.1587
CE-PD				
Negatif	0.0707	0.1089	0.9484	0.0889
Pozitif	0.9293	0.1089	0.0516	0.0889

Birinci sınıf vakaların % 62'sini içermektedir. İkinci sınıfta ise % 38'i yer almaktadır. Birinci sınıfta üç tanı testi değerlendirmesi sonucunda model tarafından pozitif (benin) bulunan lezyonlar, ikinci sınıfta ise model tarafından negatif (malin) olarak sınıflandırılan lezyonlar yer almaktadır. Hastalık (+) gruba ait prevalans değeri 0.62 olup, standart hata değeri 0.1042 olarak elde edilmiştir. Sütun yüzdeleri ile hastalık (+) sınıflandırılan lezyonların % 78.38'inin MG-US tanı testi ile % 96.97'sinin PD tanı testi ile % 92.93'ünün CE-PD tanı testi ile pozitif tanı aldığı elde edilmiştir. Bu değerler birinci sınıfta yer alan gözlemlerin belirleyicilerin düzeylerine göre koşullu olasılıkları olup tanı testlerine ait duyarlılık değerlerini tanımlamaktadır. Benzer şekilde tanı testlerine ait seçicilik değerleri ikinci sınıfta, hastalık (-) sınıfında yer alan lezyonların belirleyici değişken olarak alınan her bir tanı testi ile negatif tanı almaları koşullu olasılık değerlerinden elde edilebilir. Buna göre, MG-US tanı testinin seçicilik değeri % 54.31, PD' in seçiciliği % 62, CE-PD testinin seçicilik değeri % 94.84 olarak elde edilmiştir. Tabloda her bir parametreye ait standart hata değerleri de verilmektedir. Yanlış sınıflandırma hata olasılıkları ise birinci sınıfta yer alan bir lezyonun her bir tanı testi ile negatif değer alma olasılıkları, yanlış negatif olasılığı ve ikinci sınıfta yer alan lezyonların her bir tanı testi ile pozitif tanı alma olasılıkları, yanlış pozitif olasılığı ile verilmektedir. Sınıflar ve değişkenler (tanı testleri) arasındaki ilişkiyi ortaya koymanın bir başka yolu, satır yüzdelerinden elde edilen kısmi koşullu olasılık değerlerinin elde edilmesidir. Bir gözlemin belirleyici değişkenin herhangi bir kategorisinde yer alma koşullu olasılığı verildiği durumda, gizli sınıflardan birinde yer alma olasılık değeri tanı testlerinin GSA' de tahmini pozitif ve tahmini negatif

değerlerini tanımlamaktadır. Tablo 3.4'de tahmini pozitif ve tahmini negatif değerler verilmektedir. Buna göre, MG-US ile pozitif tanı alanların % 73.97'si hastalık (+) sınıfında yer alırken, % 26.03'ü hastalık (-) sınıfında yer almaktadır. PD için tahmini pozitif değer (ptd), % 81.22, CE-PD için % 97.67'dir. Benzer şekilde tablodan MG-US için negatif tahmini değer (ntd), % 61.12, PD için % 93.61, CE-PD için % 90.09 olarak bulunmuştur.

Tablo 3.4. Latent GOLD Prob Means Sonuçları,
(Tahmini Pozitif ve Tahmini Negatif Değerler)

	1. Sınıf Hastalık (+) Grup	Std. Hata	2. Sınıf Hastalık (-) Grup	Std. Hata
Tüm Olasılık	0.62	0.1042	0.38	0.1042
MG-GSS				
Negatif	0.3888	0.0797	0.6112	0.1332
Pozitif	0.7397	0.0797	0.2603	0.1332
PD				
Negatif	0.0639	0.0412	0.9361	0.1587
Pozitif	0.8122	0.0412	0.1878	0.1587
CE-PD				
Negatif	0.0991	0.1089	0.9009	0.0889
Pozitif	0.9767	0.1089	0.0233	0.0889

Kümeler, gizli sınıflar üçgenin iki dikey doğrusu ile verilmiştir. Genel sınıf olasılıkları kullanılarak her bir sınıfı temsil eden üçgenin kenarlarını kesecek şekilde sınıf olasılık değerleri birleştirilir. Ayrıca her bir negatif tanıyı gösteren sembollerin yan yana ve daha çok ikinci sınıfta yer aldıkları görülmektedir. Tri-Plot grafiği ile negatif ve pozitif tanıların hangi sınıflarda kümelenme yaptıkları görülebilmektedir. Ayrıca, grafikte her bir tanı testini belirleyen sembollerin birbirlerine yakın olması bu testlere ait sınıf üyelik olasılıklarının benzer olması ile açıklanabilir. Laheij ve ark. (2000), gold standart test olmadığı durumda *Helicobacter pylori* taramasında kullanılan biyopsi tabanlı üç testin hassaslık parametrelerini 7 bilinmeyenli 7 tane non-lineer regresyon denklemini en küçük kareler yöntemi ile çözerek elde etmişlerdir ve tanı testlerinin birleştirilmesi ile hassaslık parametrelerinin artırılacağını göstermişlerdir. Buna göre, MG-US + PD, MG-US + CE-PD ve PD + CE-PD değerlendirmeleri yapıldıktan sonra, GSA ile hassaslık parametreleri elde edilmiştir. GSA ile elde edilen tanı testlerine ait parametre

kestirimleri ve standart hata değerleri Tablo 3.5'de özetlenmiştir. Histopatoloji sonuçlarına göre hassaslık parametreleri ise Tablo 3.6'da verilmektedir.

Tablo 3.5. GSM' e Göre Tanı Testlerine ve Birleştirilmiş Testlere Ait Parametre Kestirimleri (GSM ile Prevalans % 62.0; Testlerin Birleştirilmesi ile Prevalans % 86.32)

Tanı Testi	Duyarlılık (%)	Std. Hata	Seçicilik (%)	Std. Hata	Ptd (%)	Ntd (%)
MG-US	78.38	0.0797	54.31	0.1332	73.97	61.12
PD	96.97	0.0412	62.23	0.1587	81.22	93.61
CE-PD	92.93	0.1089	94.84	0.0889	97.67	90.09
MG-US + PD	99.84	0.0062	92.31	0.1119	99.82	99.99
MG-US + CE-PD	90.17	0.0465	93.47	0.1007	99.90	61.92
PD + CE-PD	87.73	0.0512	93.63	0.0997	99.89	56.50

Tablo 3.6. Histopatoloji Sonuçlarına Göre Tanı Testlerine ve Birleştirilmiş Testlere Ait Parametre Değerleri (Histopatoloji Sonuçlarına Göre Prevalans % 53.2)

Tanı Testi	Duyarlılık (%)	Seçicilik (%)	Ptd (%)	Ntd (%)	Prevalans (%)
MG-US	100.0	72.7	80.6	100.0	66.0
PD	84.0	36.4	60.0	36.4	74.5
CE-PD	80.0	63.6	71.4	63.6	59.6
MG-US + PD	100.0	27.3	61.0	100.0	87.2
MG-US + CE-PD	100.0	45.5	67.6	100.0	78.7
PD + CE-PD	88.0	36.4	61.1	72.7	76.6

Gold standart olarak kabul edilen histopatoloji sonuçlarına göre test sonuçları değerlendirildiğinde, hastalık + olan kişilerin oranı (prevalans değeri), MG-US ile % 66, PD ile % 74.5, CE-PD ile % 59.6 elde edilirken, GSM ile % 62, testlerin birleştirilmesi ile % 86.3 olarak bulunmuştur. Gold standart olarak kabul edilen histopatoloji sonuçlarına göre 47 hastadan 25'ine hastalık (+) tanısı konmuş ve prevalans değeri 53.2 elde edilmiştir.

Fakat biyopsi tabanlı testlerde bile hata olabileceği değerlendirmelerde göz önünde bulundurulmalıdır. Testlerin birleştirilmesi sonucu elde edilen prevalans değerleri gerçek prevalans değeri % 53.2'den oldukça yüksek elde edilmiştir. Tanı testlerine ait yanıt düzenekleri için gözlenen ve kestirilen gözlem sayıları Tablo 3.7'de, birleştirilmiş tanı testlerine ait gözlenen ve kestirilen gözlem sayıları ise Tablo 3.8'de verilmiştir. Tablolardan da görüldüğü üzere, gözlenen ve kestirilen değerler arasındaki fark oldukça az olup, standart artık değerleri küçük bulunmuştur. Bu durum, elde edilen modellerin veriye uygunluğunun bir başka değerlendirmesidir.

Tablo 3.7. Tanı Testlerine Ait Yanıt Düzenekleri İçin Gözlenen ve Kestirilen Gözlem Sayıları

MG-US	PD	CE_PD	Gözlenen	Kestirilen	Std. Artıklar
Negatif	Negatif	Negatif	6	5.7	0.11
Negatif	Pozitif	Negatif	4	3.9	0.05
Negatif	Pozitif	Pozitif	6	5.9	0.06
Pozitif	Negatif	Negatif	5	4.9	0.06
Pozitif	Negatif	Pozitif	1	0.9	0.09
Pozitif	Pozitif	Negatif	4	4.5	-0.23
Pozitif	Pozitif	Pozitif	21	20.8	0.05

Tablo 3.8. Birleştirilmiş Tanı Testlerine Yanıt Düzenekleri İçin Gözlenen ve Kestirilen Gözlem Sayıları

MG-US+PD	MG-US+CE-PD	PD+CE_PD	Gözlenen	Kestirilen	Std. Artıklar
Negatif	Negatif	Negatif	6	5.2	0.35
Pozitif	Negatif	Pozitif	4	3.5	0.25
Pozitif	Pozitif	Negatif	5	4.5	0.23
Pozitif	Pozitif	Pozitif	3	3.2	-0.01

Tablo 3.9'da ve Tablo 3.10'da tanı testlerinden ve birleştirilmiş tanı testlerinden elde edilen olası yanıt düzeneklerine göre kestirilen sınıf üyelik olasılıkları ve sınıf atamaları verilmiştir.

Tablo 3.9. GSM ile Tanı Testleri Sonuçlarına Göre Sınıflandırma ve Atama Olasılıkları

MG-US	PD	CE_PD	Model	1. Sınıf Hastalık (+)	2. Sınıf Hastalık (-)
Negatif	Negatif	Negatif	-	0.0024	0.9976
Negatif	Pozitif	Negatif	+	0.1106	0.8894
Negatif	Pozitif	Pozitif	-	0.9677	0.0323
Pozitif	Negatif	Negatif	-	0.0101	0.9899
Pozitif	Negatif	Pozitif	+	0.7105	0.2895
Pozitif	Pozitif	Negatif	-	0.3489	0.6511
Pozitif	Pozitif	Pozitif	+	0.9923	0.0077

Her bir olası yanıt düzeni için sınıf olasılıkları hesaplandıktan sonra en yüksek olasılık değerine sahip sınıfa atama gerçekleştirilmektedir. Örneğin; her üç tanı testinde pozitif olan lezyonlar birinci sınıfa atanırken, testlerden en az ikisinin uyumlu olması durumunda sınıflandırma bu ortak tanı yönünde olmaktadır.

Tablo 3.10. GSM ile Birleştirilmiş Tanı Testleri Sonuçlarına Göre Sınıflandırma ve Atama Olasılıkları

MG-US+PD	MG-US+CE-PD	PD+CE-PD	Model	1. Sınıf Hastalık (+)	2. Sınıf Hastalık (-)
Negatif	Negatif	Negatif	-	0.0001	0.9999
Pozitif	Negatif	Pozitif	+	0.9916	0.0084
Pozitif	Pozitif	Negatif	+	0.9933	0.0067
Pozitif	Pozitif	Pozitif	-	0.9999	0.0001

Gizli sınıf modeli ile elde edilen sınıflandırmaların histopatoloji sınıflandırmaları ile karşılaştırması Tablo 3.11'de, birleştirilmiş testlere ait sınıflandırmaların karşılaştırması ise Tablo 3.12'de verilmektedir.

Tablo 3.11. Tanı Testlerinin Sonuçlarına Göre Gizli Sınıf Modeli ile Elde Edilen Sınıflandırmaların Histopatoloji Sınıflandırmaları ile Karşılaştırması

Gizli Sınıf Modeli Sınıflandırması		Histopatoloji Sınıflandırması		Toplam
		Hastalık (-)	Hastalık (+)	
Hastalık (-)	Sayı	14	5	19
	Satır yüzdesi	73.7	26.3	100.0
	Sütun Yüzdesi	63.6	20.0	40.4
Hastalık (+)	Sayı	8	20	28
	Satır yüzdesi	28.6	71.4	100.0
	Sütun Yüzdesi	36.4	80.0	59.6
Toplam	Sayı	22	25	47
	Satır yüzdesi	46.8	53.2	100.0
	Sütun Yüzdesi	100.0	100.0	100.0

Tablo 3.12. Birleştirilmiş Tanı Testlerinin Sonuçlarına Göre Gizli Sınıf Modeli ile Elde Edilen Sınıflandırmaların Histopatoloji Sınıflandırmaları ile Karşılaştırması

Gizli Sınıf Modeli Sınıflandırması (Birleştirilmiş Testlere Göre)		Histopatoloji Sınıflandırması		Toplam
		Hastalık (-)	Hastalık (+)	
Hastalık (-)	Sayı	6		6
	Satır yüzdesi	100		100
	Sütun Yüzdesi	27.3		12.8
Hastalık (+)	Sayı	16	25	41
	Satır yüzdesi	39.0	61	100.0
	Sütun Yüzdesi	72.7	100	87.2
Toplam	Sayı	22	25	47
	Satır yüzdesi	46.8	53.2	100.0
	Sütun Yüzdesi	100.0	100.0	100.0

Her bir tanı testi değerlendirmeleri kullanılarak elde edilen gizli sınıf modeli ile 28 hastaya pozitif tanı (% 59.6), 19 hastaya negatif tanı konmuştur (% 40.4). Histopatoloji sonuçlarına göre 25 hasta pozitif tanı (% 53.2), 22 hasta negatif tanı (% 46.8) almıştır. Her iki sınıflandırmada toplam 34 hastanın tanısı ortak konmuş ve gizli sınıf modelin doğru sınıflandırma yüzdesi % 72.4 olarak elde edilmiştir. Gizli sınıf modeli ile elde edilen

tanıların duyarlılığı % 80, seçiciliği ise % 63.6 bulunmuştur. Tahmini pozitif değer % 71.4, tahmini negatif değer ise % 73.7 hesaplanmıştır. Tanı testlerinin birleştirilmesiyle oluşturulan gizli sınıf modeli ile 41 hastaya (% 87.2) pozitif tanı, 6 hastaya (% 12.8) negatif tanı konmuştur. Elde edilen sonuçların duyarlılığı % 100, seçiciliği % 27.3, tahmini pozitif oranı % 61, tahmini negatif oranı % 100 bulunmuştur. Bu gizli sınıf modeli sınıflandırmaları histopatoloji sınıflandırmaları ile karşılaştırıldığında doğru sınıflandırma yüzdesi % 66 olarak bulunmuştur. Gizli sınıf modeli ile oluşturulan model parametrelerinin istatistiksel önem kontrolü Wald istatistiği ile gerçekleştirilmektedir. Tablo 3.13'de parametrelere ait Wald istatistik değerleri verilmektedir. Her üç teste ve birleştirilmiş testlere ait parametre kestirimleri önemli bulunmuştur ($p=0.045$; $p=0.010$; $p=0.024$). Ayrıca belirleyicilik yüzdeleri de R^2 değeri ile verilmektedir. Hastalığı tanımlamadaki açıklayıcılık değeri en yüksek CE-PD testi ile birleştirilmiş testlerde ise MG-US + PD ile elde edilmiştir.

Tablo 3.13. GSM İle Elde Edilen Testlere ve Birleştirilmiş Testlere Ait Parametrelerin Önem Kontrolü

	1. Sınıf Hastalık (+) Grup	2. Sınıf Hastalık (-) Grup	Wald İstatistik değeri	P değeri	R ²
MG-US					
Negatif	-0.3652	0.3652	4.0170	0.045	0.1122
Pozitif	0.3652	-0.3652			
PD					
Negatif	-0.9911	0.9911	6.5577	0.010	0.4343
Pozitif	0.9911	-0.9911			
CE-PD					
Negatif	-1.3715	1.3715	5.1240	0.024	0.7535
Pozitif	1.3715	-1.3715			
MG-US + PD					
Negatif	-2.2363	2.2363	4.3992	0.036	0.9007
Pozitif	2.2363	-2.2363			
MG-US + CE-PD					
Negatif	-1.2191	1.2191	7.9285	0.0049	0.4932
Pozitif	1.2191	-1.2191			
PD + CE-PD					
Negatif	-1.1636	1.1636	7.1696	0.0074	0.4361
Pozitif	1.1636	-1.1636			

4. Sonuç ve Öneriler

GSA'ne ait çalışmalar ve uygulamalar yöntemin kolaylığı ve hazır yazılım programları ile uygulanabiliyor olması nedeniyle günden güne artmaktadır. Literatürde yabancı çalışmalarda GSA'nin çok farklı alanlarda çok sayıda uygulaması ile karşılaşılmışına karşın ülkemizde bu konu ile ilgili özellikle biyoistatistik alanında ve tıp uygulamalarında çalışmalar oldukça azdır. Bu nedenle çalışmada GSA ile verilen teorik bilgilerin ve açıklamaların faydalı olacağı düşünülmektedir. Özellikle tanı testlerinin değerlendirilmesinde verilen uygulamanın klinisyenlere farklı bir bakış açısı kazandırarak karar verme aşansında faydalı bilgiler sağlayacağı düşünülmektedir. GSA rahatlıkla pek çok alanda çalışmacılar tarafından uygulanabilecek bir yeni analiz yöntemidir. İleriye yönelik olarak önerilecek çalışmalar teorik ve uygulama çalışmaları olarak ayrılabilir. Araştırmacılar her türlü alanda uygulayacakları geniş çaplı bir anket çalışmasından elde ettikleri gözlemlerin değerlendirmesinde ve farklı grup yapılarının oluşturulmasında GSA yöntemini inceleyebilir. Bununla beraber oluşturulan yeni bir ölçüm skalasının veya bir testin geçerlilik ve güvenilirlik incelemesinde GSA yaklaşımını kullanabilir. Bu çalışmada verilen birden fazla tanı testlerinin hassaslık parametrelerinin değerlendirmesi her türlü uyum verisinin elde edildiği çalışmalara uygulanabilir. "Altın standart" test olmaması durumunda verilen literatürde yer alan diğer çalışmalar taranarak GSA ile karşılaştırmaları farklı uygulamalar üzerinde verilebilir. Ayrıca çalışmada verilen her bir uygulama alanında ve ortak kullanım alanında GSA rahatlıkla uygulanabilir. Faktör analizi, küme analizi ile diğer benzer ve farklı yönleri verilen yöntemler teorik açıdan karşılaştırılabilir, uygulamalar üzerinde bu karşılaştırmalar verilebilir. Teoride verilen lokal bağımsızlık ve kullanılacak testler arasındaki bağımlılık durumu incelenebilir. Parametre tahmininde ve standart hataların hesaplanmasında verilen yöntemlerin teorik karşılaştırmaları incelenebilir.

Kaynaklar

1. Hagenars, A.Jacques., McCutcheon, L.Allan. (2002). Applied Latent Class Analysis. Cambridge University Press.
2. Rynard, V. L. T., Walter, S. D. "Effects of dependent errors in assesment of diagnostic test performance", Statistics in Medicine, 16, 2157-2175, 1997.
3. Alonzo, T. A., Pepe, M. S. "Using a Combination of reference tests to assess the accuracy of a new diagnostic test", Statistics in Medicine, 18, 2987-3003, 1999.

4. Uebersax, S.John., Grove, M.William. (1990). Latent Class Analysis of Diagnostic Agreement. *Statistics in Medicine*, Vol:9, 559-572.
5. Lazarsfeld, P. F ve Henry, N.W. *Latent Structure Analysis*, Houghton Mifflin, Boston, Massachusetts, 1968.
6. Goodman, L. A. "Exploratory latent structure analysis using both identifiable and unidentifiable models", *Biometrika*, 61, 215-231, 1974b.
7. Haberman, S. J. *Analysis of Qualitative Data, Volume 2: New Developments*, New York: Academic Press, 1979.

**ALTIN STANDART YÖNTEMİN BULUNMADIĞI
ÇALIŞMALARDA TANI TESTLERİNİN ETKİNLİĞİNİN
SAPTANMASINDA GİZLİ SINIF ANALİZİNİN
KULLANILMASI**

**E.Arzu Kanık¹, Bahar Taşdelen², Murat Arslan³, Mahmood Morshedi⁴,
Sergio Oehninger⁵**

Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Mersin, Türkiye¹

*Mersin Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Mersin,
Türkiye²*

*Mersin Üniversitesi, Tıp Fakültesi, Kadın Hastalıkları ve Doğum Anabilim
Dalı, Mersin, Türkiye³*

*The Jones Institute for Reproductive Medicine, Department of Obstetrics and
Gynecology, Eastern Virginia Medical School, Norfolk, VA, USA⁴*

*The Jones Institute for Reproductive Medicine, Department of Obstetrics and
Gynecology, Eastern Virginia Medical School, Norfolk, VA, USA⁵*

[e-mail:arzukanik@mersin.edu.tr](mailto:arzukanik@mersin.edu.tr)

Özet

Tanı testlerinin etkinliğini konu alan araştırmalarda altın standart bir testin bulunmadığı veya doğrudan görülmediği durumlarda tanı testlerinin performanslarını değerlendirmek oldukça önemli bir problemdir. Bu çalışmada bu problemin çözümü için gizli sınıf analizi yaklaşımı önerilmiş ve uygulamalar örnek bir veri seti üzerinde gösterilmiştir. Ayrıca tanı testlerinin prevelanslarının gizli sınıf analizi sonuçlarını nasıl etkilediği N=1000 olan popülasyonlar üzerinde yapılan simülasyon çalışmalarıyla izlenmiştir. Bu veri setinde tüm dünyada güncelliğini koruyan bir problem olan infertilite verileri kullanılmıştır. Sebebi açıklanamayan veya erkek kaynaklı infertil olduğu saptanan tedavi altındaki 82 çifte vücut dışında dölleme uygulamalarına göre daha kolay ve ucuz olan intrauterin inseminasyon (aşılama yöntemi) 3 ay süreyle uygulanmıştır. 3 ay sonunda anne adaylarında gebelik oluşmaması, yumurtaların döllemediğini tam olarak açıklayamadığından bu araştırmada dölleme olup olmaması gizli bir altın standart olarak kabul edilmektedir. Bu veri setinde sperm kalitesini gösteren 4 binary değişken (motility, sperm concentration, morphology and hemizona index (HZI)) ile gizli sınıf analizi yapılmış ve bu değişkenlerin kombinasyonlarına göre

fertilite olasılıkları hesaplanmıştır. Simülasyon çalışmalarında 2, 3 ve 4 adet bağımsız Bernoulli değişkeni farklı prevelans kombinasyonlarında denenmiştir. Hesaplamalarda LCAP v2.34 latent class analysis program kullanılmıştır. Yapılan simülasyon çalışmaları sonuçları tablolarda özetlenmiştir. Sonuçların altın standart yöntemin bulunmadığı durumlarda kullanılacak tanı testlerinin seçiminde yol göstereceği düşünülmektedir.

Anahtar Kelimeler: Gizli sınıf, altın standart, prevelans, simülasyon.

TELE-RADYOLOJİ UYGULAMASI ETKİLİLİĞİNİN GİZLİ SINIF (LATENT CLASS) ANALİZİ İLE DEĞERLENDİRİLMESİ

Meriç Çolak¹, Vildan Sümbüloğlu², Cenk Sözen¹, Korkut Ersoy¹

Sağlık Kurumları İşletmeciliği Bölümü, Başkent Üniversitesi, ANKARA¹
Biyoistatistik Ana Bilim Dalı, Tıp Fakültesi, Kara Elmas Üniversitesi,
ZONGULDAK²

e-mail: meric@baskent.edu.tr

Özet

Çalışmanın amacı, Tele-Radyoloji uygulamasında, gözlemcilerarası uyumun, güvenilirliğin değerlendirilmesinde olasılıksal model yaklaşımının Gizli Sınıf Analizi (GSA) ile verilmesidir. Gizli Sınıf Analizi (GSA), doğrudan gözlenemeyen (gizli) bir teorik kavramın, gizli değişkenin mükemmel belirleyicileri olmayan gözlenen değişkenler ile arasındaki kovaryasyonu açıklamada uygulanabilir bir yöntemdir. Amaç, gözlenen yanıtlar arasındaki ilişkiyi açıklayacak birbirinden tamamen ayrık ve bağımsız sınıfları tanımlamaktır. Gizli sınıf modelleri tanı koyucuların, gözlemcilerin ve tanı testlerinin hassaslıklarını ve değişen hastalık kategorilerine göre prevalans değerini gözlenemeyen parametre olarak elde etmeyi amaçlar. Gözlenen veriye dayanarak bu parametreler kestirilir. GSA ile, gözlemcilerarası uyumun, güvenilirliğin karakteristikleri verilebilir, test edilebilir ve gözlemcilere ait hata oranları ve standart hatalar hesaplanabilir. GSA ile bu çalışmada, Tele-Radyoloji uygulaması sonucu dört radyolog tarafından 8 farklı Tüberküloz (TB) bulgularının mevcudiyetine dair oluşturulan yanıt düzenekleri bir arada değerlendirilerek, bulguların varlığını veya yokluğunu gösteren gizli sınıflar 8 farklı radyolojik bulgu için ayrı ayrı ortaya konulmuştur. GSA ile, gözlemcilerarası uyum, güvenilirlik, uyum karakteristikleri verilmiş, gözlemcilerarası uyum test edilmiş ve gözlemcilere ait hata oranları ve standart hatalar hesaplanmıştır.

Çalışmanın verilerini Tele-Radyoloji yöntemi ile elde edilen Tüberküloz'un 8 farklı radyolojik bulgusuna ait, radyologların verdikleri yanıt düzenekleri oluşturmaktadır. Başkent Üniversitesi Ankara Hastanesinin arşivinden (TB) tanısı konulmuş hastaların taranması sonucu elde edilen filmlerin içinden 40 adet ayakta Postero Anterior Akciğer grafisi seçilmiştir. Orijinal imajlar, Başkent Üniversitesi Ankara Hastanesi Radyoloji Bölümü'ne bağlı 2 uzman radyolog tarafından incelenmiş ve anket üzerinden raporlanmıştır. Gülhane Askeri Tıp Akademisi Haydarpaşa Hastanesine bağlı 2

uzman Radyolog ise, dijital fotoğraf makinesi sistemi ile elde edilen dijital imajları, oluşturulan web sayfası üzerinden incelemiş ve web sitesi üzerinde oluşturulan on-line anket ile raporlamıştır. On-line anket ile elde edilen veriler Başkent Üniversitesi Bilgi İşlem Departmanına ait verita-banında depolanmış, çalışma sonunda veriler toplu halde elde edilerek SPSS veri formatına dönüştürülmüştür. Radyologlar oluşturulan anket ile grafilerde Tüberkülozun 8 adet radyolojik bulgusunun (İnfiltrasyon, Kavite formasyonu, Pnömotoraks, Plevral efüzyon, Kalsifikasyon, Skar dokusu, Lenfadenopati, Nodül veya kitle) mevcudiyetini likert ölçek ile skorlamışlardır. Radyologlar ayrıca dijital imajlar ve orijinal grafiler arasında radyografik kalite açısından anlamlı bir fark olup olmadığını değerlendirmek üzere imajlar ve grafileri; Netlik, Kontrast, Pozisyon ve Santralizasyon kriterleri açısından da skorlamışlardır.

GSA uygulamalarında, analizlerde kolaylık sağlamak amacı ile skorlar üç grupta birleştirilmiştir. TB bulgularına ait değişkenlerin düzey sayısının üç olması nedeni ile, 3 sınıflı gizli sınıf modeli oluşturulmuştur. $3^4=81$ olası yanıt düzeni olmasına karşın, gözlenemeyen yanıt düzenleri gizli sınıf analizinde değerlendirmeye alınmamıştır.

Anahtar Kelimeler: *Gizli Sınıf Analizi, Gözlemcilerarası Uyum, Tele-Radyoloji.*

Giriş

Çalışmanın amacı, Tele-Radyoloji uygulamasında, gözlemcilerarası uyumun, güvenilirliğin değerlendirilmesinde olasılıksal model yaklaşımının Gizli Sınıf Analizi (GSA) ile verilmesidir. Gizli Sınıf Analizi (GSA), doğrudan gözlenemeyen (gizli) bir teorik kavramın, gizli değişkenin mükemmel belirleyicileri olmayan gözlenen değişkenler ile arasındaki kovaryasyonu açıklamada uygulanabilir bir yöntemdir. Amaç, gözlenen yanıtlar arasındaki ilişkiyi açıklayacak birbirinden tamamen ayrık ve bağımsız sınıfları tanımlamaktır. Gizli sınıf modelleri tanı koyucuların, gözlemcilerin ve tanı testlerinin hassaslıklarını ve değişen hastalık kategorilerine göre prevelans değerini gözlenemeyen parametre olarak elde etmeyi amaçlar. Gözlenen veriye dayanarak bu parametreler kestirilir. GSA ile gözlemcilerarası uyumun, güvenilirliğin karakteristikleri verilebilir, test edilebilir ve gözlemcilerarası hata oranları ve standart hatalar hesaplanabilir. GSA ile bu çalışmada, Tele-Radyoloji uygulaması sonucu dört radyolog tarafından 8 farklı Tüberküloz (TB) bulgularının mevcudiyetine dair oluşturulan yanıt düzenekleri birarada değerlendirilerek, bulguların varlığını veya yokluğunu gösteren gizli sınıflar 8 farklı radyolojik bulgu için ayrı ayrı ortaya konulmuştur. GSA ile, gözlemcilerarası uyum, güvenilirlik, uyum karakter-

istikleri verilmiş, gözlemcilerarası uyum test edilmiş ve gözlemcilere ait hata oranları ve standart hatalar hesaplanmıştır.

1.1 Tele - Tıp ve Tele Radyoloji

Tele-tıbbın sakla ve gönder teknolojisine dayalı uygulamaları, gelişmekte olan ülkelerdeki doktorlara klinik verileri ve dijital görüntüleri, uzaktaki radyologlara uzmanlık bilgilerine danışmak için göndermede düşük maliyetli alternatifler sunmaktadır. Mevcut iletişim hatları ile internet üzerinden bu verileri ve dijital görüntüleri, e-maillere iliştiyerek ya da bunun için tasarlanmış internet sayfaları üzerinden erişime hazır hale getirerek, Türkiye'nin de karşı karşıya kaldığı sağlık profesyonellerinin ülke çapında heterojen dağılımından kaynaklanan ve hem hizmet sunumunun etkililiği hem de maliyetinde yaşanan sorunlar azaltılabilir. 30 yıldan daha fazla zamandır, klinisyenler, sağlık hizmeti araştırmacıları ve diğer sağlık profesyonelleri gelişmiş telekomünikasyon sistemlerinin ve bilgi teknolojilerinin sağlık düzeyini yükseltmek amacıyla nasıl kullanılabileceğini araştırmaktadırlar. Bu çabaların kesiştiği yerde yenilikçi ve önde gelen teknolojilerin bir kombinasyonu olan "tele-tıp" ortaya çıkmıştır. Bu tanıma göre tele-tıp, elektronik bilgisi ve iletişim teknolojilerinin, sağlık profesyonelleri arasındaki bilgi paylaşımının mesafe faktörü yüzünden engellendiği durumlarda sağlık hizmeti sağlamak ve mevcut hizmeti desteklemek amacı ile kullanılmasıdır. Başka bir deyişle tele-tıp, hekimlerin hastaları iletişim teknolojileri sayesinde uzaktan inceleyebildiği bir sağlık hizmetleri dağıtım sistemidir. Tele-tıp, video konferans gibi uygulamalar klinisyenlere uzaktaki hastalarını görme, işitme, muayene etmenin yanında hastalara soru sorma ve direktifte bulunma gibi gerçek zamanlı teşhis ve tedavi uygulamaları fırsatını verirken; "depola ve gönder" teknolojisine dayalı olan uygulamalar dijital görüntü ve diğer bilgilerin hafızaya alınıp değerlendirilmeyi yapacak olan danışman yada sağlık profesyonellerine ucuz ve hızlı bir yol olan internet yada network sistemi üzerinden iletilmesi imkanı sağlar.

Başka bir hastanede tıbbi teknik personel tarafından hazırlanmış olan radyolojik görüntülerinin ikinci bir görüş almak üzere bu konuda uzmanlaşmış başka bir hastanedeki uzmanlara gönderilebilmesi sağlık hizmeti sunumunda fiziki sınırları ortadan kaldırmıştır. İdeal bir dünyada her yerde bir radyoloji uzmanının bulunması ve röntgen çekme imkan ve kabiliyetine sahip olunması beklenmekte, ancak bu gerçekte pek mümkün olamamaktadır. Geleneksel çözüm, röntgen filmlerinin radyoloji uzmanına götürülmesi ve onun bunları inceleyerek değerlendirmesi ile elde edilmektedir. Bant genişliği imkanlarının artması ve donanım fiyatlarının hızla düşmesi sayısal veri transferi imkanlarının artmasını sağlamıştır. Bu sayede, röntgen filmlerinin elektronik transferi olan tele-radyoloji kavramı ortaya

çıkılmıştır. Gerekli cihazlarla donatılması durumunda radyoloji uzmanı hastayı görmeden elektronik olarak kendisine iletilen röntgen filmlerini inceleyerek bir değerlendirme yapma imkanına kavuşmuştur. Bir tele-radyoloji sistemi cerrahlara da hastayı görmeksizin, ameliyat öncesi ve sonrası röntgen filmlerini inceleme fırsatı vermektedir. Ayrıca, birinci basamak sağlık hizmeti veren aile hekimi uzmanlarına röntgen dahil, hasta verilerine daha kolay ulaşma imkanı sunmaktadır. Böylelikle uzmanlarla yapılacak video-konferans görüşmesi öncesinde gerekli tüm bilgilere ulaşarak hastaların gereksiz yere yolculuk etmesi önlenerek tedaviyi hızlandırma imkanı yaratılmış olmaktadır.

Tele-Radyolojinin en önemli yönü maliyet ve klinik olarak etkinliğidir. Burada önemli nokta sistemlerin güvenilirliği, iletilen resimlerin kalitesi, resimlere ulaşmadaki hız ve kullanım kolaylığıdır (Klein ve Manning, 1995). Sistem röntgen filmlerinin çıktıları alınıncaya kadar resimleri elektronik ortamda en az bir hafta saklayabilecek kapasitede olmalıdır. Birçok ülkede yılları kapsayacak bir periyotta saklamak yasal bir gerekliliktir. Aynı zamanda eski filmlerle beraber görüntüleme ve istenilen bölümü büyütme imkanı olması istenilen bir özelliktir. Ayrıca, röntgen filmi ile ilgili raporların, bilhassa sayısal ses kayıtlarının da resimlerle birlikte tutulması arzu edilen bir özellik olarak günümüzde göze çarpmaktadır. Tele-radyoloji uygulamalarında, sakla ve gönder (Store-and-forward) adı verilen bir yöntem uygulanır. Bu yöntem temelde multimedya elektronik postalardan ve web sayfalarından oluşur. Resimlerin transferi (örneğin CT taramaları veya Röntgen filmleri), laboratuvar verileri, geçmiş hasta kayıtları ve muayene bulgularının tek bir e-maile eklenmesi ya da web sayfasına konulması ve bunlarla ilgili görüşlerin geri gönderilmesi esasına dayanmaktadır. Bir Tele-Radyoloji sistemi, bir haberleşme ağına bağlı resim elde etme bölümü ile resim gösterme/yorumlama bölümünden oluşur. Tele-radyolojinin kardeş teknolojisi olan resim arşivleme ve haberleşme sistemi ise (Picture Archiving and Communication System-PACS), hastanedeki sayısal imajları saklama, arşivleme ve aynı zamanda transferinde kullanılır. Aralarındaki fark PACS yerel alan ağı (Local Area Network- LAN) kullanırken tele-radyoloji geniş alan ağlarını (Wide Area Networks-WAN) kullanmaktadır

Bu araştırmada Tele-Tıbbın pratik bir uygulaması olan sakla ve gönder tekniğinden yararlanılmıştır. Bu teknikte hastanın filmi dijitalize edilip, anamnez bilgilerini içeren bir form ile beraber bu amaçla oluşturulan web sayfasına konulmuştur. Fakat imajların web sayfasında çok yer kaplaması bu tekniğin önündeki potansiyel bir problemdir ve imajların sıkıştırılmasını gerektirmektedir. Bundan önce yapılan çalışmalar doğru kriterlerde yapılan sıkıştırılmaların imajların diagnostik kalitesinde fark edilir bir değişikliğe neden olmadığını göstermiştir. Bu araştırma sakla ve gönder tekniğinin geniş çapta uygun, etkili bir tele-radyoloji uygulaması olup olmadığını değerlendirmek

için tasarlanmış olup, radyologlararası uyum, Gizli Sınıf Analiz yaklaşımı ile elde edilerek, yöntemin güvenilirliği incelenmiştir.

1.2 Gözlemcilerarası Uyum Verilerinin Değerlendirilmesi

Uyum verisi vakaların aynı veya farklı gözlemciler tarafından değerlendirildiği araştırma düzenlerinden elde edilebilir. Gözlemcilerin aynı olması durumu sabit gözlemcili panel çalışmaları, gözlemcilerin farklı olması durumu ise değişen gözlemcili panel çalışmaları olarak adlandırılmaktadır. Sabit gözlemcili veya çoklu tanı testi sonuçlarının elde edildiği panel çalışmalar, k tane gözlemcinin her vakayı değerlendirdiği veya k tane tanı testi ile vakaların sınıflandırıldığı tamamı ile çapraz düzenlerdir. Değişen gözlemcili panel çalışmalarına benzer çalışmalarda, modelde her bir gözlemcinin veya testin koşullu sınıflama olasılıklarını dikkate alan ek parametreler gerekmektedir. Uebersax ve Grove (1990)'de bu modellerin Dawid ve Skene, Dillon ve Mulani, Walter ve Irwig ile Espeland ve Handelman tarafından tartışıldığı verilmektedir. Hastalık durumu kesikli veya sürekli olabilir. Kesikli hastalık modellerinde, bir vakanın sınıflandırmada pozitif yanıt verme olasılığının aynı kategoride yer alan diğer vakalar arasında sabit, fakat farklı kategoride yer alanlar arasında farklı olduğu varsayılır. Sürekli hastalık modelleri ise, hastalığın değişen düzeylerinin ve vakaların pozitif tanı alma olasılıklarının sürekli olmasına olanak tanır. Sürekli hastalık modelleri her ne kadar daha hassas olsalar da bu modellerin hesaplamaları oldukça karmaşıktır (Uebersax ve Grove, 1990).

Medikal araştırmalarda tanı koyucular arasındaki uyumun değerlendirilme ihtiyacı oldukça ortak bir konudur. Uyumun ölçülmesinin en önemli nedenlerinden bir tanesi yöntemin verilen tanıların hassaslıkları ile ilgili hipotez testlerine olanak tanımasıdır. Örneğin; yüksek derecede bir uyum her ne kadar tek tek gözlemcilere ait tanıların doğruluğunu ifade etmese de, uyum yetersizliği konulan tanıların genellikle doğru olmadığını ifade eder.

Medikal enformasyon araştırmacıları sıkça uyum ölçütlerini iki veya daha fazla gözlemcinin verdiği sınıflamalar arasındaki benzerliğin boyutunu ortaya koymak amacı ile kullanırlar. Fakat doğru bir standardın olmadığı durumda böyle bir referansın kullanılmasından evvel, referans standardın kalitesin değerlendirilmesi için ölçüm çalışmasının yapılması gerekmektedir. Bu ölçütlerden bir tanesi gözlemcilerin verdiği kararların güvenilirliğidir. Kategorik yanıt yapıları için uyum, güvenilirliğin bir ölçütü olarak kullanılabilir. Eğer gözlemcilerarasında çok az bir uyum var ise, oluşturdukları yanıt düzenleri güvenilir değildir ve referans standardın kalitesi de bu duruma bağlı olacaktır. Literatürde çok sayıda uyum ölçütleri önerilmektedir. Bunlardan en popüler üçü gözlenen uyum, özel uyum ve kappa'dır. Gözlenen uyum, gözlemcilerarası uyum yüzdesidir. Özel uyumda, her bir kategori için ayrı ayrı uyumun derecesi ortaya konulur. İkili durumlarda pozitif

ve negatif özel uyum hesaplanabilir. Bir kategorideki gözlemcilerin uyum gösterdikleri durumların oranı ile verilmektedir. Kappa istatistiği ise, gözlenen uyum ve şansa bağlı beklenen uyum cinsinden ifade edilir.

Kategorik verilerin bulunduğu güvenilirlik çalışmalarında, uyum ölçütleri sıkça kullanılmaktadır. Gözlenen uyum, özel uyum gibi basit ölçütler, örnek hakkında iyi bir fikir verebilir. Kappa istatistiği intraclass korelasyon katsayısına uyumu ve hesaplama kolaylığı nedeni ile oldukça sık kullanılmaktadır. Fakat Kappa istatistiği kullanılan diagnostik yöntemin duyarlılık ve seçicilik değerinden ve incelenilen olayın prevalansından etkilenir. Eğer amaç gözlemcilerarası uyumun veya tanı testinin güvenilirliğini artırmak ise, uyumsuzluğun bileşenlerinin ayrılmasında faydalı bir ölçüttür. Gizli sınıf modelleri, Gizli trait modelleri, polychoric ve tetrachoric korelasyon gibi modellemeye dayalı karar verme süreçleri bu noktada faydalı olabilir. Farklı uyum yapılarının davranışlarının anlaşılmasında bu modeller kullanılabilir. Örneğin, iki mevcut kategoriden bir tanesinde gözlenen yanıtların prevalansı düşük ise, gözlemcilerin gözlemleri ayrıştırma kabiliyeti için örnekleme yetersiz bilgi olduğu söylenebilir ve kappa istatistiği doğru uyumu olduğundan daha az tahminleyebilir ve gözlenen uyum değerinin üzerinde tahminleyebilir (Hripscak, 2002).

GSA kullanılarak ortaya atılan stratejilerde bu problemler ortaya konulmuştur. Ayrıca kappa istatistiği uyumsuzluğun kaynağının belirlenmesine olanak tanımaz ve kappa istatistiği tanının en iyi tahmininin güvenilirliği ile ilgili faktörlerin tanımlanmasına izin vermez. (Flaherty, 2002). Gözlemciler, gözlemlerin belirli kategorilere atanmasında önemli bir hata kaynağı oluşturmaktadır. Bu nedenle hangi atamaların birbiri ile uyum halinde olduğunun belirlenmesi önem taşımaktadır.

Gözlemcilerarası uyumun değerlendirmesinde iki temel yaklaşım ayırt edilebilir. Birinci yaklaşım gözlemcilerarası uyumu indeksleyecek özet ölçümlerin ortaya konulmasında odaklanırken, ikinci yaklaşımda gözlemcilerin verdiği yanıt yapılarının arasındaki ilişkinin modellenmesi hedeflenmektedir. Modelleme yaklaşımında üç temel model grubu incelenebilir; Gizli sınıf modelleri, basit quasi simetrik uyum modelleri ve karma modeller (Schuster, 2002).

Bu çalışmada kategorik uyum verilerinin gizli sınıf analizi ile değerlendirmesi verilmektedir. Bu yaklaşım direk olarak diagnostik hassaslıkla ilgili parametreler cinsinden uyumu formülüle eder ve bireysel değerlendirmelerin hassaslıklarının tahminlenmesi ve çoklu değerlendirmeleri birarada ele alarak hassaslığın artırılması gibi pek çok pratik uygulamalara da olanak tanır. GSA diagnostik uyum verilerinin analizinde güçlü ve esnek bir yöntem olup, kullanımı oldukça pratik bir yöntemdir. Diagnostik uyum verilerinin kesikli modelleri GSA'nın özel bir durumu olarak ele alınabilir (Lazarsfeld ve Henry, 1968; Goodman L. A., 1974; Haberman, 1979). Gizli sınıf modelleri

hastalıklara ait semptom verilerinin analizinde de oldukça sık uygulanmaktadır (Uebersax ve Grove, 1990).

Uyum verileri ile ilgili literatürde yer alan daha önceki çalışmalarda daha çok kappa ve benzeri uyum indisleri ve bu indislerle ilgili çalışmalar veya varyans bölümüne dayalı yöntemlerle ilgili çalışmalar yer almaktadır. Fakat bu yöntemlerde verilen tanımlar ile bir bulgunun varlığı veya yokluğu yönünde ayırım yapılamamaktadır ve farklı tabanlı çalışmaları içeren örnekleri karşılaştırmada uygun yöntemler değildir. Buradaki temel amaç ise, gözlemciler tarafından konulan tanımların hassaslığı ile ilgili bilgiyi uyum verilerinden nasıl elde edilebileceği ve bu bilginin en etkili bir şekilde nasıl kullanılacağı sorularının cevaplanmasıdır. Bu ise duyarlılık ve seçicilik gibi uyum verilerinin, tanımların hassaslığı ile ilgili olasılıksal ölçütler cinsinden parametrelendirilmesi ile gerçekleşmektedir. Genel olarak, varyans bölümüne dayalı istatistiksel yöntemler ve uyum indisleri bunu gerçekleştirmez ve böylece tanı hassaslığı ile ilgili direkt olarak çıkarsamalara da olanak vermez.

GSA ile her bir gizli sınıfta yer alan vakaların oranı ve her bir gizli sınıfta yer alan vakaların gözlemciler veya tanı testleri ile öngörülen düzeye ait olma olasılıkları belirlenmektedir. Gizli sınıf modeli oluşturulduğunda elde edilecek bu parametreler gözlemcilerarası uyumun değerlendirilmesinde, tanı testlerinin performanslarının değerlendirilmesinde veya diğer pek çok uygulamalarda kullanılabilir.

1.3. Gizli Sınıf Modelleri

Diagnostik uyum çalışmalarında gizli yapı modellerinin ve ilgili modellerin kullanımı hızla artmaktadır. Gizli sınıf modelleri, kategorik değişkenler arasındaki birleşik dağılımı, bileşenlerinin değişkenler arasındaki bağımsızlığı sağladığı, dağılımların karmaşıklık olarak tanımlar. Goodman (1974) tarafından en çok olabilirlik tahminlerinin gizli sınıf modellerin parametrelerinin tahminlenmesinde kullanımını ortaya koymasından sonra gizli sınıf modellerin uygulama alanları oldukça artmıştır (Agresti ve Lang, 1993).

Model yaklaşımlarından GSA verinin iki veya daha fazla gizli sınıfta yer aldığını varsayar. Gizli sınıf modelleri ikili, nominal ve bazı durumlarda en az üç gözlemcinin olması durumunda sıralı veriler için uygundur. En basit durumda ki sınıf bulunmaktadır; gerçek pozitif ve gerçek negatif durumlar. Gözlemcilerin gözlemleri ayırt edebilme kabiliyeti pozitif gözlemleri pozitif, negatif gözlemleri negatif olarak sınıflandırma olasılığı ile tanımlanabilir. Gizli sınıf modeli doğru durum (gizli sınıf) üzerinde kararların şartlı olarak bağımsız olduklarını varsayar. Bu modeller sadece uyumu belirlemek değil, aynı zamanda uyumsuzluğun kaynağını da anlamak amacı ile kullanılabilir. Araştırmacılar uyumu genellikle gözlemcilerin kalitesi ve tutarlılığını ortaya

koymak amacı ile sayısallaştırırlar. Gözlemcilerin birbirlerinden bağımsız oldukları varsayılarak, kategoriler arasında uygun bir şekilde ayrışma yapma üzerinde uyum göstereceklerdir. Daha iyi ayrıştırabilen gözlemciler daha iyi uyum sağlayacaklardır (Hripcsak, 2002).

Gözlemlerin iki veya daha fazla gizli sınıfa ait oldukları varsayılır. Gizli sınıf en basit şekilde, bir gözlemin sınıf üyeliğinin direk olarak gözlenememesi anlamına gelmektedir. Bir gözleme bir yanıt yapısının atanması olasılığı o gözlemin içerisinde bulunduğu gizli sınıfa bağlıdır. Örneğin, iki gizli sınıf olabilir; hastalık pozitif gözlemlerin sınıfı ve hastalık negatif gözlemlerin sınıfı. Bir gözlemin pozitif veya negatif olarak bir gözlemci tarafından tanı konulması, o gözlemin hangi sınıfa dahil olduğuna bağlı olacaktır. Gizli sınıf analizi, her bir gizli sınıftaki gözlemlerin oranının ve her bir gizli sınıf için, bu sınıfta yer alan gözlemlerin gözlemciler tarafından öngörülen yanıt yapısına sahip olma olasılıklarının tahmin edilmesini amaçlar. Gizli sınıf modeli oluşturulduğunda, bu parametreler gözlemcilerin verdikleri yanıtları değerlendirmede kullanılabilir ve aynı zamanda değerlendirmelerin hassaslığı hakkında bilgi sağlar veya başka pratik kullanımlarda yer alabilir (Uebersax, 2001).

Gizli yapı modellerinde bir uyum veya uyumsuzluk yapısını ortaya koymak için gerekli olan hassaslık düzeyi nedir sorusuna cevap aranmaktadır. Her bir tan koyucunun hassaslığı ile değişen hastalık düzeylerine göre prevalans değerlerini gözlenemeyen parametreler olarak ele alır ve gözlenen veriye dayalı olarak bu parametreler tahminlenir. Tam tanımlı bir standardın olmadığı durumda, sınıflandırma işlemi sırasında hassaslığın değerlendirilmesinde, kesikli ölçümler sadece rölatif ifadelerle değerlendirilebilir. Eğer bu kesikli süreci tanımlayan kuvvetli varsayımlar yapılabılırsa, gizli sınıf modelleri gözlemcilerarası uyum/uyumsuzluk'un değerlendirilmesine imkan verir. Aynı zamanda ortak bir prevalans değeri de elde edilebilir. Bu hipotetik standart gizli faktör olarak hizmet eder. Kesikli veriler tamamlanmamış olarak ele alınır ve log lineer modeller gizli sınıf modelleri ve bu modellerin uzantılarını parametrelendirme de kullanılabilir (Espeland ve Handelman, 1989).

Standart gizli sınıf modeli, incelenilen olayın belirleyicileri olarak kabul edilen kategorik gözlenen değişkenler yardımıyla bir veya daha fazla kategorik gizli gözlenemeyen değişkenlerin ölçülmesi için oluşturulmuş bir modeldir (Hagenaars ve ark., 2002). Gizli değişkenler ve gözlenen değişkenler nominal kabul edilir. Bu modeller ilk olarak Lazarsfeld ve Henry (1968) tarafından ortaya konulmuş olup Goodman (1974a), Haberman (1979) ve Clogg (1981)'un çalışmaları ile kullanımı artmıştır (Heinen, 1996). Gizli sınıf olasılıkları ve koşullu yanıt olasılıkları GSA'nın iki temel parametresidir. Gizli sınıflar, bir değişken setine göre bireyleri yapılandıran boyutlardır. Bütün gizli sınıflar kontrol altına alındığında sadece değişkenler arasındaki rasgele ilişki mevcut olacaktır. GSA, gözlemleri, denekleri,

vakaları, bireyleri birbirlerine koşullu bağımsız olan gizli sınıflara ayırmaktadır. Bu ise ilgilenilen değişkenlerin herhangi bir sınıf içerisinde birbirleri ile ilişkili olmamaları anlamına gelmektedir. Sınıflar gizlidir. Çünkü sınıflar doğrudan gözlenememekte, gözlenen belirleyiciler setinden oluşan bir fonksiyona dayanarak belirlenmektedir. Gizli sınıflar koşullu olasılıkla belirlenmektedir. Şöyle ki; her bir gizli sınıfta, her bir değişken diğer değişkenlerle istatistiksel olarak bağımsızdır. Gizli sınıf içerisinde bir semptomun varlığının ya da yokluğunun diğerlerinden bağımsız olarak ele alınması örnek olarak verilebilir. Bir başka deyişle; gizli sınıflar öyle tanımlanmaktadır ki, eğer veride gizli sınıf üyeliği yok edilirse sadece rasgelelik mevcut olacaktır. Gizli sınıf olasılıklarının yorumlanmasında her bir sınıfın büyüklüğü de önem taşımaktadır. Gizli sınıf olasılıklarının göreceli büyüklükleri, evrenin bu sınıflar içerisinde eşit bir şekilde dağılıp dağılmadığını veya bazı sınıflarda yığılmalar olurken bazı sınıflarda daha az kümelenmelerin olup olmadığını gösterir. Bütün sınıflar üzerinden gizli sınıf olasılıklarının toplamı bire eşittir. Koşullu yanıt olasılıkları, gizli değişken X'in t'inci sınıfında bulunan bir kişinin gözlenen değişkenlerin belirli bir düzeyinde bulunma olasılıklarıdır. Koşullu yanıt olasılıkları ile her bir maddenin, değişkenin veya gizli sınıfın kombinasyonunun olasılığından bahsedilmektedir. Koşullu olasılıklar her bir gizli sınıf tarafından tanımlanan yapının karakterize edilmesine yardımcı olur. Her bir gizli sınıf içerisinde, koşullu olasılıklar t'inci gizli sınıf içerisindeki gözlemlerin her bir gözlemlenen değişken ile benzer veya benzer olmayan karakteristiklere sahip olup olmadıklarını belirler. Bu nedenden dolayı koşullu yanıt olasılıkları faktör analizindeki faktör skorlarına karşılık gelmektedir.

GSA'de model parametrelerinin kestiriminde gözlemlenmiş veri kullanılır. Temel parametreler gizli sınıf prevalans değerleri ve yanıt olasılıkları olup deneklerin açıklayıcı değişkenlerle göstermiş olduğu çapraz tablolardaki gözlemlenen frekans değerlerinden kestirilmektedir. Kısıtlanmamış gizli sınıf modellerinde parametrelerin en çok olabilirlik kestirimi değişik şekillerde gerçekleştirilmektedir. En sık kullanılan yöntemler Haberman tarafından ortaya konulan Newton-Raphson algoritması, yine Haberman tarafından oluşturulan skorlama algoritması ve Dempster, Laird ve Rubin ile Goodman tarafından önerilen E-M algoritmasıdır. Goodman tarafından ortaya konulan daha genel gizli sınıf modellerin en çok olabilir kestirimlerinin elde edildiği yöntem ilk yöntemlerden biraz farklıdır. Basit, adımsal, oransal uyum yönteminin en çok olabilirlik parametrelerinin elde edilmesinde kullanılabileceğini göstermiştir. Bu yöntem E-M algoritması olarak adlandırılmıştır. Goodman tarafından tanımlanan en çok olabilirlik yöntemine göre parametre kestirimi gerçekleştirilen E-M algoritmasında, bazı matematiksel ölçütleri optimize edecek parametre değerlerinin elde edilmesinde en çok olabilirlik kestirimi kullanılmaktadır.

Diğer gizli sınıf gözlemcilerarası güvenilirlik modellerinin özel durumları, Guttman ölçeklemede ve doğrusal hiyerarşik modellerden çıkarılabilir. Guttman ölçeklemede, yanıt hatalarına olanak tanınmaz. Bmyle bir modelde, gözlemcilerin tam uyum içerisinde olduğu durum söz konusudur. Guttman modelin değişik bir durumu Goodman (1975) tarafından incelenmiştir. Bu model gizli sınıf modellerinin quazi bağımsızlık modeli olarak ele alınmaktadır. Eğer atama kategorileri arasında örneğin, pozitif, nötr ve negatif gibi bir hiyerarşi söz konusu ise, oluşturulacak gizli sınıflarda bu hiyerarşiye uyacaktır. (William R. Dillion, 1984).

Ölçümlerin güvenilirliğinin değerlendirilmesi direk olarak gizli sınıf modelinin tahminlenmesi sonucu elde edilen tahminler ile gerçekleşmektedir. Klasik test teorisine dayanmamasına rağmen, burada verilecek olan güvenilirlik prosedürleri ile amaç, ölçümlerin ne kadar tutarlı ve ne kadar bağımlı olduğunun ortaya konması ve ölçüm hataları ile maddeler arasındaki iç tutarlılık incelenerek, incelenen karakteristiğinin güvenilir belirleyicilerin ortaya konmasıdır. Her bir gözlemciye ve gözlemcilerin tamamına ait güvenilirliğin tahminlenmesi için bir yöntem verilecektir. Burada verilecek yöntemlerle araştırmacılar, hangi sınıftaki gözlemlerin daha iyi ölçüldüğünü hangilerinin ölçülemediğini ve hangi değişkenlerin iyi performans verdiğini hangilerinin veremediğini, sınıfları belirlemede ve ölçmede hangi maddelerin eklenmesi hangi maddelerin çıkarılması gerektiği sorularına ait cevapları bulabilecektir (Flaherty, 2002).

1.4 Olasılıksal Model

J tane gözlemcinin n tane gözlemi C tane kategoriye veya sınıfa ataması durumu ele alındığında; Gözlemcilerin yapmış oldukları atamaların çapraz sınıflaması C^J tane hücreli olan J-boyutlu bir tablo oluşturacaktır. Bu tablonun her bir hücresi $U_s, s= 1,2,\dots, C^J$ ile gösterilecektir. Herhangi bir U_s , her bir gözlemcinin yapmış olduğu yanıt düzenliğini temsil edecektir. Örneğin; 3 gözlemci ve 3 kategorinin olması durumunda, 3-boyutlu tablo da 27 hücre bulunacaktır. Böyle bir durumda, yanıt vektörü örneğin, $U_s = (1,1,2)$ birinci gözlemcinin ve ikinci gözlemcinin 1. kategoriye, 3. gözlemcinin ise 2. kategoriye atama yaptığı durumu ifade etmektedir. Bu hücrede yer alan sayı ise böyle bir yanıt yapısının toplam sayısını verecektir. 27 hücreli bu tablodaki $U_1 = (1,1,1)$, $U_{14} = (2,2,2)$ ve $U_{27} = (3,3,3)$ hücreleri tam uyumu temsil eden hücreler iken, $U_2 = (1,1,2)$ hücresi kısmi uyumu, $U_6 = (1,2,3)$ hücresi ise uyum olmadığını ifade etmektedir. V_t , tüm yanıt düzenleri içerisinde, doğru, tam uyum olan sınıfları temsil etmektedir. $V_1 = (1,1,1)$, $V_2 = (2,2,2)$, $V_3 = (3,3,3)$. Bir veya daha fazla gözlemcinin yapmış oldukları hatalı atamalar sonucu manifest yanıt vektörü U_s ile gizli V_t yanıt vektörlerinden herhangi birisi ile birebir uyumlayabilir. Aşağıda verilecek olan olasıksal modelin temel prensibi, gözlenen yanıt vektörünün lokal bağımsızlık kuralına

uyan yanıt hatalarının sonucu olarak gizli yanıt yapılarından ortaya çıktığı yönündedir. Olasılıklar için P(.) notasyonunu kullanarak, model formül (1) ile verilebilir (Dillion ve Mulani, 1984).

$$P(U_s) = \sum_{t=1}^T P(U_s | v_t) \cdot \theta_t \quad (1)$$

Her bir gizli sınıf için olasılıklar θ_t ile gösterilmektedir ve toplamları da biri vermektedir.

$$\sum_{t=1}^T \theta_t = 1 \quad (2)$$

Şartlı olasılıklar $P(U_s | v_t)$ recruitment olasılıklar olarak adlandırılmaktadır. Çünkü bu olasılıklar, gözlenen yanıt vektörünü gizli sınıflara bağlamaktadır. Bu şartlı olasılıklar, hata olasılıkları olarak formül (3) de ki gibi ifade edilebilirler.

$$P(U_s | v_t) = \prod_{j=1}^J \gamma_{j,t}^{a_j} (1 - \gamma_{j,t})^{1-a_j} \quad (3)$$

$\gamma_{j,t}$ = t ninci gizli sınıf için j ninci gözlemcinin hatalı bir atama yapma olasılığı

$(1 - \gamma_{j,t})$ = t ninci gizli sınıf için j ninci gözlemcinin doğru atama yapma olasılığı.

$a_j = 1$ eğer U_s yanıt vektöründeki j ninci gözlem v_t vektöründeki j ninci gözlem ile uyum sağlamaz ise;

$$(U_{sj} \neq v_{tj})$$

$a_j = 0$, eğer U_s yanıt vektöründeki j ninci gözlem v_t vektöründeki j ninci gözlem ile uyum sağlar ise

$$(U_{sj} = v_{tj})$$

Sonuç olarak, gözlemcilerin yanıtlarına ait herhangi bir gözlenen yanıt yapısının olasılığı, θ_i ağırlıkları kullanarak tam uyum sınıfı verildiğinde, şartlı olasılıkların ağırlıklı toplamlarına eşit olacaktır. Şartlı olasılıklar her bir gözlemciye ait hata olasılıklarını ifade etmektedir. Örneğin; J=3 gözlemci ve C=3 kategori ve gözlenen yanıt vektörü $U_s = (1,1,2)$ olması durumunda, Birinci tam uyum sınıfı $V_1 = (1,1,1)$ için, üçüncü gözlemcinin hata yapması durumunda, böyle bir yanıt yapısına ait şartlı olasılık aşağıdaki gibi verilecektir.

$$\gamma_{1,1}^0(1-\gamma_{1,1})\gamma_{2,1}^0(1-\gamma_{2,1})\gamma_{3,1}(1-\gamma_{3,1})^0 = (1-\gamma_{1,1})(1-\gamma_{2,1})\gamma_{3,1}$$

İkinci tam uyum sınıfı $V_2 = (2,2,2)$ için, ilk iki gözlemcinin hatalı sınıflandırması bulunmaktadır ve, şartlı olasılık aşağıdaki gibi olacaktır.

$$\gamma_{1,2}(1-\gamma_{1,2})^0\gamma_{2,2}(1-\gamma_{2,2})^0\gamma_{3,2}^0(1-\gamma_{3,2}) = \gamma_{1,2}\gamma_{2,2}(1-\gamma_{3,2})$$

Üçüncü tam uyum sınıfı $V_3 = (3,3,3)$ için, tüm gözlemcilerin hatalı sınıflandırması bulunmaktadır ve, şartlı olasılık aşağıdaki gibi olacaktır.

$$\gamma_{1,3}(1-\gamma_{1,3})^0\gamma_{2,3}(1-\gamma_{2,3})^0\gamma_{3,3}(1-\gamma_{3,3})^0 = \gamma_{1,3}\gamma_{2,3}\gamma_{3,3}$$

gözlenen yanıt yapısı $U_s = (1,1,2)$ için şartlı olmayan olasılık değeri, gizli sınıf olasılıklarını ağırlık değerleri alarak, uygun şartlı olasılıkların ağırlıklı toplamları şeklinde elde edilebilir. Böylece,

$$\begin{aligned} P(U_s = 1,1,2) &= P(U_s = 1,1,2 | v_1 = 1,1,1) \theta_1 + \\ &P(U_s = 1,1,2 | v_2 = 2,2,2) \theta_2 + \\ &P(U_s = 1,1,2 | v_3 = 3,3,3) \theta_3 \\ &= (1-\gamma_{1,1})(1-\gamma_{2,1})\gamma_{3,1} \theta_1 + \gamma_{1,2}\gamma_{2,2}(1-\gamma_{3,2})\theta_2 + \gamma_{1,3}\gamma_{2,3}\gamma_{3,3} \theta_3 \end{aligned}$$

olacaktır.

1.5 Genel Modelin Kısıtlanmış Durumu

Formül (1) ve Formül (3)'de verilen modelde, şartlı olasılıklar üzerinde hiç bir kısıtlama bulunmamaktadır. Gözlemcilerarası güvenilirlik çalışılırken, modelin kısıtlanmış formunun dikkate alınması gerekebilir. Tahminlemede ihtiyaç duyulan parametrelerin sayısının azaltılması açısından, verilen yanıtların hata yapılarını dikkate alan çeşitli hipotezlerin ortaya konulması açısından, kısıtlanmış modeller bir araç olabilir. Örneğin, gözlemcilerin hatalı bir değerlendirme yapma olasılıklarının her bir gözlemci için farklı, fakat her bir gözlemcinin gizli sınıflar boyunca hatalı değerlendirme yapma olasılıklarının aynı olması durumunda,

$$\gamma_{1.1} = \gamma_{1.2} = \gamma_{1.3} = \gamma_1, \gamma_{2.1} = \gamma_{2.2} = \gamma_{2.3} = \gamma_2, \gamma_{3.1} = \gamma_{3.2} = \gamma_{3.3} = \gamma_3$$

olacaktır. Genel model bu durumda formül (4) ile verilecektir.

$$P(U_s) = \sum_{t=1}^T \left[\theta_t \cdot \left(\prod_{j=1}^J \gamma_j^{a_j} (1 - \gamma_j)^{1 - a_j} \right) \right] \quad (4)$$

Eğer her bir gözlemcinin hatalı değerlendirme yapma olasılıklarının aynı olması durumunda

$$\gamma_{1.1} = \gamma_{1.2} = \gamma_{1.3} = \gamma_{2.1} = \gamma_{2.2} = \gamma_{2.3} = \gamma_{3.1} = \gamma_{3.2} = \gamma_{3.3} = \gamma$$

model daha basitleşecek ve formül (5) de verilen hali alacaktır.

$$P(U_s) = \sum_{t=1}^T \left[\theta_t \cdot \gamma^{a_j} (1 - \gamma)^{J - a_j} \right] \quad (5)$$

a_j , gözlemcilerin gizli sınıf vektörü v_t için verdikleri yanıt düzenlerindeki, U_s , toplam hata sayısı. Genel olarak böyle bir kısıtlama tahminlemede gerekli olan parametre sayısının azalmasını sağlayacaktır.

1.6 Parametre Tahmini

$$n = \sum_{s=1}^{C^J} n_s \text{ yanıtları için olabilirlik, } L, \text{ formül (6) ile verilmektedir.}$$

$$L = \sum_{s=1}^{n_s} P(U_s) \quad (6)$$

U_s , n_s tane yanıt içeren gözlenen yanıt vektörüdür. Şartsız maksimum olabilirlik parametre tahminleri, bir seri kısmi türevlerin çözümü ile elde edilebilir; formül (7)

$$\frac{\partial \ln L}{\partial \theta_t}, t = 1, 2, \dots, T \quad \text{ve} \quad \frac{\partial \ln L}{\partial \gamma_{j,t}} = 0, j = 1, 2, \dots, J; t = 1, 2, \dots, T \quad (7)$$

Bu kısmi türevler, Fisher algoritması E-M algoritması veya benzer birtakım Newthon Raphson gibi adimsal algoritmalar ile çözümlenebilir (Çolak M, 2003)

1.7 Olasılık Modelinin Gizli Sınıf Modeline Dönüşümü

Formül (1) ve formül (3)'de gösterilen olasılıksal model gizli sınıf modelidir. Formül (4) ve formül (5)'de verilen kısıtlanmış model ise, kısıtlanmış gizli sınıf modele karşılık gelir. A ve B 'nin iki ilişkili kategorik değişken olduğu durumda, sisteme yeni bir X kategorik değişkenin eklenmesi durumunda, X'in her bir düzeyinde A ve B değişkenleri ilişkisiz hale geliyorlar ise, X değişkeninin A ve B değişkenleri arasındaki ilişkiyi açıkladığı söylenebilir. GSA ile böylesi bir gözlenemeyen gizli X değişkeninin ortaya konulması amaçlanır. Faktör analizinde faktörlerin sürekli gözlenen değişkenleri arasındaki yapıyı ortaya koyması ile benzer bir durumdur. Genel olarak pek çok düzeyi olabilen tek bir gizli değişken X olması durumu ele alınmasına karşın, oluşturulan modeldeki parametreler üzerinde bir takım kısıtlar getirilerek, X değişkenin birden fazla gizli değişkeni temsil etme durumu da söz konusu olabilir.

Gizli sınıf modelin iki tür parametresi vardır bunlar, her bir gizli sınıfta bulunmanın şartsız olasılıkları ve bir sınıftaki üyeliğin verilmesi durumunda, bir yanıt yapısının gözlenme şartlı olasılığı. Goodman (1974) tarafından verilen notasyona göre, şartsız olasılıklar π_1^X, π_2^X, \dots , ile gösterilmektedir. π_t^X , herhangi bir gözlemin gizli değişken X'in t'inci kategorisinde yer alma olasılığıdır. Şartlı olasılıklar π_{it}^{AX} , ile verilmektedir ve gizli sınıf X'in t'inci kategorisinde yer alan bir gözlemin A maddesine i yanıtını verme şartlı olasılığını tanımlamaktadır. Bu parametrelerden A, B, C maddeleri için herhangi bir yanıt düzeneğinin elde edilme olasılığı, her bir sınıftaki

gözlemler için maddelere verilen yanıtların bağımsız olduğu varsayımı ile hesaplanabilmektedir. Şartlı bağımsızlık, kısıtlayıcı bir varsayım olup, gizli değişken X 'in verideki gözlenen ilişkiyi ortaya koymasını anlatır. Faktör analizindeki terminolojide, faktörlerin tamamı ile gözlenen kovaryansları açıklamasına karşılık gelir.

Goodman (1974), parametrelerin en çok olabilirlik yöntemleri ile nasıl tahminlenebileceğini göstermiştir. Latent GOLD (Vermunt, Magidson, 2000) ve benzeri programlar ile parametreler tahminlenebilmektedir. Olasılıksal modelin gizli sınıf modeline dönüştürülmesi oldukça açıktır. $J=3$ gözlemcinin ve $C=3$ kategorinin olması durumu ve formül (5)'de verilen kısıtlı model ele alınırsa, aşağıda verilecek olan eşitlik kısıtları ile gizli sınıf modeli tahminlenebilir. J_1, J_2, J_3 3 gözlemciye ait kararları ifade etmektedir. Aşağıda verilen eşitlik kısıtları ile birlikte gizli sınıf modeli paket programlar yardımı ile tahminlenebilir.

$$\pi_{11}^{J_1 X} = \pi_{11}^{J_2 X} = \pi_{11}^{J_3 X} =$$

$$\pi_{22}^{J_1 X} = \pi_{22}^{J_2 X} = \pi_{33}^{J_3 X} =$$

$$\pi_{33}^{J_1 X} = \pi_{33}^{J_2 X} = \pi_{33}^{J_3 X}$$

Birinci tam uyum sınıfı ele alındığında, doğru sınıfa atama sadece birinci kategoriye atama gerçekleştiğinde olacaktır. Bu nedenle, ikinci veya üçüncü sınıfa atama yapıldığında, $(1 - \pi_{11}^{J_1 X}) = (1 - \pi_{11}^{J_2 X}) = (1 - \pi_{11}^{J_3 X}) = \gamma$ hata olasılıkları gerçekleşecektir. Benzer şekilde diğer tam uyum gerçekleşen sınıflar içinde gözlemcilerle ait hata olasılıkları gerçekleşecektir. Her bir gözlemci farklı hata olasılık değerlerine sahiptir. Fakat, her bir gözlemciye ait hata değerleri gizli sınıflar boyunca aynıdır.

Diğer gizli sınıf gözlemcilerarası güvenilirlik modellerinin özel durumları, Guttman ölçeklemede ve doğrusal hiyerarşik modellerden çıkarılabilir. Guttman ölçeklemede, yanıt hatalarına olanak tanınmaz. Bmyle bir modelde, gözlemcilerin tam uyum içerisinde olduğu durum söz konusudur. Guttman modelin değişik bir durumu Goodman (1975) tarafından incelenmiştir. Bu model gizli sınıf modellerinin quazi bağımsızlık modeli olarak ele alınmaktadır. Eğer atama kategorileri arasında örneğin, pozitif, nötr ve negatif gibi bir hiyerarşi söz konusu ise, oluşturulacak gizli sınıflarda bu hiyerarşiye uyacaktır.

1.8 Gizli Sınıf Üyeliklerinin Kestirilmesi ve Gizli Sınıflara Atamaların Yapılması

Gizli sınıf modeli kestirildikten sonra, "recruitment" olasılıklar yardımıyla, gözlemler en olası gizli sınıfa atanarak sınıflandırılırlar. "Recruitment" olasılık; gizli sınıfın rastgele seçilmiş bir gözlemi için verilen yanıt düzeyinin gözlemlenme olasılığıdır. "Recruitment" olasılıklar, kestirimi koşullu yanıt olasılıklarından hesaplanmaktadır (Lazersfeld ve Henry, 1968). "Recruitment" olasılıklardan, her bir gizli sınıfın prevalans değeri ile, her bir sınıftaki gözlemlerin sonsal üyelik olasılıkları hesaplanabilmektedir. Daha sonra, en yüksek sonsal olasılıklara sahip gözlemler, birimler, vakalar gizli sınıflara atanmaktadır (model ataması, model belirlenmesi) veya sınıflamayı bulanık olarak (fuzzy) bırakarak derecelendirmeye yönelik üyelik fonksiyonu tanımlanmaktadır.

3 gözlemci ve bir gizli değişkenli model durumunda,

$$\pi_{ijkt}^{ABCX} = \pi_{it}^{\bar{A}X} \times \pi_{jt}^{\bar{B}X} \times \pi_{kt}^{\bar{C}X} \times \pi_t^X \text{ formülünün } \lambda_{ijk} = \sum_t \lambda_{ijkt}^{ABCX}$$

formülüne konulması ile i, j, k düzeylerindeki bir gözlemin, deneğin, yanıtlayıcının gizli değişken X 'in t 'inci düzeyinde bulunması olasılığı elde edilir.

$$P(X = t | ABC) = \pi_{ijkt}^{ABC\bar{X}} = \pi_{ijkt}^{ABCX} / \sum_t \pi_{ijkt}^{ABCX} \quad (8)$$

Formül 8 en büyük koşullu olasılık $\pi_{ijkt}^{ABC\bar{X}}$ değerine sahip olan gözlemlerin gizli sınıflara atanmasına olanak tanır.

İlkolarak, $\pi_{ijkt}^{ABCX} = \pi_{it}^{\bar{A}X} \times \pi_{jt}^{\bar{B}X} \times \pi_{kt}^{\bar{C}X} \times \pi_t^X$, formülünün yardımıyla, π_{ijkt}^{ABCX} olasılıkları her bir gizli sınıf için hesaplanır. Bu kestirimler kullanılarak, gözlemlerin her bir gizli sınıfa atanmasındaki koşullu olasılıklar elde edilmektedir. Daha sonra maksimum değeri veren sınıfa gözlemlerin ataması yapılır. Paket programlarda model uyumu testi yapılabilmekte ve gizli sınıflara atamalar yapılabilmektedir. Gizli sınıflara atamaların yapılması olasılıksal olarak gerçekleştirilir. Bu durum yapılacak işlemlerde hataların olabileceğini de beraberinde getirmektedir. Clogg (1979; 1981), bu hataların

kestiriminde iki yöntem önermiştir. Doğru bir şekilde sınıflandırmanın yüzdesi Formül (9) ile verilmektedir;

$$100 \times \sum_{ijk} \left(\pi_{ijkt}^{ABC\bar{X}} \times p_{ijk} \right) \quad (9)$$

$\pi_{ijkt}^{ABC\bar{X}}$ hücre için model olasılığı, p_{ijk} ise hücredeki evren yüzdesidir. Goodman ve Kruskal (1954) tüm gözlemlerin en çok olasılık değeri ile elde edilen gizli sınıfa atanmasından kaynaklanan hata oranı değeri λ 'yı Formül 10 ile tanımlamaktadır; λ , gizli değişken X ile birleşik değişken (ABC) arasındaki ilişkinin yakınlığını gösterir

$$I = \frac{E_1 - E_2}{E_1}, \quad E_1 = 1 - p_t^X, \quad E_2 = \sum_{ijk} \left(1 - p_{ijkt}^{ABC\bar{X}} \right) p_{ijk} \quad (10)$$

1.9 Model Uyumu ve Uyum Testi

Modelin tanımlı olup olmadığını kontrol etmenin kolay bir yolu, her bir parametre için değişik başlangıç modelleri ile modelin kestirilmesi. Eğer analizler sonucunda parametre kestirimlerinin değerleri her seferinde farklı elde ediliyor, fakat kestirilen frekans değerleri ile model uyum katsayısı ki-kare değeri aynı elde ediliyorsa model tanımlı değildir. Bununla birlikte bu durum, GSA'de parametrelerin tek bir şekilde belirlenmemesi nedeni ile gerekli bir koşul değildir. Goodman (1974b) lokal tanımlılık için yeterli koşulu oluşturmuştur. Lokal tanımlılığı kontrol etmek için satırlarında model tarafından tanımlanan yanıt olasılıklarının, sütunlarında ise model parametrelerinin yer aldığı Hessian matrisi oluşturulur. Matrisin hücrelerinde model parametrelerine göre alınan birinci türevler yer almaktadır. Birinci türevlerin örnek kestirimleri, formüllerde birinci türevler için en çok olabilirlik kestirimlerinin yerleştirilmesi ile elde edilir. Oluşturulan matrisin rank değerinin hesaplanması ile lokal tanımlılık için test gerçekleştirilebilmektedir. Modelin lokal tanımlı olabilmesi için matrisin tam rank olması gerekmektedir. Lokal tanımlılık aynı zamanda beklenen enformasyon matrisinin kestirimi kullanılarak da oluşturulabilir. Dayton ve Mcready (1980), teorik varyans kovaryans matrisinin tam rank olması gerektiğini belirtmiştir. Dolayısıyla, kestirilen asimptotik varyans kovaryans matrisinin rank değeri ile lokal tanımlılık test edilebilmektedir. Forman (1985), beklenen ikinci dereceden kısmi türev matrisinin, olabilirlik fonksiyonunun lokal maksimuma ulaşabilmesi için sahip olması gerekli özellikleri tartışmıştır. Bu matrisin tüm

özdeğerlerinin negatif olması durumunda matrisin negatif tanımlı olacağını ve modelin de lokal tanımlı olacağını belirtmiştir. Tanımlı olmayan modeller, parametreler üzerine kısıtlar getirilerek tanımlı olabilmektedir. Modelin evren için geçerli olması yönündeki hipotez için olabilirlik oranı, asimptotik olarak bazı düzenleyici koşulların sağlanması durumunda, ki-kare dağılımı göstermektedir. Bu ise kısıtlanmamış gizli sınıf modeli için log lineer modellerdeki test işlemlerinin kullanılabilceğini göstermektedir (Heinen, 1996).

Gizli sınıf modellerin değerlendirilmesinde çeşitli model değerlendirme ölçütleri bulunmaktadır. Tüm bu ölçütler model altında beklenen değerlerin orijinal değerlerle ne derece uyum sağladığı esasına dayanmaktadır. Uygulamada bazı karşılaştırmalı çeşitli modeller ve bu kapsamda da çeşitli ölçütler bulunmaktadır. Bu ölçütlerden Pearson ki-kare (X^2) istatistiğine dayanan ki-kare önemlilik testi, olabilirlik ki-kare oranına (L^2) dayanan ki-kare önemlilik testi, Akaike enformasyon ölçütü (AIC) ve Bayes enformasyon ölçütüne (BIC) dayanan göreceli uyum ölçütü ve Rudas, Clogg, Lindsay (1994) tarafından geliştirilen model uyum indisi π^* oldukça sık kullanılan ölçütlerdendir. Bu ölçütler, kestirilen model parametreleri ile verilen hücrelere ait beklenen frekans değerleri ile (f_{ijkl}), gözlenen frekans değerlerinin (F_{ijkl}) karşılaştırmasına dayanmaktadır. Gözlenen frekans değerlerinden çok uzaklaşan beklenen değerler veren modeller kabul edilmez iken yakın değerler veren modeller kabul edilir. Genellikle daha fazla gizli sınıfın olduğu çok fazla parametre içeren modeller veriye daha iyi uyum sağlarlar.

1.10 Standart Hataların Hesaplanması

GSA'da varyans-kovaryans matrisinin kestirimi iki şekilde yapılabilir. Yöntemlerden birisi, beklenen enformasyon matrisinin negatifini almaktır. Diğer bir yöntem ise parametrelerin log olabilirlik değerinin ikinci türevini almaktır. Bu matrisin negatifinin tersi, varyans kovaryans matrisinin kestirimidir. Gözlenen enformasyon matrisi en çok olabilirlik kestirimleri için hesaplanan ikinci türevlerin negatifidir. Varyans kovaryans matrisinin kestirimindeki en sık kullanılan yöntem gözlenen veya beklenen enformasyon matrisini kullanarak ilave hesaplamaların yapılmasıdır. Her iki yöntem de asimptotik olarak hemen hemen aynı sonuçları vermektedir (Heinen, 1996). Bazı bilgisayar programlarında gizli sınıf oranları ve koşullu olasılıklar için standart hatalar hesaplanabilmektedir. GSA'da kestirilen standart hataların iki temel kullanımı bulunmaktadır. İlk olarak, parametre kestirimlerinin güven aralıklarının oluşturulmasında kullanılmaktadır. İkinci olarak ise, kestirilen standart hatalar kullanılarak önemlilik testleri gerçekleştirilmektedir. Verilen bir parametrik değerın sifıra eşit olup olmadığı

test edilebilir ve geniş örneklerde z testi uygulanabilir. $H_0: \pi_{11}^{\bar{A}X} = 0$

şeklinde olup, z istatistiği $z = \frac{\pi_{11}^{\bar{A}X}}{\sigma_h}$ ile verilmektedir. σ_h , kestirilen standart

hata olup, $\sigma_h = \pi_{11}^{\bar{A}X}$ ile verilmektedir. Kestirilen standart hataların hesaplanmasında kullanılan iki deneysel yöntem Jackknife ve Bootstrap yöntemleridir.

Gereç ve Yöntem

Tele-Radyoloji yönteminin potansiyel yararının ve etkililiğinin, gözlemcilerarası uyumun GSA yöntemi ile saptanarak ortaya konulması amacıyla yapılan bu araştırmada, 40 adet Postero Anterior Akciğer Direkt Grafisi 5.0 megapiksellik bir dijital fotoğraf makinesi yardımı ile dijital ortama aktarılmıştır. 4 bağımsız radyolog orjinal imajları ve JPEG formatında dijitalize edilmiş hallerini incelemiş ve Tüberkülozun (TB) 8 radyolojik göstergesinin mevcudiyetini ve buna dair görüşlerinin kesinlik derecesini farklı skorlarla değerlendirmiş, diğer patolojik bulguları raporlamıştır. Radyologlar ayrıca dijital imajlar ve orjinal grafiler arasında radyografik kalite açısından anlamlı bir fark olup olmadığını değerlendirmek üzere imajlar ve grafileri; Netlik, Kontrast, Pozisyon ve Santralizasyon kriterleri açısından da skorlamışlardır. Daha sonra bu skorlar ve raporların istatistiksel analizleri çıkarılmış, dijital ve orjinal imajların değerlendirilmesinde istatistiksel açıdan anlamlı bir farkın olup olmadığı değerlendirilmiştir. GSA ile bu çalışmada, Tele-Radyoloji uygulaması sonucu dört radyolog tarafından 8 farklı Tüberküloz (TB) bulgularının mevcudiyetine dair oluşturulan yanıt düzeyleri birarada değerlendirilerek, bulguların varlığını veya yokluğunu gösteren gizli sınıflar 8 farklı radyolojik bulgu için ayrı ayrı ortaya konulmuştur. GSA ile, gözlemcilerarası uyum, güvenilirlik, uyum karakteristikleri verilmiş, gözlemcilerarası uyum test edilmiş ve gözlemcilerle ait hata oranları ve standart hatalar hesaplanmıştır.

Başkent Üniversitesi Ankara Hastanesinin arşivinden (TB) tanısı konulmuş hastaların taranması sonucu elde edilen filmlerin içinden 40 adet ayakta Postero Anterior Akciğer grafisi seçilmiştir. Orijinal imajlar, Başkent Üniversitesi Ankara Hastanesi Radyoloji Bölümü'ne bağlı 2 uzman radyolog tarafından incelenmiş ve anket üzerinden raporlanmıştır. Gülhane Askeri Tıp Akademisi Haydarpaşa Hastanesine bağlı 2 uzman Radyolog ise, dijital fotoğraf makinesi sistemi ile elde edilen dijital imajları, oluşturulan web sayfası üzerinden incelemiş ve web sitesi üzerinde oluşturulan on-line anket ile raporlamıştır. On-line anket ile elde edilen veriler Başkent Üniversitesi

Bilgi İşlem Departmanına ait veritabanında depolanmış, çalışma sonunda veriler toplu halde elde edilerek SPSS veri formatına dönüştürülmüştür. Radyologlar oluşturulan anket ile grafilerde Tüberkülozun 8 adet radyolojik bulgusunun (İnfiltrasyon, Kavite formasyonu, Pnömotoraks, Plevral efüzyon, Kalsifikasyon, Skar dokusu, Lenfadenopati, Nodül veya kitle) mevcudiyetini 5'li likert ölçek ile skorlamışlardır. GSA uygulamalarında, analizlerde kolaylık sağlamak amacı ile skorlar üç grupta birleştirilmiştir. TB bulgularına ait değişkenlerin düzey sayısının üç olması nedeni ile, 3 sınıflı gizli sınıf modeli oluşturulmuştur. Bunlardan (+) sınıf, bulgunun varlığını gösteren sınıf, (0) sınıfı, bulgunun varlığı konusunda kararsız olan sınıfı, (-) sınıfı ise, bulgunun yokluğunu gösteren sınıfı tanımlamaktadır. İlk iki radyolog (A ve B), Başkent Üniversitesinde görevli olup, orijinal imajları değerlendirirken, diğer iki radyolog (C ve D), Tele-Radyoloji uygulaması ile dijital görüntüler üzerinden değerlendirmeleri gerçekleştirmişlerdir. $3^4 = 81$ olası yanıt düzeni olmasına karşın, gözlenemeyen yanıt düzenleri gizli sınıf analizinde değerlendirmeye alınmamıştır. GSA sonuçları Latent GOLD paket programı kullanılarak elde edilmiştir.

Bulgular

Radyologların dijital imajlar ve orijinal grafilere arasında radyografik kalite açısından anlamlı bir fark olup olmadığını değerlendirmek üzere yapmış oldukları; Netlik, Kontrast, Pozisyon ve Santralizasyon kriterlerinin 5'li Likert skalasına göre (1=Çok iyi, 2=İyi, 3=Vasat, 4=Kötü, 5=Çok kötü) ve kalitesi açısından "1" den (en kötü) "10" a (en iyi) kadar olan skor değerlendirmelerine ait tanımlayıcı istatistikler verilmiştir. Tüm radyologlar her bir kriter yönünden oldukça olumlu sonuçlar vermişlerdir. Bu kriterler yönünden değerlendirmenin olumlu olması Tele-Radyoloji yöntemin etkinliğinin bir göstergesi olarak düşünülebilir. Netlik skorlamasının amacı, orijinal grafilere ile, dijital edilmiş halleri arasında dönüşüm sürecinde hatanın olup olmadığının tespit edilmesidir. Kontrast birbirine komşu iki yapının yoğunluğu arasındaki fark olup, ne kadar yüksek olursa, film üzerinde komşu yapılar arasındaki geçişler o kadar belirgin olur. Akciğer grafilere çekilirken ki pozisyon standardındaki bozulmalar patolojilerin görülmemesine neden olabilir ve dijital grafilere elde edilirken fotoğraflama aşamasında yapılabilecek çekim hataları nedeni ile dokular arasındaki simetrik bozulabilir. Bu nedenle pozisyon değerlendirmesi de oldukça önemlidir. Santralizasyon, incelenmek istenen alanın grafinin sınırları içerisinde tamamen görüntüye alınması kavramıdır. Santralizasyondaki hatalar nedeni ile görüntü dışında kalan bir patoloji gözden kaçabilir. Fotoğraflama aşamasında bu tür artefaktlar oluşabilir.

Tablo 1. Filmin / İmajın Kalitesine Yönelik Değişkenlerin Tanımlayıcı İstatistikleri

Kalite	N	Ortalama	Std. Sapma	Minimum	Maksimum
Radyolog A	40	9,07	,80	6	10
Radyolog B	40	7,68	,97	6	9
Radyolog C	40	8,28	1,15	6	10
Radyolog D	40	7,25	1,28	4	9
Netlik	N	Ortalama	Std. Sapma	Minimum	Maksimum
Radyolog A	40	1.50	.72	1	4
Radyolog B	40	2.03	.28	1	3
Radyolog C	40	1.40	.59	1	3
Radyolog D	40	2.10	.71	1	3
Kontrast	N	Ortalama	Std. Sapma	Minimum	Maksimum
Radyolog A	40	1.45	.68	1	4
Radyolog B	40	2.35	.48	2	3
Radyolog C	40	1.38	.49	1	2
Radyolog D	40	2.20	.79	1	4
Pozisyon	N	Ortalama	Std. Sapma	Minimum	Maksimum
Radyolog A	40	1.63	.77	1	4
Radyolog B	40	2.70	.94	1	5
Radyolog C	40	1.62	.70	1	3
Radyolog D	40	2.08	.80	1	3
Santralizasyon	N	Ortalama	Std. Sapma	Minimum	Maksimum
Radyolog A	40	1.52	.72	1	4
Radyolog B	40	2.23	.62	1	4
Radyolog C	40	1.53	.60	1	3
Radyolog D	40	2.05	.78	1	4

Tüberkülozun 8 adet radyolojik bulgusunun (İnfiltrasyon, Kavite formasyonu, Pnömotoraks, Plevral efüzyon, Kalsifikasyon, Skar dokusu, Lenfadenopati, Nodül veya kitle) mevcudiyetinin likert ölçek ile skorlama sonuçlarının değerlendirmesine yönelik her bir değişken için oluşturulan 3 sınıflı gizli sınıf model sonuçları Tablo 2’de verilmektedir. GSA uygulamalarında, analizlerde kolaylık sağlamak amacı ile skorlar üç grupta birleştirilmiştir. L^2 , model uyum indisi’ne göre her bir değişken için oluşturulan gizli sınıf modellerinin uyumu oldukça iyidir. Ayrıca her bir değişkene göre gizli sınıf modeli sınıflandırmasında gözlemcilerle ait

değerlendirmelerin parametre tahminlerinin Wald istatistikleri hesaplanmış ve tüm değişkenlerde istatistiksel olarak önemli bulunmuştur.

Tablo 2. Değişkenlere Ait Gizli Sınıf Model Sonuçları

Değişkenler	L ²	X-kare	BIC _L ²	Serb.der
Efüzyon	18,99	32,53	-51,09	19
İnfüzyon	15,93	30,78	-54,16	19
Kalsifikasyon	9,82	7,55	-60,74	19
Kavite	3,89	2,88	-25,62	8
Lenfodönopati	22,98	81,07	-109,82	36
Nodül veya Kitle	16,53	217,08	-116,27	36
Pnömotoraks	2,46	1,76	-1,23	1
Skar Dokusu	14,93	23,20	-117,87	36

Tablo 3 –Tablo 10 arasında ki tablolarada, Latent GOLD Profile çıktıları yer almaktadır. Bu tablolarda gizli dağılım, gizli sınıf olasılıkları ve gizli sınıf şartlı olasılıkları, her bir değişken için parametrelerin standart hata değerleri ile beraber verilmiştir.

Tahmini gizli sınıf parametreleri ile radyologların gizli güvenilirlik değerleri elde edilebilir. Örneğin;Efüzyon değişkeni için model sonucunda, dört radyoloğu n değerlendirmeleri sonucunda model tarafından kesinlikle mevcut değil olarak sınıflandırılan gözlemler bulunmaktadır. Gözlemlerin % 61'i negatif sınıfta yer alırken, % 25'i pozitif sınıfta ve % 14'ü de kararsız sınıfta yer almıştır.

Negatif sınıfta yer alan gözlemlerin % 97'si 1. radyolog, % 99'u 2. radyolog, % 88'i 3. radyolog, % 99'u ise 4. radyolog tarafından da negatif olarak değerlendirilmiştir. Negatif sınıfta, 2. ve 3. radyologların gözlemleri gizli sınıflama ile daha yüksek uyum içerisindedir. Pozitif sınıfta ise, gözlemlerin % 86'sı 1. radyolog, % 88'i ikinci radyolog, % 97'si 3. radyolog ve % 98'i 4. radyolog tarafından pozitif olarak değerlendirilmiştir. Pozitif sınıflamada, 4. ve 3. radyologların uyumu daha yüksek elde edilmiştir (Tablo 3). Kararsız gözlemlerin yer aldığı 3. sınıfta ise, 3. ve 4. radyologların tanıları bulunmayıp, 1. ve 2. radyologlardan ise en yüksek olasılık 2. radyolog da elde edilmiştir (0,14).

Radyologlara ait hata değerleri olasılık tahminleri de uyumsuzluk bulunan sınıflandırmalardaki olasılıkların toplanması ile elde edilebilir. 1. radyolog için negatif sınıftaki hata olasılık tahmini (0,0147+0,0152= 0,03). 2.

radyolog için $(0,004+0,0086=0,009)$. 3. radyolog için $(0,12)$. 4. radyolog için $(0,005)$. Negatif sınıfta 3. radyoloğun hatalı sınıflandırma olasılığı daha fazladır. Benzer hata olasılıkları diğer sınıflar içinde hesaplanabilir. Radyologlara ait taminlenen genel hata oranı tüm sınıflar üzerinden ağırlıklı toplam alınarak hesaplanabilir. Örneğin, 1. radyolog için bireysel ağırlıklı ortalama hata oranı;

$$0,61*(0,0147+0,0152)+0,25*(0,0373+0,0998)+0,14*(0,17+0,79)=0,19$$

Benzer şekilde 2. radyolog için hata oranı;

$$0,61*(0,0004+0,0086)+0,25*(0,0374+0,0798)+0,14*(0,3774+0,443)=0,15$$

ve 3. radyolog için hata oranı benzer şekilde hesaplandığında, $0,08$, 4. radyolog için, $0,009$ olarak hesaplanabilir. En yüksek tahmini hata olasılık değeri 1. radyolog da elde edilmiştir. Tüm radyologlar için elde edilen hata oranları kabul edilebilir değerlerdedir.

Ayrıca, her bir kategoriye ait genel güvenilirlik değeri de verinin kalitesi ve özellikle de hataların yapısı açısından oldukça önemlidir. Her bir kategoriye ait güvenilirlik profile çıktılarında elde edilebilir. Örneğin; pozitif tam uyum sınıfında, tahmini toplam negatif hata olasılık değeri;

$$0,25* (0,0998 + 0,0374 + 0,0283 +0,0233) = 0,05,$$

tahmini toplam nötr sınıf hata olasılık değeri,

$$0,25* (0,0373 + 0,0798) = 0,03 \quad \text{olacaktır.}$$

Verilen tam uyum sınıfı için tahmini kategorilere atama güvenilirlik değerleri bu hata değerlerinden hesaplanabilir. Örneğin pozitif sınıf için, ağırlıklı güvenilirlik,

$$1- (0,05 +0,03) = 0,92 \quad \text{olacaktır.}$$

Benzer şekilde diğer kategoriler içinde ağırlıklı güvenilirlik değerleri hesaplanabilir. Değişkenlerin sayısının çok olması ve yorumların benzer olacağı düşüncesi nedeni ile diğer değişkenlere ait gizli sınıf modellemesi sonucuna göre elde edilen parametre tahminleri ve hata olasılık tahminleri yorumu verilmemiştir. İlgili tablolarda bu değerler görülmektedir.

Tablo 3. Efüzyon Değişkenine Ait Parametre Tahminleri

<i>Efüzyon Değişkeni</i>	Sınıf 1(-)	s.e.	Sınıf 2 (+)	s.e.	Sınıf 3 (0)	s.e.	
Gizli Sınıf Prevelans Değeri	0,6105	0,0921	0,2529	0,0687	0,1366	0,0751	
A	Pozitif	0,0147	0,0234	0,8629	0,1064	0,1669	0,1734
	Kararsız	0,0152	0,0169	0,0373	0,0387	0,0461	0,0489
	Negatif	0,9701	0,0341	0,0998	0,0924	0,7871	0,1919
B	Pozitif	0,0004	0,002	0,8827	0,1027	0,3774	0,255
	Kararsız	0,0086	0,0199	0,0798	0,0666	0,1796	0,1361
	Negatif	0,991	0,0218	0,0374	0,0576	0,443	0,2871
C	Pozitif	0,1196	0,0674	0,9717	0,0601	0,046	0,1361
	Negatif	0,8804	0,0674	0,0283	0,0601	0,954	0,1361
D	Pozitif	0,0045	0,0142	0,9767	0,0477	0,3682	0,2523
	Negatif	0,9955	0,0142	0,0233	0,0477	0,6318	0,2523

Tablo 4. İnfüzyon Değişkenine Ait Parametre Tahminleri

<i>İnfüzyon Değişkeni</i>	Sınıf 1(+)	s.e.	Sınıf 2 (-)	s.e.	Sınıf 3 (0)	s.e.	
Gizli Sınıf Prevelans Değeri	0,5954	0,1094	0,3501	0,0749	0,0545	0,0845	
A	Pozitif	0,891	0,0794	0,0079	0,0207	0,3076	0,5847
	Kararsız	0,0311	0,0312	0,0104	0,0155	0,0523	0,0571
	Negatif	0,0779	0,0699	0,9817	0,0337	0,6401	0,6077
B	Pozitif	0,9104	0,0642	0,0044	0,0129	0,5762	0,4493
	Kararsız	0,0581	0,0434	0,0224	0,0315	0,1391	0,1135
	Negatif	0,0315	0,0392	0,9733	0,0427	0,2848	0,4012
C	Pozitif	0,9613	0,0495	0,0835	0,0734	0,4294	0,7165
	Negatif	0,0387	0,0495	0,9165	0,0734	0,5706	0,7165
D	Pozitif	0,9952	0,014	0,0183	0,0393	0,9365	0,2109
	Negatif	0,0048	0,014	0,9817	0,0393	0,0635	0,2109

Tablo 5. Kalsifikasyon Değişkenine Ait Parametre Tahminleri

Kalsifikasyon Değişkeni	Sınıf 1 (-)	s.e.	Sınıf 2 (+)	s.e.	Sınıf 3 (-)	s.e.	
Gizli Sınıf Prevelans Değeri	0,6021	0,0758	0,2998	0,0766	0,0981	0,0549	
A	Pozitif	0,0076	0,0132	0,6338	0,1355	0,7516	0,2291
	Kararsız	0,032	0,0269	0,1414	0,0806	0,1175	0,0843
	Negatif	0,9604	0,0373	0,2248	0,1161	0,1309	0,173
B	Pozitif	0,0543	0,0448	0,8633	0,1422	0,0122	0,0494
	Kararsız	0,1381	0,0679	0,1065	0,0847	0,0697	0,1282
	Negatif	0,8076	0,0783	0,0302	0,0665	0,9181	0,1755
C	Pozitif	0,0054	0,015	0,9837	0,0359	0,9389	0,1324
	Negatif	0,9946	0,015	0,0163	0,0359	0,0611	0,1324
D	Pozitif	0,0837	0,0555	0,9841	0,0358	0,7057	0,243
	Negatif	0,9163	0,0555	0,0159	0,0358	0,2943	0,243

Tablo 6. Kavite Değişkenine Ait Parametre Tahminleri

Kavite Değişkeni	Sınıf 1(-)	s.e.	Sınıf 2 (0)	s.e.	Sınıf 3 (+)	s.e.	
Gizli Sınıf Prevelans Değeri	0,4462	0,6214	0,4462	0,6214	0,1077	0,0491	
A	Pozitif	0,0005	0,005	0,0005	0,005	0,2284	0,2006
	Negatif	0,9995	0,005	0,9995	0,005	0,7716	0,2006
B	Pozitif	0,0309	0,0912	0,0309	0,0912	0,6731	0,2205
	Kararsız	0,0215	0,0358	0,0215	0,0358	0,0539	0,0548
	Negatif	0,9476	0,1213	0,9476	0,1213	0,273	0,2092
C	Pozitif	0,0014	0,0089	0,0014	0,0089	0,6852	0,2274
	Negatif	0,9986	0,0089	0,9986	0,0089	0,3148	0,2274
D	Pozitif	0,0019	0,0108	0,0019	0,0108	0,9129	0,1513
	Negatif	0,9981	0,0108	0,9981	0,0108	0,0871	0,1513

Tablo 7. Lenfodenöpati Değişkenine Ait Parametre Tahminleri

<i>Lenfodenöpati Değişkeni</i>	Sınıf 1(-)	s.e.	Sınıf 2 (+0)	s.e.	Sınıf 3 (-)	s.e.	
Gizli Sınıf Prevelans Değeri	0,491	0,1478	0,4236	0,0783	0,0854	0,1289	
A	Pozitif	0,0001	0,0004	0,7662	0,1038	0,0043	0,027
	Kararsız	0,0124	0,0246	0,2044	0,0975	0,0858	0,2361
	Negatif	0,9876	0,0249	0,0293	0,0452	0,9099	0,2619
B	Pozitif	0,0005	0,0021	0,6458	0,1156	0,0138	0,0676
	Kararsız	0,0095	0,0199	0,1561	0,0851	0,0495	0,1166
	Negatif	0,99	0,0219	0,1981	0,0968	0,9367	0,1826
C	Pozitif	0,0972	0,1005	0,9908	0,0229	0,6741	0,8122
	Negatif	0,9028	0,1005	0,0092	0,0229	0,3259	0,8122
D	Pozitif	0,1225	0,1662	0,7755	0,1	0,7178	0,4999
	Kararsız	0,0224	0,0233	0,0271	0,0273	0,0295	0,0327
	Negatif	0,855	0,1747	0,1974	0,0951	0,2527	0,4851

Tablo 8. Nodül Değişkenine Ait Parametre Tahminleri

<i>Nodül Değişkeni</i>	Sınıf 1(-)	s.e.	Sınıf 2 (+)	s.e.	Sınıf 3 (-)	s.e.	
Gizli Sınıf Prevelans Değeri	0,6274	0,0819	0,2765	0,0699	0,0962	0,0559	
A	Pozitif	0,0017	0,0044	0,8093	0,1158	0,0018	0,0089
	Kararsız	0,0422	0,0372	0,1608	0,104	0,0427	0,0969
	Negatif	0,9561	0,0399	0,0299	0,0483	0,9555	0,105
B	Pozitif	0,0186	0,0442	0,9083	0,0833	0,9073	0,1505
	Kararsız	0,0448	0,0375	0,0587	0,0564	0,0592	0,0791
	Negatif	0,9367	0,0703	0,0329	0,0489	0,0335	0,0856
C	Pozitif	0,0009	0,0037	0,9139	0,0815	0,7467	0,3416
	Kararsız	0,007	0,0137	0,0454	0,0505	0,0837	0,0915
	Negatif	0,9921	0,0169	0,0407	0,056	0,1696	0,3099
D	Pozitif	0,0036	0,0118	0,978	0,0443	0,0249	0,0816
	Negatif	0,9964	0,0118	0,022	0,0443	0,9751	0,0816

Tablo 9. Pnömotoraks Değişkenine Ait Parametre Tahminleri

<i>Pnömotoraks Değişkeni</i>		<i>Sınıf 1(-)</i>	<i>s.e.</i>	<i>Sınıf 3 (+)</i>	<i>s.e.</i>
Gizli Sınıf Prevelans Değeri		0,4835	0,6192	0,0331	0,0282
A	Pozitif	0,0004	0,0046	0,7439	0,3853
	Negatif	0,9996	0,0046	0,2561	0,3853
B	Pozitif	0,0004	0,0046	0,7439	0,3853
	Negatif	0,9996	0,0046	0,2561	0,3853
C	Pozitif	0,0004	0,0046	0,7439	0,3853
	Negatif	0,9996	0,0046	0,2561	0,3853
D	Pozitif	0,0256	0,1427	0,7624	0,3686
	Negatif	0,9744	0,1427	0,2376	0,3686

Tablo 10. Skar Dokusu Değişkenine Ait Parametre Tahminleri

<i>Skar Dokusu Değişkeni</i>		<i>Sınıf 1(+)</i>	<i>s.e.</i>	<i>Sınıf 2 (0)</i>	<i>s.e.</i>	<i>Sınıf 3 (-)</i>	<i>s.e.</i>
Gizli Sınıf Prevelans Değeri		0,5436	0,2025	0,2739	0,0714	0,1825	0,2
A	Pozitif	0,8846	0,0741	0,0168	0,0394	0,9019	0,1716
	Kararsız	0,0298	0,0304	0,0138	0,0193	0,0274	0,0353
	Negatif	0,0856	0,0641	0,9694	0,0544	0,0708	0,1477
B	Pozitif	0,9299	0,121	0,049	0,0655	0,3075	0,6552
	Kararsız	0,0579	0,0758	0,1104	0,0954	0,2097	0,1522
	Negatif	0,0122	0,05	0,8406	0,1129	0,4828	0,5863
C	Pozitif	0,9737	0,1164	0,1386	0,1025	0,7274	0,2614
	Kararsız	0,0126	0,033	0,0368	0,0425	0,0443	0,0562
	Negatif	0,0137	0,0853	0,8247	0,1132	0,2283	0,2327
D	Pozitif	0,9526	0,0478	0,1917	0,1261	0,9847	0,0534
	Negatif	0,0474	0,0478	0,8083	0,1261	0,0153	0,0534

Sonuç ve Öneriler

Tele-Radyoloji uygulamasında, gözlemcilerarası uyumun, güvenilirliğin değerlendirilmesinde olasılıksal model yaklaşımı, Gizli Sınıf Analizi (GSA) bu çalışmada dört farklı radyolog tarafından 8 farklı TB tanısı için verilen karar düzenleri verisi üzerinde uygulanmıştır. GSA ile üç sınıflı gizli sınıf modeli oluşturulmuş, model uyum istatistikleri, her bir değişken ait tanımlayıcı istatistikler ve her bir değişken için ayrı ayrı parametre tahminleri

standart hata değerleri ile elde edilmiştir. Parametre tahminlerinden güvenilirlik değerleri ve hata oranları her bir radyolog için hesaplanmış ve ayrıca her bir kategoriye ait hatalı sınıflandırma oranları ile, ağırlıklı güvenilirlik değerleri elde edilmiştir. Elde edilen sonuçlar değerlendirildiğinde görüntülerin dijital imajlarını web tabanlı sistem üzerinden değerlendiren radyologların doğru sınıflandırma olasılıkları da hemen hemen tüm değişkenlerde oldukça yüksek olup, bazı değişkenlerde hatta orijinal imajları değerlendiren radyologlardan daha yüksek olduğu belirlenmiştir. GSA, gözlemcilerarası uyum verilerinin değerlendirilmesinde model tabanlı bir yaklaşım olup, kullanım alanları oldukça geniştir. Kullanımı kolay paket programlarla analiz edilebilir olması ve diğer uyum indislerine göre avantajlarının bulunması nedeni ile karar vericilere çalışmalarında önerilmektedir.

Kaynaklar

1. Uebersax, S.John., Grove, M.William. (1990). "Latent Class Analysis of Diagnostic Agreement", *Statistics in Medicine*, Vol:9, 559-572.
2. Hripcsak G. (2002). "Measurement Agreement in Medical Informatics Reliability Studies", *Biomedical Informatics*, Vol:35(2), 99-110.
3. Flaherty, P., Brian, P. (2002). "Assessing Reliability of Categorical Substance Use Measures with Latent Class Analysis", *Drug and Alcohol Dependence*, Vol:68, 7-20.
4. Schuster, C. (2002). "A Mixture Model Approach to Indexing Rater Agreement", *Vol:15*, 289-303
5. Lazarsfeld, P. F ve Henry, N.W. *Latent Structure Analysis*, Houghton Mifflin, Boston, Massachusetts, 1968.
6. Goodman, L. A. "Exploratory latent structure analysis using both identifiable and unidentifiable models", *Biometrika*, 61, 215-231, 1974b.
7. Haberman, S. J. *Analysis of Qualitative Data, Volume 2: New Developments*, New York: Academic Press, 1979.
8. Agresti, A., Lang, B. (1993). "Quasi-Symmetric Latent Class Models, with Application to Rater Agreement", *Biometrics*, 49, 131-139.
9. Uebersax, S.John. (2001), "Latent Class Models for Multiple Ratings or Tests", *Statistics in Medicine*
10. Espeland, M., Handelman, S. (1989), "Using Latent Class Models to Characterize and Assess Relative Error in Discrete Measurements", *Biometrics*, 45, 587-599.
11. Heinen, T. *Latent Class and Discrete Latent Trait Models: Similarities and Differences*, Thousand Oaks, California: Sage, 1996.

12. Clogg C. C., Goodman, L. A. "latent structure analysis of a set of multidimensional contingency tables", Journal of the American Statistical Association, 79, 762-771, 1984.
13. Goodman, L. A. "A new model for scaling response patterns: an application of the quasi-independence concept", Journal of the American Statistical Association, 70, 755-768, 1975.
14. Dillon, W. R. Ve Mulani, N. "A probabilistic latent class model for assessing inter-judge reliability", Multivariate Behavioral Research, 19, 438-458, 1984.
15. Çolak, M. (2003), Gizli Sınıf (Latent Class) Analizi ve Tanı Testlerinin Değerlendirilmesi, Sağlık Bilimleri Enst, Biyoistatistik Ana Bili Dalı, Doktora Tezi, Hacettepe Üniv, Ankara.
16. Vermunt, J. K., Magidson, J. Latent GOLD User's Guide. Belmont, Mass. Statistical Innovations, Inc, 2000.
17. Clogg, C. C., Unrestricted and restricted maximum likelihood latent structure analysis: a manual for users. Working paper 1977-09, Pennsylvania State University, Population Issues Research Center, 1981.
18. Clogg, C. C. "Some latent structure models for the analysis of Likert-type data", Social Science Research, 8, 287-301, 1979.
19. Dayton, C. M. ve Macready, G. B. "A scaling model with response errors and intrinscally unscalable respondents", Psychometrika, 45, 343-356, 1980.
20. Formann, A. K. "Constrained latent class models: theory and applications", British Journal of Mathematical and Statistical Psychology, 38, 87-111, 1985.
21. Rudas, T., Clogg, C. C., Lindsay, B. G. "A new index of fit based on mixture methods for the analysis of contingency tables", Journal of the Royal Statistical Society, Series B, 56, 623-639, 1994.

İLERİYE DÖNÜK ÇALIŞMALARDA TANI TESTLERİNİN DUYARLILIK VE SEÇİCİLİKLERİ İÇİN ÖRNEKLEM BÜYÜKLÜĞÜ VE GÜÇ HESAPLAMASI

Jale Karakaya, Erdem Karabulut, Sevilay Karahan, Osman Saraçbaşı

*Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara
e-mail : jalek@hacettepe.edu.tr*

Özet

Bir hastalık tarama çalışmasının düzeni, testlerin duyarlılık ve seçicilikleri için hipotez testlerine dayalı olabilir. Olgu-kontrol çalışmalarında, bireyler çalışmaya katıldığında hastalık durumlarının (hastalık var-yok) bilinmesi gerekir. Ancak, ileriye yönelik çalışmalarda, tarama testinden sonra, önceden belirlenmiş bir tarihte altın standart test uygulanana kadar bireylerin hastalık durumları kesin olarak bilinmez. Bu nedenle, hasta ve hasta olmayan bireylerin sayısının rasgele olduğu düşünülür. Olgu-kontrol çalışması düzeni için kullanılan yöntemler ileriye yönelik çalışmalar için uygun değildir. Toplam örneklem genişliğinin belirlenmesi için çeşitli ön test güç hesaplama işlemleri araştırmacılar tarafından sıklıkla kullanılır. Ancak, literatürde bu yöntemlerin özelliklerinin çok iyi anlaşılmadığı belirtilmektedir. Li ve Fini tarafından; test istatistiklerinin özelliklerinin koşulsuz gücüne dayalı örneklem genişliği ve güç hesaplama yöntemi geliştirilmiştir. Tek örneklem, bağımlı ve bağımsız iki örneklem düzeni için duyarlılık-seçicilik değerlerinin karşılaştırılmasında örneklem büyüklüğü ve güç hesaplama yöntemi önerilmiştir.

Tek Örneklem Düzeni:

H_0 : Duyarlılık = Duyarlılık₀ ($Se=Se_0$)

H_1 : Duyarlılık \neq Duyarlılık₀ ($Se \neq Se_0$)

n_1 : Hasta olan kişiler

s_1 : n_1 hastaların içinde test sonucu pozitif olanların sayısı

Hipotezini test etmek için gerekli örneklem büyüklüğü ve güç hesaplaması 4 farklı yöntemle yapılmıştır. Bu yöntemler;

Yöntem 1

$$n(\alpha, \beta) = n_1(\alpha, \beta) / p$$

alternatif bir yaklaşım;

$$P[n_1 > n_1(\alpha, \beta) / n, p] = \phi,$$

Yöntem 2; Normal dağılım yaklaşımı,

Öncelikle hasta bireyler sayısı aşağıdaki formülle hesaplanır.

$$n_1(\alpha, \beta) = \left(\frac{Z_\alpha \sqrt{Se_0 * (1 - Se_0)} + Z_\beta \sqrt{Se_1 * (1 - Se_1)}}{Se_1 - Se_0} \right)^2$$

Yöntem 3; Binom dağılımı yaklaşımı

Aşağıdaki iki eşitliğin eşanlı çözülmesi gerekir.

$$P[s_1 > s(\alpha, \beta) / n_1(\alpha, \beta), Se_0] = \alpha$$

$$P[s_1 > s(\alpha, \beta) / n_1(\alpha, \beta), Se_1] = 1 - \beta$$

$$\sum_{s_1=[s(\alpha, \beta)+1]}^{n_1(\alpha, \beta)} \binom{n_1(\alpha, \beta)}{s_1} Se^{s_1} (1 - Se)^{n_1(\alpha, \beta) - s_1}$$

Yukarıdaki eşitliklerin sol tarafı aşağıdaki gibi gösterilebilir.

Yöntem 4; Koşulsuz güç $(1 - \beta^*)$ ile hesaplama

$$\min : \left\{ n : \sum_{k=0}^n P[s_1 > s(k) / k, Se_1] P(n_1 = k / n, p > 1 - \beta^*) \right\}$$

Güç fonksiyonu $1 - \beta^*$ dan büyük olana kadar n değeri artırılarak hesaplamalar yapıldığından, eşitsizliğin sol tarafı n'nin monoton artan bir fonksiyonudur.

İki Örneklem Düzeni:

H_0 : Duyarlılık1=Duyarlılık2 ($Se_1=Se_2$)

H_1 : Duyarlılık1 \neq Duyarlılık2 ($Se_1 \neq Se_2$)

Bağımsız ve bağımlı iki örneklem için de aynı yaklaşımlar kullanılmıştır.

Sonuç olarak bu çalışmada, çeşitli prevalans (0,02; 0,05; 0,10; 0,20; 0,25) ve duyarlılık değerleri için ileriye yönelik çalışmalarda örneklem genişliği ve güç için 4 ayrı yöntemle hesaplamalar yapılarak sonuçları değerlendirilmiştir.

Anahtar Kelimeler: Duyarlılık, seçicilik, ileriye yönelik çalışmalar, örneklem büyüklüğü, güç

Kaynaklar

1. Li J ve Fine J, On sample size for sensitivity and specificity in prospective diagnostic accuracy studies. *Statistics in Medicine*: 2004; 23:2537-2550.
2. Alonzo T, Pepe M ve Moskowitz C, sample size calculations for comparative studies of medical tests for detecting presence of disease. *Statistics in Medicine* : 2002; 21:835-852.
3. Fleiss J, *Statistical Methods For Rates and Proportions*. John Wiley & Sons 1981. Canada.
4. Obuchowski NA and Zhou XH. Prospective studies of diagnostic test accuracy when disease prevalence is low. *Biostatistics* 2002; 3:477-492.

KULLBACK-LEIBLER UZAKLIĞI VE LORENZ EĞRİSİNDEN ELDE EDİLEN ÖLÇÜLER KULLANILARAK TANI TESTLERİNİN DEĞERLENDİRİLMESİ

Umur Arslan¹, Barış Kaya², Necla Özer², Ergun Karağaoğlu¹

H.Ü. Tıp Fakültesi Biyoistatistik Anabilim Dalı¹

H.Ü. Tıp Fakültesi Kardiyoloji Anabilim Dalı²

e-mail: umutars@hacettepe.edu.tr

Özet

Bu çalışmanın amacı, tanı testlerinin performanslarının değerlendirilmesinde kullanılan Kullback-Leibler uzaklığı ve Lorenz eğrisinden elde edilen ölçüler hakkında bilgi vermek, ve bunları karşılaştırmalı olarak bir örnek üzerinde göstermektir. Geleneksel yöntemlerin yanı sıra yeni geliştirilen yöntemlerden biri Kullback-Leibler uzaklığıdır. Kullback-Leibler uzaklığı; "hastalığın varlığının kanıtlanması" (rule in) ve "hastalığın yokluğunun kanıtlanması"nda (rule out) hangi tanı testinin tercih edilmesi gerektiğini gösteren bir yöntemdir. Bir diğer yöntem Lorenz Eğrisi ve bu eğriden türetilen Pietra ve Gini indeksi değerleridir. Pietra indeksi, test tarafından sağlanan hastalık olasılığındaki ortalama mutlak değişimi verirken, Gini indeksi rasgele seçilen iki deneğin son test olasılıklarındaki ortalama mutlak farkı vermektedir. Ayrıca her iki indeks maksimum prognostik bilgi yüzdesini de ölçmektedir.

Bu amaçla uygulamada, kalp krizi tanısında kullanılan, mitral anüler sistolik velosite (MASV) ölçümü, strain skor indeksi (SSİ) ve strain rate skor indekslerinin (SRSİ) ayırıcılık güçleri ROC curve analizi, Kullback-Leibler yöntemi ve Lorenz eğrisi indeksleri ile değerlendirilmeye çalışılmıştır.

Anahtar kelimeler: Tanı Testleri, Kullback-Leibler Uzaklığı, Lorenz Eğrisi, Pietra İndeksi, Gini İndeksi.

Giriş

Tanı testlerinin, belli bir hastalığa sahip olanla olmayanı ayırmadaki başarısını gösterebilmek için farklı performans ölçüleri kullanılmaktadır. Tanı testlerinin performansı genellikle duyarlılık, seçicilik, olabilirlik oranları ve ROC eğrisi altında kalan alanlar ile tanımlanır. Ancak bu ölçütler bir tanı

testinin özelliklerini ayrıntılı bir şekilde incelemek için tek başına yeterli değildir. Tanı testlerinin ayırıcılık gücünün daha ayrıntılı değerlendirilmesi için yeni yöntemler üzerinde çalışılmaktadır. Bu yöntemlerden birisi hasta olanlarla olmayanların olasılık dağılımları arasındaki uzaklığın hesaplanması ile elde edilen ve Bilgi Kuramı Yaklaşımı içerisinde yer alan Kullback-Leibler uzaklığıdır. Bu uzaklık "hastalığın varlığının kanıtlanması" (rule in) ve "hastalığın yokluğunun kanıtlanması"nda (rule out) hangi tanı testinin tercih edilmesi gerektiğini gösteren bir yöntemdir. Yine tanı testlerinin değerlendirilmesinde daha ayrıntılı bilgi verebilen bir diğer ölçüm Lorenz Eğrisi ve bu eğriden türetilen Pietra ve Gini İndeksidir.

Gereç ve Yöntem

Kullback-Leibler Uzaklığı:

Kullback-Leibler uzaklığı istatistik ve Bilgi Kuramı'ndan ortaya çıkan özet bir kavramdır. Bu uzaklık test öncesi "hastalığın varlığının kanıtlanması" (rule in) ve "hastalığın yokluğunun kanıtlanması" (rule out) potansiyelini iyi bir biçimde tanımlamaktadır. Kullback-Leibler uzaklığı hasta olan ve olmayanların olasılık dağılımları (f ve g dağılımları) arasındaki uzaklığı ölçer. Sıralı (ordinal) test sonuçlarına sahip bir tanı testinin sonuçları "i" ile gösterilsin. $i=1,2,\dots,K$. Bu durumda i. test kategorisinde hastalara ilişkin olasılık dağılımı f_i , hasta olmayanlara ilişkin olasılık dağılımı g_i 'dir

($\sum_{i=1}^K f_i = \sum_{i=1}^K g_i = 1$). Hasta olmayanların dağılımı referans olarak

alındığında Kullback-Leibler uzaklığı $D(f||g)$ ile hasta olanların dağılımı referans olarak alındığında $D(g||f)$ gösterilir. $D(f||g)$ uzaklık değeri büyük olan tanı testi hasta olanları belirlemede, $D(g||f)$ uzaklık değeri daha büyük olan tanı testi ise hasta olmayanları belirlemede daha başarılıdır. Bu uzaklıklar aşağıda ki eşitlikte gösterilmiştir.

$$D(f || g) = \sum_{i=1}^K f_i \cdot \log \frac{f_i}{g_i} = \sum_{i=1}^K f_i \cdot \log LR_i$$

$$D(g || f) = \sum_{i=1}^K g_i \cdot \log \frac{g_i}{f_i} = \sum_{i=1}^K g_i \cdot \log(1/LR_i) \quad \text{Burada } LR_i: i.$$

sınıfın olabilirlik oranı

İki sonuçlu tanı testleri için Doğru Pozitiflik Oranı (DPO) ve Yanlış Pozitiflik Oranı (YPO) kullanılarak bu uzaklık değerleri:

$$D(f \parallel g) = (1 - DPO) \log \frac{(1 - DPO)}{DNO} + DPO \log \frac{DPO}{(1 - DNO)}$$

$$D(g \parallel f) = DNO \log \frac{DNO}{1 - DPO} + (1 - DNO) \log \frac{(1 - DNO)}{DPO}$$

Hastalığın varlığının kanıtlanmasında $D(f||g)$ değeri daha büyük olan tanı testi, hastalığın yokluğunun kanıtlanmasında $D(g||f)$ değeri daha büyük olan tanı testleri tercih edilir. Ayrıca bu uzaklıklardan yararlanarak P_{in} ve P_{out} gibi iki farklı değer elde ederek tanı testlerinin hastalığın varlığını ve yokluğunu gösterebilme potansiyeli yorumlanabilir. $P_{in} = \exp D(f||g)$, $P_{out} = \exp D(g||f)$ dir. P_{in} rasgele seçilen hasta bir denek için, hastalık odds'undaki artışı gösterirken, P_{out} , rasgele seçilen hasta olmayan bir denek için, hastalık odds'undaki azalmayı gösteren bir değerdir.

Lorenz Eğrisi, Pietra ve Gini İndeksi:

Sıralı (ordinal) test sonuçlarına sahip bir tanı testinin sonuçları "i" ile gösterilsin. $i=1,2,\dots,K$. Pietra ve Gini indekslerinin hesaplanabilmesi için olabirlik oranları monoton olarak artmalıdır. Monoton artış göstermeyen olabirlik oranları küçükten büyüğe tekrar sıralanmalıdır. i. test kategorisinde hasta olmayanlara ilişkin yığılımlı olasılıklar X_i , hastalara ilişkin yığılımlı olasılıklar Y_i ile gösterildiğinde; Lorenz Eğrisi, hasta olmayanların yığılımlı yüzdelerine karşı, hasta olanların yığılımlı yüzdelerinin çiziminden elde edilen eğridir. Lorenz eğrisi ile referans (diagonal) çizgi arasında kalan üçgenlerden alanı en büyük olanın iki katı Pietra indeksini, Lorenz eğrisi ile referans çizgi arasında kalan alanın iki katı ise Gini indeksini verir.

$$Pietra = \max_{1 \leq i \leq K-1} \frac{X_i}{1} \frac{Y_i}{1}$$

$$Gini = \sum_{i=1}^{K-1} \left| \begin{array}{cc} X_i & Y_i \\ X_{i+1} & Y_{i+1} \end{array} \right|$$

Test uygulanmadan önce bir kişinin hasta olup olmadığını tahmin etmek için prevalans değeri kullanılır. Fakat populasyon heterojen denekler içerebilir bu nedenle prevalans kaba bir göstergedir. Testten sonra “test sonrası olasılık” prevalans değerinden düşük ya da yüksek olabilir. Pietra indeksi kullanılarak “test tarafından sağlanan hastalık olasılığındaki ortalama mutlak değişim (ΔP)” hesaplanabilir.

$$DP = \sum_{i=1}^K P(T_i) |P(D|T_i) - P(D)|$$

$DP = 2P(D)[1 - P(D)]$ Pietra Burada, $P(D)$: Test öncesi olasılık (prevalans)

Bir diğer ölçüm rasgele seçilen iki deneğin son test olasılıkları arasındaki ortalama mutlak farktır (ΔP^*). Gini indeksi kullanılarak ΔP^* değeri hesaplanabilir.

$$DP^* = \sum_{i=1}^K \sum_{j=1}^K P(T_i)P(T_j) |P(D|T_i) - P(D|T_j)|$$

$$DP^* = 2P(D)[1 - P(D)] Gini$$

ΔP ve ΔP^* değerleri altın standart testte maksimumdur. Altın standart testten elde edilen ΔP ve ΔP^* değerleri maksimum prognostik bilgidir.

$$\max DP = 2P(D)[1 - P(D)]$$

$$\max DP^* = 2P(D)[1 - P(D)]$$

Buna göre altın standartla karşılaştırılan bir tanı testi için maksimum prognostik bilgi yüzdesi Pietra ve Gini indekslerini vermektedir.

$$Pietra = \frac{DP}{\max DP}$$

$$Gini = \frac{DP^*}{\max DP^*}$$

Pietra ve Gini indeksleri tanı testlerinin performanslarının farklı bir açıdan değerlendirilmesini sağlamaktadır. Pietra test sonrası hastalık olasılıklarındaki beklenen değişimi ölçerken, Gini test sonrası olasılıkların denekten deneğe olan değişimini ölçer. Test öncesi hastalık olasılıkları eşit olan bir grup birey için en olası ayırımı yapmak gerektiğinde Gini indeksi büyük olan bir test tercih edilmelidir. Bu özellik Pietra için de geçerlidir.

Yukarıdaki yöntemlerin uygulanabilmesi için, Hacettepe Üniversitesi Tıp Fakültesi Kardiyoloji Anabilim Dalı'na göğüs ağrısı şikayeti ile başvuran 101 hastanın kalp krizi geçirip geçirmediği incelenmiş, buna göre kalp krizi geçirme tanısı alan 71 hasta ve 30 sağlıklı birey için kalbin bölgesel kasılmasını ölçebilen yeni değişkenler değerlendirilmeye çalışılmıştır.

Bulgular

Kalp krizi tanısında kullanılan mitral anüler sistolik velosite (MASV), strain skor indeksi (SSİ) ve strain rate skor indeksi (SRSİ) değişkenlerinin tanı performansları ROC analizine göre değerlendirildiğinde elde edilen değerler sırasıyla 0,884, 0,908, ve 0,885'tir. Kullback-Leibler uzaklığına göre bulunan sonuçlar Tablo 1'de, Lorenz eğrisi indekslerine göre elde edilen sonuçlar Tablo 2'de verilmiştir.

Tablo 1. Kalp Krizi Tanısında Kullanılan MASV, SSİ ve SRSİ Değişkenlerinin Kullback-Leibler Uzaklığı P_{in} ve P_{out} Değerleri

Kalp Krizi Tanısında Kullanılan Değişkenler	$D(f g)$	P_{in}	$D(g f)$	P_{out}
MASV	1,708	5,517	1,647	5,189
SSİ	2,098	8,152	2,136	8,464
SRSİ	2,061	7,855	2,037	7,664

Tablo 2. Kalp Krizi Tanısında Kullanılan MASV, SSİ ve SRSİ Değişkenlerinin Lorenz Eğrisi İndeksleri

Kalp Krizi Tanısında Kullanılan Değişkenler	Pietra İndeksi	Gini İndeksi
MASV	0,764	0,812
SSİ	0,837	0,863
SRSİ	0,811	0,857

Tartışma ve Sonuç

Kalp krizi tanısında kullanılmak istenen ve kalbin bölgesel kasılmasını ölçen üç değişken Kullback-Leibler uzaklığı ve Lorenz Eğrisi indekslerine göre değerlendirildiklerinde en iyi sonuç veren değişken SSİ'dir. SSİ değişkeninin Kullback-Leibler uzaklıkları diğer iki değişkene göre daha büyüktür ve bu değişken hem hastalığın varlığının gösterilmesi hem de yokluğunun gösterilmesinde MASV ve SRSİ değişkenlerine göre daha başarılıdır. SSİ değişkeni rasgele seçilen hasta bir bireyin test sonrası hasta olma odsunu test öncesi hasta olma odsuna göre 8,152 kat artırmaktadır. Benzer şekilde hasta olmayan bir bireyin test sonrası hastalık oddsu test öncesi odsuna göre 8,464 kat azalmaktadır. Yukarıdaki üç değişken Pietra ve Gini indeksleri ile değerlendirildiklerinde, yine en iyi değişken SSİ en kötü değişken MASV değişkenidir. SSİ değişkeni, test öncesine göre test sonrası hastalık olasılıklarındaki beklenen değişimi en fazla olan değişkendir ve prognostik bilgi yüzdesi 0,837'dir. Yine aynı test için son test olasılıklarında, denekten deneğe olan değişim 0,863'tür ve denekten deneğe değişebilecek maksimum prognostik bilgi yüzdesi 0,863'tür.

Klinik tıpta hastalığın varlığını ya da yokluğunu göstermek amacıyla altın standart gibi kesin sonuç vermeyen tanı testleri kullanılmaktadır. Bu testlerin ayrıcalık gücünün değerlendirilmesi için en çok kullanılan yöntemler duyarlılık seçicilik ve ROC eğrisi analizidir. Bunun dışında test performanslarının değerlendirilebilmesi için daha ayrıntılı bilgi veren ölçülere ihtiyaç vardır. Yukarıda sözü edilen performans ölçülerine göre Kulback-Leibler uzaklığı hastalığın varlığının ve yokluğunun gösterilmesi ile ilgili daha ayrıntılı bilgi verirken, Lorenz eğrisi indeksleri farklı şekillerde yorumlanabildiği için yine testlerin performanslarının karşılaştırılması ve değerlendirilmesinde oldukça kullanışlıdır.

Kaynaklar

1. Wen-Chung Lee. Selecting Diagnostic Tests for Ruling Out or Ruling In Disease: The Use of the Kullback-Leibler Distance. International Journal of Epidemiology. 1999; 28: 521-525.
2. Wen-Chung Lee. Probabilistic Analysis of Global Performances of Diagnostic Tests: Interpreting the Lorenz Curve-Based Summary Measures: Statistics in Medicine 1999; 18: 455-471.

ROC EĞRİSİ ALTINDAKİ ALANIN PARAMETRİK VE PARAMETRİK OLMAYAN YÖNTEMLERLE HESAPLANMASI

Sema Erden, Bahar Taşdelen

Mersin Üniversitesi Biyoistatistik Anabilim Dalı
e-mail: semaerden@mersin.edu.tr

Özet

Bu çalışmada, ROC (Receiver Operating Characteristics) eğrisi altında kalan alanın parametrik ve nonparametrik yöntemler ile hesaplama adımlarının gerçek ve simulasyon yöntemiyle üretilmiş iki tip veri seti kullanılarak, gösterilmesi amaçlanmıştır. ROC eğrisi genellikle tanı testlerinin diagnostik performanslarını karşılaştırmak ya da özetlemek için kullanılır. ROC eğrisinin çizimi, bireylerin hasta olup olmama ya da bir koşulu sağlayıp sağlayamama durumunu kontrol etmek için kullanılan, tanı testinin ayırıcılıktaki doğruluğunu grafik üzerinde gösteren popüler bir yöntemdir. Tanı testinin yararlılığını ölçebilmek için o testi yorumlayanların performansının da ölçülebilmesi gereklidir. ROC eğrileri yorumlayıcı performansını incelemek içinde iyi bir temel oluşturur.

ROC eğrisi altında kalan alan ise tanı testlerinin etkinliğini ortaya koymak için ve testin doğruluğunu tek bir sayısal değerle özetlemek için oldukça önemlidir. Bu alanın tahmini için kullanılan parametrik yöntem normal varsayımlar yöntemidir (normal assumptions methods). Normallik varsayımı sağlanmadığı takdirde parametrik olmayan (nonparametrik), Mann-Whitney istatistiği ve Kernel Smoothing yöntemi kullanılabilir. Bu yöntemin algoritmaları ve hesaplama adımlarının detaylı olarak açıklanabilmesi amacıyla bu çalışmada, Mersin Üniversitesi Tıp Fakültesi Biyokimya laboratuvarına çeşitli nedenlerden dolayı kan vermiş 49 kalp hastası ve 73 kontrol grubuna ait trigliserid değerleri kullanılmıştır. Bu değerlere $\lambda=0$ değeri için Box-Cox transformasyonu uygulanarak normalliğe yaklaşması sağlanmış ve Kernel Smoothing yöntemi ile eğri altında kalan alan hesaplanmıştır. Diğer yöntemlerde ise simulasyon yöntemi kullanarak veri üretilmiştir. Yapılan işlemler sonucunda alanlar arasında istatistikî olarak önemli bir fark saptanmamıştır.

Anahtar Kelimeler: ROC eğrisi, Kernel Smoothing, Box-Cox, Mann-Whitney.

FARKLI KORELASYON YAPILARINA SAHİP VERİ SETLERİNDE BAZI REGRESYON DİAGNOSTİKLERİNİN TARTIŞILMASI

Ayşe Canan Yazıcı¹, Arzu Kanık², Fikret Gürbüz³

Başkent Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı¹

Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı²

Ankara Üniversitesi Ziraat Fakültesi, Zootekni Bölümü, Biyometri-Genetik Anabilim Dalı³

e-mail: acyazici@baskent.edu.tr

Özet

Çoklu regresyon analizi yönteminin uygulanması sırasında bazı problemlerle karşılaşabilmektedir. Söz konusu problemler hipotez kontrollerinde önceden kararlaştırılmış olan 1. tip hatanın artmasına ve tahminlerdeki isabet derecesinin önemli düzeyde düşmesine neden olabilir. Bu gibi durumlarda söz konusu problemlerin varlığı araştırılmalı, veri seti ayrıntılı bir biçimde incelenmelidir. Regresyon analizinde model ya da veri seti ile ilgili problemlerin saptanmasında kullanılan istatistikler, genel olarak “regresyon diagnostikleri” olarak isimlendirilirler. Bu çalışmada bazı aykırı gözlem ve etkili gözlem istatistikleri ele alınarak farklı korelasyon yapılarından etkilenip etkilenmedikleri araştırılmıştır. Bu amaçla, simülasyonla farklı korelasyon matrislerine göre üretilen sekiz adet veri setine söz konusu diagnostik istatistikleri uygulanmıştır. Sonuç olarak çalışmada yer verilen istatistiklerin farklı korelasyon yapılarından etkilenmediği görülmüştür.

Anahtar Kelimeler: Regresyon diagnostikleri, artıkların analizi, etki istatistikleri.

REGRESYON MODELLERİNE ALTERNATİF BİR YAKLAŞIM: MARS

Gülhan Orekici Temel¹, Handan Çamdeviren², A. Canan Yazıcı³

¹Mersin Üniversitesi, İktisadi ve İdari Bilimler Fakültesi

²Mersin Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı

³Başkent Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı

e-mail: gulhan_orekici@hotmail.com

Özet

Bu çalışmada, MARS tekniğinin temel özellikleri ve uygulama adımları üzerinde durulmuştur. Ayrıca bu tip karmaşık algoritmaların kullanımının getireceği avantajlar açıklanmıştır. Bu teorik bilgilere ilaveten kanserli ve Behçetli hastaların ailelerinden tespit edilmiş bir takım davranış ve psikolojik test sonuçlarının bu kişilerdeki Beck depresyon durumunu tahmin etmede kullanılabilecek bir MARS modeli elde edilmiştir. Bu modelin özellikleri açıklanmış ve tahmindeki başarısı oldukça yüksek bulunmuştur.

Anahtar Kelimeler: MARS, etkileşim, temel fonksiyon, veri madenciliği.

Giriş

İki ya da daha fazla değişken arasındaki ilişkilerin incelenmesinde regresyon modelleri sıkça kullanılmaktadır. Regresyon modellerinde, gözlenen bir sonuç değişkeni değerlendirilirken hangi etkenlerin sonuç üzerinde etkili olduğunun ve etkili olan değişkenler içerisinde hangi değişkenlerin etkisinin daha fazla olduğunun araştırılması amaçlanmaktadır. Modelde, etkilenen değişken bağımlı, etkileyen değişkenler ise bağımsız değişken olarak adlandırılır (1).

Değişkenler arasındaki ilişkiyi ortaya koymak amacıyla kullanılan modeller, doğrusal, doğrusal olmayan veya karışık yapıda olabilir. Kurulan modellerin gerçek yapıyı daha iyi yansıtmaları için özellikle son yıllarda çok sayıda bağımsız değişken bir arada değerlendirilmektedir. Günümüzde bilgisayar teknolojisindeki gelişmelere paralel olarak çok sayıda değişkeni bir arada değerlendirebilen karmaşık algoritmaların kullanılabilirliği artmıştır.

Bu algoritmaları kullanan regresyon yöntemlerinden birisi de **MARS** (*Multivariate Adaptive Regression Splines*, Çok Değişkenli Uyumlu Regresyon Uzanımları) tekniğidir. Bu teknik, pratikte yaygın kullanılan doğrusal ve doğrusal olmayan regresyon modellerinin aksine bağımsız değişkenin farklı aralık değerleri için farklı katsayılar türetir ve modele interaksiyon terimlerini de katarak gerçek yapıyı daha iyi yansıtan modeller geliştirir. Ayrıca bağımlı ve bağımsız değişkenlerin dağılımları üzerine her hangi bir varsayım gerektirmez ve değişken tipi de önemli değildir (2).

Bu çalışmanın amacı MARS yönteminin teorik temelleri üzerinde durmak ve söz konusu modelin tıp biliminde uygulanabilirliğini bir veri seti üzerinde göstermeye çalışmaktır.

Çalışmanın birinci bölümünde, MARS analizinin teorik temelleri üzerinde durulmuştur. İkinci bölümde analizin avantajları ve dezavantajları tartışılmış ve üçüncü bölümde uygun bir veri setinin tanıtımı yapılmış ve son olarak dördüncü bölümde veri setinden elde edilen sonuçlar yorumlanmaya çalışılmıştır.

1. BÖLÜM

1.1. Tanım

1991 yılında Stanford'da fizikçi ve istatistikçi Jerome Friedman tarafından MARS tekniği ilk olarak ortaya atılmıştır. Mars adimsal bir regresyon yöntemidir ve regresyon setinin performansını geliştirmek için tekrarlamalı ayırma metodunun ve adimsal lineer regresyonun geliştirilmiş bir hali olarak görülebilir (3). Mars, bağımlı ve bağımsız değişkenler arasındaki doğrusal olmayan ilişkileri doğrusal yapıya dönüştürme amacıyla uygun transformasyonlar bulmada ve bağımsız değişkenler arasındaki interaksiyonları belirlemede optimum bir yeniliktir (4). Mars tekniği bağımlı ve bağımsız değişkenler arasında oluşturulan fonksiyonel ilişkiye ait hiçbir varsayım içermeyen parametrik olmayan bir yöntemdir. Bağımsız değişkenlerin farklı aralıklarına karşılık gelen temel fonksiyonları kullanarak mars esnek bir regresyon modeli inşa eder.

1.2. Regresyon Modeli

X bağımsız değişken vektörü, Y bağımlı değişken vektörü ve ϵ hata terimi olmak üzere, X ve Y arasındaki ilişkinin denklem ile ifadesi 1 nolu eşitlikteki gibidir. Bu gösterime genel olarak regresyon denklemi ya da regresyon modeli denir (5).

$$Y=f(X)+\epsilon \quad (1)$$

Bu modelde $f\{X\}$, doğrusal, logistik veya doğrusal olmayan bir ilişki içinde olabilir. Örnek olarak Tablo 1'de $f\{X\}$ fonksiyonu farklı regresyon modelleri için tanımlanmıştır.

Tablo 1. Regresyon modelleri

$f\{X\}$	İlişki şekli
$f\{X\}=\beta_0+\beta_1X_1+\dots$	Doğrusal Model
$f\{X\}=\beta_0+\beta_1X_1+\beta_2X_2^2+\dots$	Parabolik Model
$f\{X\}=\beta_0+\beta_1X_1+\dots+\beta_{k-1}X_{k-1}^{k-1}+\beta_k X_k^k$	k. dereceden model

Regresyon modellerinin iki amacı vardır.

1. Bağımlı ve bağımsız değişken arasındaki yapısal ilişkiyi ortaya çıkarmak.
2. Bağımsız değişkenlere ait ölçüm vektöründen doğru ve güvenilir bir şekilde bağımlı değişkenin değerini tahmin etmek.

1. 3. MARS modeli

Parçalı temel fonksiyonlar ve bunların kombinasyonlarını kullanarak ve regresyon modellerindeki hem ileri doğru hem de geriye doğru ilerleme algoritmalarından yararlanılarak MARS modeli elde edilebilir. Model oluşumu iki adımda gerçekleşir (3).

1. Adım: Mümkün olan bütün temel fonksiyonlar oluşturulur. Çok büyük bir model bulununcaya kadar yani modelin karmaşıklığı maksimum seviyeye ulaşıncaya kadar eklenen temel fonksiyonlarla model büyür. Temel fonksiyonlar ya dönüştürülmüş tek değişkeni ya da çok değişkenli interaksyon terimlerinden meydana gelir. Bu fonksiyonlar modeldeki bağımsız değişkenlerin interaksyonlarını ve doğrusal olmayan dönüşümlerini hesaba katan bir fonksiyondur. Temel fonksiyonlar oluşturulurken tüm değişken ve değişkenlerin kombinasyonları tek tek ele alınır. Bu durumda, her bir bağımsız değişken ve bu değişkenlerin birbirleri ile olan kombinasyonlarının tanımlı bulunduğu aralıklardaki tüm olası değerler birer düğüm olarak düşünülüp, mümkün olan tüm olası temel fonksiyonlar belirlenir. Temel fonksiyonlar eklenerek model daha karmaşık ve daha esnek bir model haline gelir. Temel fonksiyon sayısı maksimum seviyeye ulaşıncaya kadar bu adım sürer. Ayrıca temel fonksiyonlar oluşturulurken aynı değişkene ait aynı temel fonksiyonuda oluşturulup modele girebilir. Aynı temel fonksiyon,

bağımlı değişkenle bağımsız değişken arasındaki dağılımın düğüm noktasına kadar olan eğiminin sıfır olduğunu gösterir.

MARS modeli (2) nolu eşitlikteki gibi tanımlanabilir.

$$Y = \beta_0 + \sum_{k=1}^K \beta_k B_k(X) \quad (2)$$

Burada;

k: Düğüm sayısını

K: Temel fonksiyon sayısını

X: Bağımsız değişkeni

β_0 : Modeldeki sabit terim

β_k : Regresyon katsayısı

$B_k(X)$: k. Temel fonksiyon

Temel fonksiyon ise aşağıdaki gibi tanımlanır.

$$B_k(X) = \prod_{j=1}^{J_k} [s_{kj} (x_{wkj} - t_{kj})]_+ \quad k=1, \dots, K \quad (3)$$

Buradaki;

J_k interaksiyon derecesini göstermektedir.

$[\cdot]_+ = \max[0, \cdot]$

$s_{kj} \in [\pm 1]$

t_{kj} : düğüm değeri ve

x_{wkj} : bağımsız değişken değerini göstermektedir (2).

Temel fonksiyonlar bağımsız değişkenlerin doğrusal olmayan dönüşümleri olabilirler. Fakat bağımlı değişken temel fonksiyonların doğrusal bir dönüşümüdür.

2. Adım: Biaslar ve varyansların optimal bir dengesi bulununcaya kadar temel fonksiyonlar modelden çıkarılır. Bu adıma budama adımı denir ve bu adımı gerçekleştirmek için MARS bir geriye doğru adım algoritması gerçekleştirir. Geriye adım algoritması gerçekleştirebilmek için aşağıdaki adımları yerine getirilmelidir. Modele yalnızca temel fonksiyon sabitini içeren en küçük modele başlanır.

- Tahmin hatasını minimize eden en uygun ölçümle her değişken ve mümkün bütün düğümler için temel fonksiyon uzayı araştırılır.
- Maksimum karmaşıklığa ulaşan bir modeli oluşturuncaya kadar bütün düğümlere ve değişkenlere bakma işlemi tekrarlanır.
- Son olarak, budama prosedürüne başvurulur.

Oluşturulan maksimum model budanarak yani önemli bağımsız değişkenler ve bu değişkenlerin interaksyonları belirlenerek, hata kareler toplamı minimum olan optimum model oluşturulur. Budama algoritması en yaygın olarak genelleştirilmiş cross validation tekniği (genelleştirilmiş hata kareler ortalaması, GCV) ile yapılır. Genelleştirilmiş cross validation hem artıkların hatasını hem de model karmaşasını hesaba katar ve 4 nolu eşitlik yardımıyla hesaplanır.

$$GCV = \frac{\sum_{i=1}^N (y_i - f(x_i))^2}{\left(1 - \frac{C}{N}\right)^2}$$

$$C = 1 + cd \quad (4)$$

Burada, N: Veri setindeki denek sayısını, d: Etkili serbestlik derecesi olup bağımsız temel fonksiyonların sayısını gösterir ve c: eklenecek temel fonksiyonların maliyet-karmaşıklık (cost-complexity) ölçüsüdür.

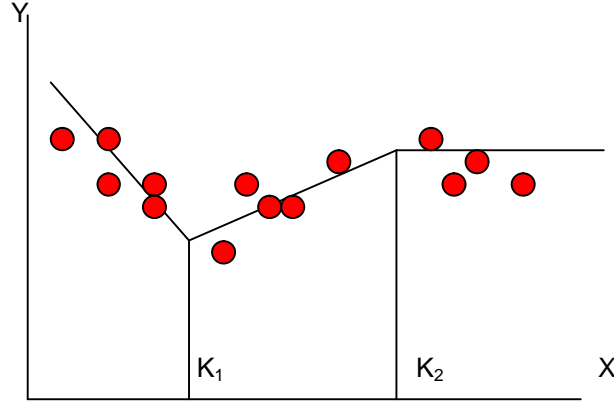
Hesaplamalar şunu göstermiştir ki en iyi C değeri için $2 < d < 3$ bulunmuştur (6).

1.3.1. Düğüm değerinin elde edilmesi:

Bağımlı ve bağımsız değişken arasındaki ilişkiler doğrusal, eğrisel ve kübik şeklinde olabilir. Aynı bağımsız değişken üzerinde, ilişkinin şeklinin değiştiği bağımsız değişken değerine düğüm değeri denir. Bir başka ifadeyle bağımsız değişken değeri, tanımlı bulunduğu aralıklarda doğrunun eğimini değiştirmeyen en son değer düğüm değeri olarak alınır. Bu düğüm değerlerinde hata kareler toplamı en küçük değerini alır.

Ardışık iki düğüm değerini birleştirerek çizilen doğrunun eğimi, Model 2'de b ile gösterilen regresyon katsayısıdır. Bu haliyle MARS modeli parçalı regresyon modeline benzer. Oluşturulan her temel fonksiyonda bağımsız değişkenden seçilen uygun düğüm değerleri bağımlı değişken ile bağımsız değişken arasında monoton transformasyonlar (logaritma, karekök...) yapmadan aralarındaki ilişkiyi doğrusal hale getirirler. Şekil 1'de iki farklı

düğüm değerine sahip bir değişkenin bağımlı değişkenle olan ilişkisi sunulmuştur. K_1 ve K_2 düğüm değerleridir. Başlangıç noktasından K_1 're kadar doğrunun eğimi aynıdır. Benzer şekilde K_1 ve K_2 arasında da doğrunun eğimi aynıdır. Fakat K_1 'den ϵ kadar uzaklaşsak bile doğrunun eğimi değişir. İşte doğrunun eğimini değiştirmeyen değer bir düğüm değeridir.



Şekil 1. X ve Y arasındaki dağılım grafiği

2. BÖLÜM:

2.1. Avantajları

- Bağımlı ve bağımsız değişkenlerin dağılımları ile ilgili herhangi bir varsayım yoktur (değişkenler kategorik veya sürekli olabilir).
- Bağımlı ve bağımsız değişkenler arasında doğrusallık varsayımı bulunmaz.
- Bağımlı değişkeni tanımlamada oluşturulan modelde hem eklemeli model hem de bağımsız değişkenlerin etkileşimine yer verilmektedir (7).
- Mars hızlıdır ve diğer doğrusal modellere göre daha az veri gerektirir (8).
- Kayıp değerlerden ve aşırı uç değerlerden çok az etkilenir(9).
- Çoklu bağlantıdan çok az etkilenir.
- Diğer çok değişkenli regresyon modellerinde yapılan simülasyon çalışmalarına göre başlangıçtaki bağımsız değişkenlerin % 90'nın modele girmesine karşın Mars modelinde daha az bağımsız değişken modele girmiştir (3).
- Modelin yorumlanması oldukça kolaydır.

2.2. Dezavantajları

- Henüz uygulaması yeni bir yöntem olmasından dolayı modelin yorumlanmasında dikkat etmek gerekir. Özellikle kayıp gözlem durumunda ve interaksiyon terimlerinin yorumlanmasında iyi bir bilgi birikimine ihtiyaç vardır.
- Uygulaması için genellikle büyük veri setlerine ihtiyaç duyulur.

3. Gereç ve Yöntem

Bu çalışmanın uygulama bölümünde, kanser ve behçet tanısı koyulmuş toplam 51 hastanın ailelerinin psikososyal durumlarına ilişkin Okyayüz ve Ünlüoğlu'nun (1998) yapmış olduğu çalışmadan alınmış birtakım ölçümler veri olarak kullanılmıştır. Bu ölçümlerde ailedeki kişilerin beç depresyon envanteri (BD) bağımlı değişken olarak alınmıştır. Yaş, durumluluk ölçeği olarak "STAT1", sürekli kaygı ölçeği olarak "STAT2", öfke kontrol ölçeği olarak "CECS-1", kaygı kontrol ölçeği olarak "CECS-2", mutsuzluk kontrol ölçeği olarak "CECS-3" ve kontrol ölçeklerinin toplamı olarak "CESCT", aile değerlendirme ölçeklerinden, problem çözme "PÇ", iletişim kurma "iletişim", roller, duygusal tepki verebilme "DUYTEP" gereken ilgiyi gösterme "GERILG", davranış kontrolü "DAVKONT", genel işlevler olarak "GENİŞL" değişkenleri ise bağımsız değişkenler olarak alınmıştır. Bağımlı ve bağımsız değişkenler belirli aralıklarda değişen puan veya toplam puanlardan oluşmaktadır. Bağımlı ve bağımsız değişkenler ve değişkenlere ait tanımlayıcı istatistikler Tablo 2'de sunulmuştur.

Tablo 2. Değişkenler ve tanımlayıcı istatistikleri

Değişkenler	Ortalama	SS	Minimum	Maksimum	N
BECKDER	14.765	9.288	3	43	51
YAS	38.333	12.227	17	62	51
STAT-1	45.745	10.971	25	71	51
STAT-2	44.725	9.525	29	69	51
CECS1	16.412	3.612	11	27	51
CECS2	16.078	3.605	9	28	51
CECS3	16.294	4.3	8	27	51
CECST	48.784	9.339	30	74	51
PROBCOZ	1.720	0.529	1	3	51
İLETİŞİM	1.889	0.548	1	3.66	51
ROLLER	1.997	0.366	1.27	2.63	51
DUYTEP	1.882	0.577	1	3.5	51
GERILG	2.379	0.414	1.42	3.57	51
DAVKONT	1.882	0.421	1.22	3.57	51
GENİŞL	1.687	0.440	1	2.58	51

SS: Standart Sapma

Bu uygulamadaki temel amaç ailelerdeki kişilerin beck depresyon puanları üzerine etkili olan etkenleri MARS modeliyle belirlemektir. Bu etkenler kapsamına alınan değişkenlerde yine bu bireylere ait birtakım davranış ve ruhsal durumu ölçen ölçek değerleridir. Hesaplamalarda MARS 2.0 paket programı kullanılmıştır.

4. Bulgular

Veri içinde çok az miktarda kayıp gözlem olduğu için bu değerler ihmal edilmiştir. Veri setinde çok sayıda kayıp gözlem olsa idi analizde kayıp gözlemler dikkate alınarak kayıp gözlemleri gösteren temel fonksiyonların modele alınması gerekirdi. Ayrıca modele alınan değişkenlerden CECST değişkeni CECS1, CECS2 ve CECS3 değişkenlerindeki puanların cebirsel toplamı olarak elde edilmiştir. Dolayısıyla bu değişken parçaları ile yüksek bir ilişki içindedir. Bu ilişki çoklu bağlantı olarak adlandırılır ve klasik regresyon modellerinde genellikle problem yaratır. Ancak MARS modelinde bu ilişki çok yüksek olmadıkça bir problem teşkil etmeyeceği daha önceki bölümde açıklanmıştı. Bu durumu göstermek amacıyla CECST değişkeni kendi parçaları ile birlikte modele alınmış ve buna ilişkin sonuçlar aşağıda açıklanmıştır. Kurulan model ana etkiler modeli olup interaksiyon terimleri modelde incelenmemiştir.

Model inşasının ilk aşamasında, ileriye doğru adım prensibi ile eklemeli maksimum model elde edilmiş ve Tablo 3'de maksimum modele ait genelleştirilmiş çapraz geçerlilik testi sonuçları (GCV), değişkenlere ait temel fonksiyonlar ve düğüm değerleri verilmiştir. Tablo 3 incelendiğinde ayna fonksiyonuna sahip değişkenler sırasıyla STAT-1, GERILG, YAŞ, DUYTEP, PROBÇÖZ, GENİŞL değişkenleridir. Ayna fonksiyonu olmayan değişkenler ise STAT-2, CECS3, CECST, CECS2 ve ROLLER'dir. Yine bazı değişkenlerin tabloda birkaç kez tekrarladığı görülmektedir. Bu ise o değişkenlere ait birden fazla düğüm noktası belirlendiğini göstermektedir. Mesela STAT-1 değişkeninin düğüm değerleri sırasıyla 39, 42, 50 ve 56 olmak üzere 4 tanedir. Düğüm değerleri kısaca şöyle yorumlanır. STAT-1 ile BD arasındaki ilişki, STAT-1' in düğüm değerlerinde farklılık göstermektedir. Bir başka ifadeyle STAT-1'in 42 ile 50 arasında değer aldığı durumda BD ile ilişkisinin şekli, 50 ile 56 arasında değer aldığı durumdan farklıdır. Bu durum diğer düğüm değerleri için de geçerlidir. Maksimum modelde sabit terimle beraber toplam 26 temel fonksiyon yer almaktadır. Modelde CECS1 değişkeni CECST değişkeni ile çok yüksek ilişki içinde olduğu için ve CECST değişkeninin modele katkısı daha yüksek olduğu için CECS1 değişkeni modelden atılmış CECST kalmıştır. Buna ilaveten İLETİŞİM ve DAVKONT değişkenleri de başlangıçta belirlenen temel fonksiyon sayısına göre modele katkısı olmadığı için modele hiç alınmamıştır. Ayrıca Tablo 3 incelendiğinde

STAT-2, CECS2, CECS3, ROLLER ve CECST değişkenlerinin düğüm değerleri 1'er tane olup bu değişkenlerin kendi değerlerinin minimumuna eşittir. Bu durum söz konusu 5 değişkenin BD ile ilişkisinin lineer yapıda olduğunu göstermektedir. Bu sonuca göre bu 5 değişkene herhangi bir transformasyon uygulanmamıştır. Maksimum model aşırı doymuş bir model olup yeni veri setlerine uyumu düşüktür.

Tablo 3. Maksimum model sonuçları

Temel Fonk.	GCV	Değişkenler	Düğüm değeri
0	87.989		
2 1	54.355	STAI_1	42
4 3	44.313	GERILG	2.28
5	40.730	STAI_2	29
6	40.658	PROBCOZ	1
7	41.570	CECS3	8
8	41.268	CECST	30
10 9	38.367	YAS	25
12 11	39.924	DUYTEP	2.5
13	37.718	GENISL	1
15 14	39.175	STAI_1	39
17 16	41.495	STAI_1	56
19 18	44.918	PROBCOZ	2.16
21 20	48.467	GENISL	1.66
22	52.186	ROLLER	1.27
24 23	57.573	STAI_1	50
25	64.378	CECS2	9

İkinci aşamada geriye doğru adım yöntemi kullanılarak, maksimum modeli daha basit bir hale getirmek için budanmıştır. Budama yöntemi olarak 10 katlı çapraz geçerlik testi tercih edilmiştir ve aşağıdaki 4 nolu tablodan oluşturulacak olan optimal (final) model elde edilmiştir. Bu modelin GCV değeri en küçüktür (GCV=33.472). Final modelin doğruluğu % 83,3'dür. Optimal modele göre modele sabit terimle beraber 12 tane temel fonksiyon girmiştir. Bu 12 temel fonksiyonlar içerisinde BF4 ve BF12 (GERILG ve DUYTEP değişkenlerine ait) ayna temel fonksiyonlarıdır. Ayrıca STAT-1 değişkeni farklı iki düğüm noktasında modele girmiştir. Final modelde bazı değişkenlere ait sadece 1 temel fonksiyon kullanılmıştır. GERILG, STAI_2,

CECS3, CECST, YAS ve GENISL değişkenleri bu grupta olup bunların düğüm değerlerinden daha büyük olan değerler modele alınmıştır. Çünkü bu değişkenlerin düğüm değerlerinden daha küçük olan değerleri için eğim sifıra eşittir. Geriye kalan değişkenlerin, modele giren temel fonksiyon sayısı ise 2'dir. Temel fonksiyon sayısının 1'den fazla olması, o değişkenin lineer bir ilişki içinde olmadığını ve belirlenen düğüm değerine göre parametrik olmayan bir transformasyonla ilişkinin lineere dönüştürüldüğünü gösterir.

Tablo 4. Optimum modele ait bilgiler

Temel Fonk.	Regresyon Katsayısı	Değişkenler	Düğüm
0	1.272		
1	2.506	STAI_1	42
3	-7.242	GERILG	2.28
4	-21.919	GERILG	2.28
5	0.228	STAI_2	29
7	1.628	CECS	8
8	-0.605	CECST	30
9	0.142	YAS	25
11	10.433	DUYTEP	2.5
12	5.094	DUYTEP	2.5
14	-1.530	STAI_1	39
20	9.926	GENISL	1.66

Bu sonuçlara göre temel fonksiyonlar ve temel fonksiyonlardan yararlanarak oluşturulan tahmin denklemi 4 nolu eşitlikteki gibidir.

$$BF1 = \max(0, STAI_1 - 42.000)$$

$$BF3 = \max(0, GERILG - 2.280)$$

$$BF4 = \max(0, 2.280 - GERILG)$$

$$BF5 = \max(0, STAI_2 - 29.000)$$

$$BF7 = \max(0, CECS3 - 8.000)$$

$$BF8 = \max(0, CECST - 30.000)$$

$$BF9 = \max(0, YAS - 25.000)$$

$$BF11 = \max(0, DUYTEP - 2.500)$$

$$BF12 = \max(0, 2.500 - DUYTEP)$$

$$BF14 = \max(0, STAI_1 - 39.000)$$

$$BF20 = \max(0, GENISL - 1.660)$$

$$Y=1.272+2.506BF1-7.242BF3-21.919BF4+0.228BF5+1.628BF7-0.605BF8+0.142BF9+ 10.433BF11+5.094BF12-1.530BF14+9.926BF20 \quad (5)$$

Bu modelin tahminlerdeki isabet derecesi (R^2) % 83.3 olarak bulunmuştur. Bu sonuç modelin oldukça başarılı tahmin yaptığını göstermektedir ($p=0.0000$). Ayrıca regresyon modelinin standart hatası ise 4.3 olarak hesaplanmıştır. Bu modelde STAT-1 değişkeninin 39'dan önceki değerleri için eğim sıfırdır, 39 ile 42 değerleri arasındaki eğimle 42'den büyük değerleri için eğim katsayıları da birbirinden farklıdır. Bu ilişkinin şekli Şekil 2'de daha açık görünmektedir.

Tablo 5'de değişkenlerin önemlilik oranları verilmiştir. Bu değerler modele alınan bütün değişkenler için hesaplanmıştır. Önemlilik değeri düşük olan değişkenin modele katkısı da azdır. Tablo incelendiğinde modele katkısı en fazla olan veya bir başka ifadeyle BD üzerine en etkili olan değişken STAT-1 değişkenidir. CECS1, PROBCOZ, ILETISIM, ROLLER ve DAVKONT değişkenlerinin önemlilik değerlerinin sıfır çıkması bu değişkenlerin BD'deki varyasyonu açıklamada anlamlı bir etkiye sahip olmadığını göstermektedir.

Tablo 5. Değişkenlerin önemlilik oranları

Değişkenler	Önemlilik
STAI_1	100.000
CECS3	72.403
GERILG	67.939
CECST	55.140
GENISL	53.168
STAI_2	31.262
DUYTEP	17.635
YAS	16.931
CECS1	0.000
CECS2	0.000
PROBCOZ	0.000
ILETISIM	0.000
ROLLER	0.000
DAVKONT	0.000

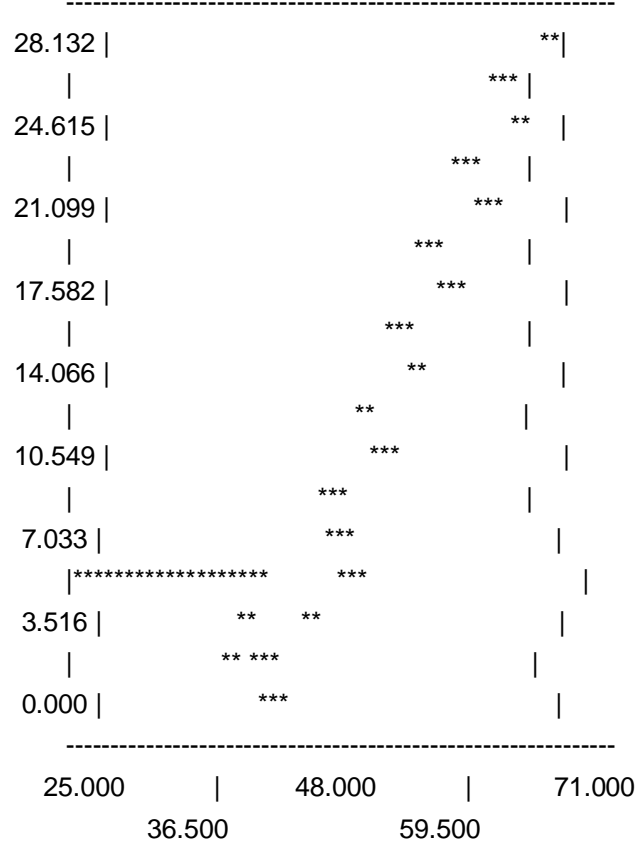
Sonuçların yorumlanmasında yardımcı olan bir diğer Tablo ise final modele giren temel fonksiyonlar (sabit terim hariç) için oluşturulan ANOVA decomposition tablosudur (Tablo 6). Bu tabloda her bir değişkenin BD üzerine genel etkisi yer almaktadır. ANOVA tablosundaki standart sapmaların büyüklüğü o değişkenin bağımlı değişken üzerine olan genel etkisinin fazla olduğunu gösterir. Mesela STAT-1 değişkeni toplam 2 temel fonksiyon ile modele girmiş ve standart sapması incelendiğinde BD üzerine en etkili olan değişkenlerden birisi olduğu görülmüştür. Bu sonuç, değişken önemlerine ilişkin Tablo 5' teki açıklamalarla paralel bir sonuçtur. Genel etkisi en düşük olan değişken ise YAŞ olarak belirlenmiştir. Ayrıca Tablo 6' nın 3. sütununda yer alan değerler, söz konusu değişkene ait temel fonksiyonları modelden çıkardığımız durumda, tahminlerde meydana gelecek kayıp göstermektedir. SS ile bu değerler zıt ilişki içindedir. SS söz konusu değişkenin kurulan modele katkısının gücünü gösterdiği için SS değeri büyük olan değişkenin modelden çıkarılması durumunda tahminlerde kötüleşme beklenir. Bu nedenle SS değeri büyük olan değişkenin kayıp değeri de büyüktür.

Tablo 6. 11 temel fonksiyon için ANOVA Decomposition sonuçları

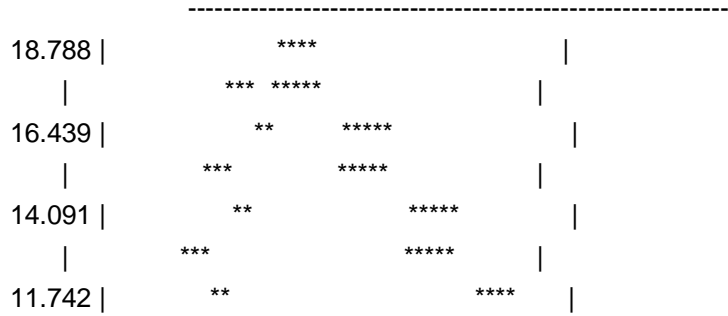
Fonksiyon	SS	Çıkarmadaki kayıp	temel fonksiyon sayısı	Etkili parametre sayısı	Değişken
1	6.582	64.180	2	3.066	STAL_1
2	4.105	47.646	2	3.066	GERILG
3	2.154	36.474	1	1.533	STAL_2
4	6.931	49.570	1	1.533	CECS3
5	5.594	42.809	1	1.533	CECST
6	1.521	34.353	1	1.533	YAS
7	2.279	34.427	2	3.066	DUYTEP
8	2.845	42.153	1	1.533	GENISL

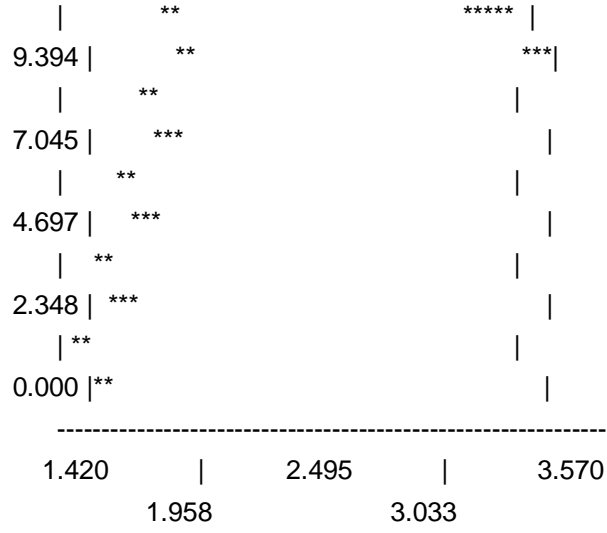
SS: Standart sapma

Bunlara ilaveten final modelde kalan değişkenlerin uygun düğüm değerleri dikkate alınarak transformasyon yardımıyla doğrusallaştırılmış hallerinin, bağımlı değişken (BD) ile ilişkilerini gösteren iki boyutlu grafikler Şekil 2 ile Şekil 9 arasında görünmektedir. Bu grafiklerde söz konusu bağımsız değişkenlerin BD ile olan ilişkilerinin şekli yer almaktadır. Ayrıca bu değişkenlere ait düğüm değerleri de grafiklerde görünmektedir.

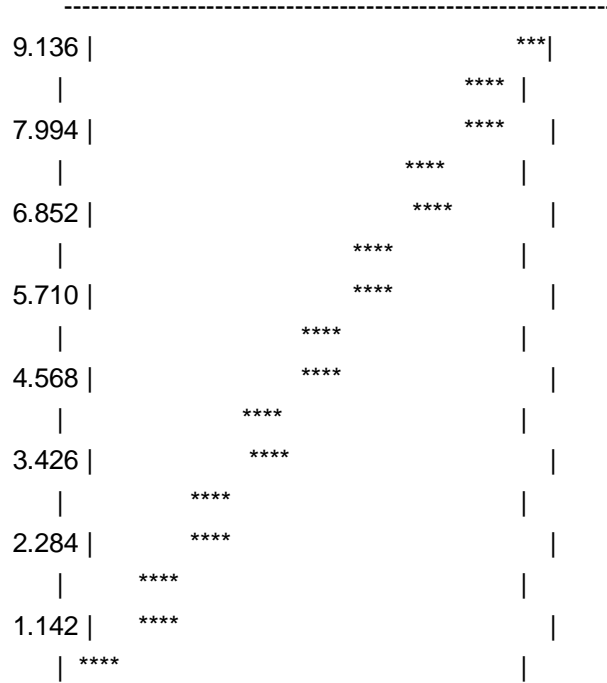


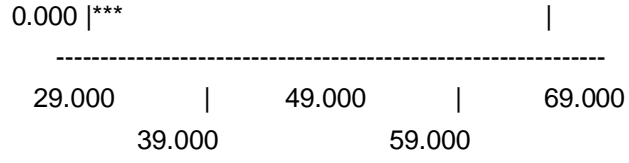
Şekil 2. STAT-1 ile BD arasındaki ilişki ve düğüm değeri



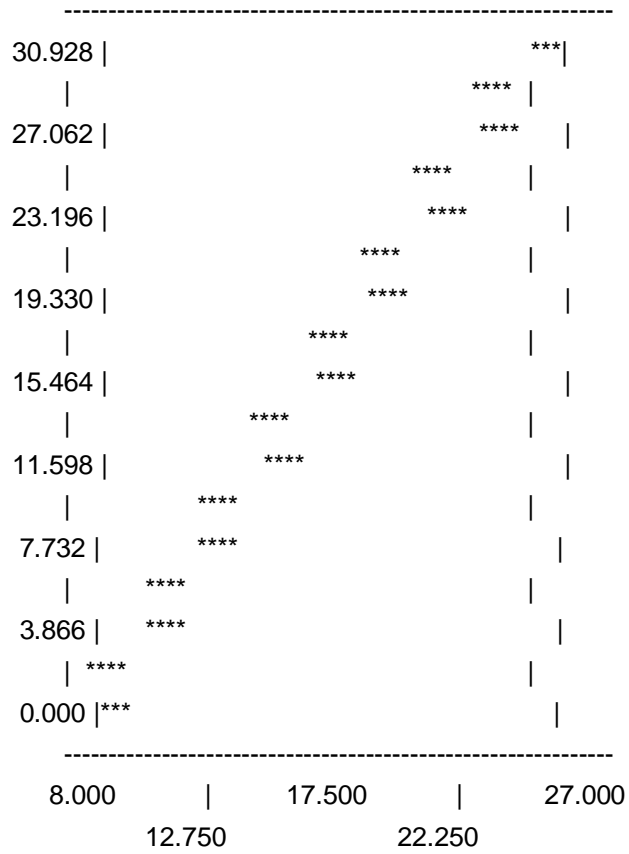


Şekil 3. GERILG ile BD arasındaki ilişki ve düğüm değeri

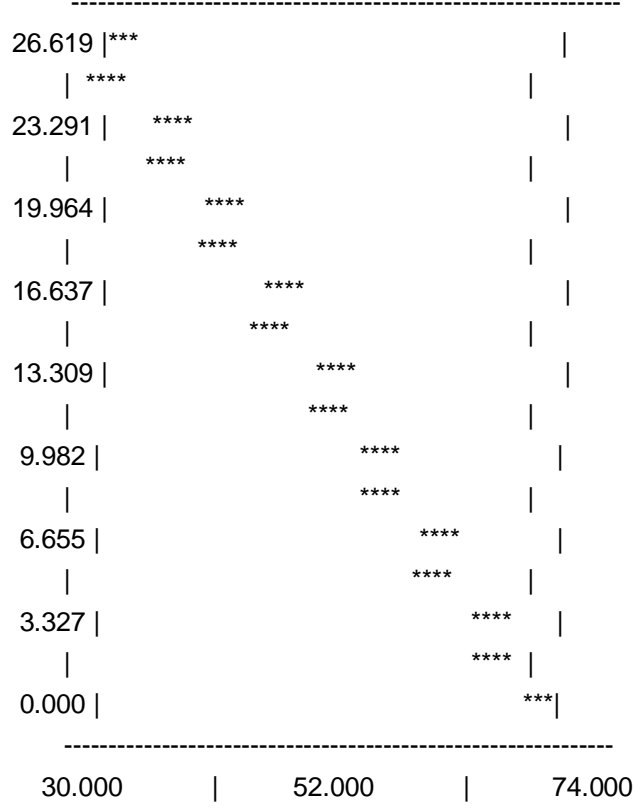




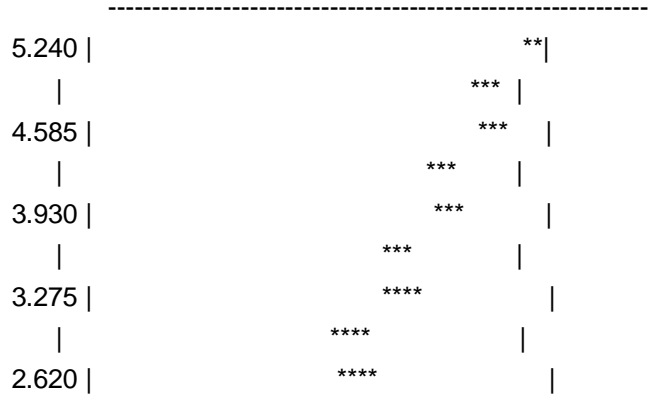
Şekil 4. STAT-2 ile BD arasındaki ilişki ve düşüm değeri

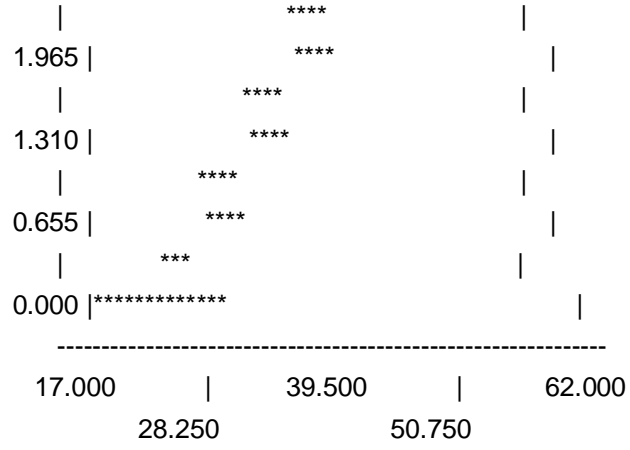


Şekil 5. CECS3 ile BD arasındaki ilişki ve düşüm değeri

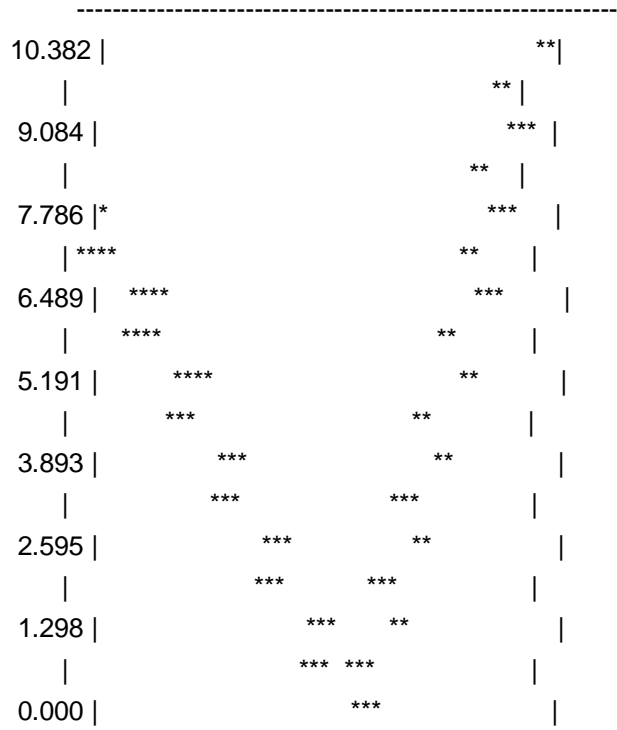


Şekil 6. CECST ile BD arasındaki ilişki ve düğüm değeri



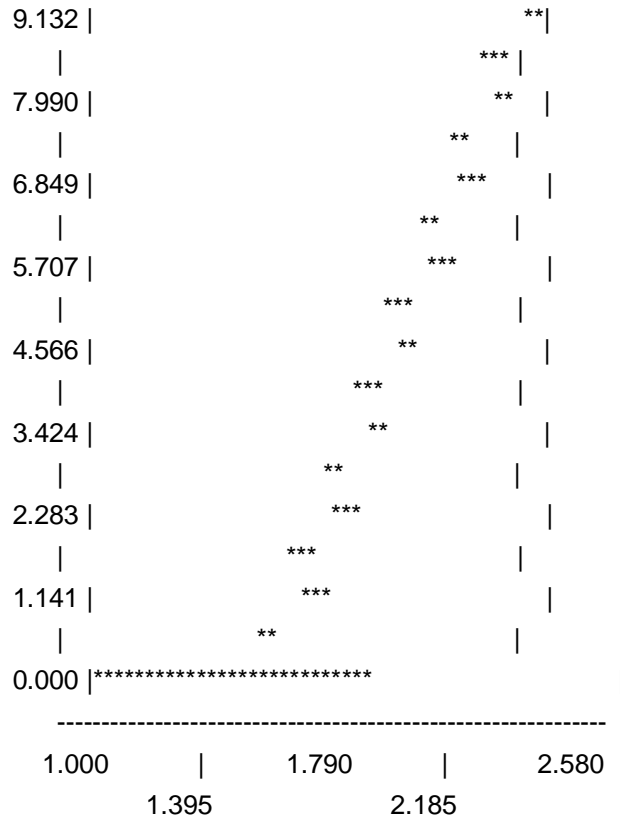


Şekil 7. YAŞ ile BD arasındaki ilişki ve düşüm değeri





Şekil 8. DUYTEP ile BD arasındaki ilişki ve düğüm değeri



Şekil 9. GENISL ile BD arasındaki ilişki ve düğüm değeri

5. Tartışma ve Öneriler

Doğrusal ilişkinin var olduğu modellerde doğrusal regresyon modeli ile MARS modelinden elde edilen eğim katsayıları aynı olmasına karşın modelin sabit terimi aynı çıkmaz.

Doğrusal modellerdeki en temel varsayım katsayı değerinin, bağımsız değişkenin bütün değerlerinde sabit kalmasıdır. Buna karşın MARS modelinde katsayı değerleri düğüm değerlerine bağlı olarak değişim gösterir. Bu durum biyolojik yapıyı daha iyi açıklama fırsatı sağlar (10).

MARS modeli doğrusallık varsayımı gerektirmeyen CART modeli ile karşılaştırıldığında CART'ın sadece interaksyon etkisini ortaya çıkardığı görülmüştür. Buna karşın MARS modeli interaksyon etkisinin yanında main effect'i ortaya çıkarır. Ayrıca MARS modelinde bir tahmin modeli elde edilirken CART analizinde görsel açıklanabilen bir şema elde edilmektedir (4).

Kaynaklar

1. Chatterjee S, Hadi A, Price B. Regression Analysis By Example. 3rd edition. Canada: John&Wiley; 2000.
2. Copyright StatSoft, Inc.; Multivariate Adaptive Regression Splines: Erişim: <http://www.statsoft.com/textbook/stmars.html>
3. Kolyshkina I., Sylvia W.; Enhancing Generalised Linear models with Data Mining: Erişim: <http://www.casact.org/pubs/dpp/dpp04/04dpp279.pdf>
4. Deichmann J, Eshghi A, Haughton D, Sayek S, Teebagy N. Application of Multiple Adaptive Regression Splines (MARS) in Direct Response Modelling. Journal of Interactive Marketing 2002; 16:15-27.
5. Draper N, Smith H. Applied Regression Analysis. 2nd edition. Canada : John&Wiley; 1981.
6. IESE; Using Multiple Adaptive Regression Splines to Understand Trends in Inspection Data and Identify Optimal Inspection Rates: Erişim: http://www.iese.fraunhofer.de/pdf_files/iese-062_00.pdf
7. Septhon P. Forecasting Recessions: Can We Do Better on MARS?. Federal Reserve Bank of St. Louis Review, 2001; 83(2), 39-49.
8. Chen I, Lee T. A Two-Stage Credit Scoring Model Using Artificial Neural Networks and Multivariate Adaptive Regression Splines 2005; 28:743-752.
9. Salford Systems, MARS User Guide, Erişim: <http://www.salford-systems.com/424.php>
10. Friedman J. Discussion. Technometrics 1991; 33: 145-148.

NEGATİF BINOM REGRESYON MODELİ

Anıl Aktaş, Osman Saraçbaşı

Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı

e-mail: anilaktas7@yahoo.com

Özet

Türkiye'de 5 bölge ve belirli yaş gruplarına göre annenin ölü doğum yapma riskini araştırmak için sayılabilir verilerden oluşan Ölü doğum sayısına Negatif Binom regresyon modeli uygulayarak bir çalışma yaptık. Bu çalışmada ortalama 72,63 varyans 2857'dir. Varyans yaklaşık olarak ortalamadan 40 kat daha büyüktür. Bu durum yayılımın çok fazla olduğunu gösterirken verilere uygun model için de bir yol göstermektedir. Yayılımın fazla olması halinde sayılabilir verilere uygulanacak Negatif Binom regresyon uygun bir yaklaşımdır. Bu nedenle Poisson regresyon yerine Negatif Binom regresyon uygulanmıştır.

Anahtar Kelimeler: Negatif Binom regresyon; sayılabilir veriler; Poisson Regresyon; aşırı yaygınlık.

Giriş

Bağımlı değişkenimiz kesikli bir dağılıma sahipse doğrusal regresyon çözümlenmesi uygulanamaz. Bu durumda nasıl bir yöntem izlenmeli?

Genel doğrusal regresyon çözümlenmesi; bağımlı değişkenin ve hata teriminin normal dağıldığı varsayımı ve diğer varsayımlar altında kullanılır. Bağımlı değişkenin kesikli dağıldığı durumda; genelleştirilmiş doğrusal modeller ailesinin özel bir durumu olan logaritmik doğrusal modeller yardımıyla Regresyon çözümlenmesi kullanılır. Bağımlı değişken $\{Y_i\}$, üstel aileden gelen, aynı dağılıma sahip, birbirinden bağımsız rastlantı değişkenleri olduğu durumlarda; bağımlı değişken ile bağımsız değişkenler arasındaki bağı doğrusal bir yapı ile açıklayan fonksiyondur. Logaritmik doğrusal modeller, Genelleştirilmiş doğrusal modeller ailesinin özel bir durumudur. Bağımlı değişken ortalaması ile bağımsız değişkenler arasındaki bağı fonksiyondur.

Sayılarla ölçülmüş değişkenlerin analizinde değişkenler negatif değer almazlar ve bu yüzden normal dağılım göstermezler. Poisson ve negatif Binom dağılımları bu tür durumlar için daha iyidir. Belli bir zaman

aralığındaki kazalar, hastaneye gelen hasta sayısı, hastalık insidansı, bir çocuğun okula gelmediği gün sayısı, bir kolonideki bakteri sayısı gibi sayılabilir verilerin analizinde ve modellenmesinde Poisson regresyon kullanılır.

Poisson dağılımında bir parametre (λ) vardır ki; bu parametre ortalama ve varyansın eşitliğini gösterir. Gerçek verilerde varyans, ortalamaya eşit değildir ve Poisson model, bu durum için uygun değildir. Poisson regresyonda iki durum bulunmaktadır. Yayılımın fazlalığı (overdispersion) ve yayılımın azlığı (underdispersion). Eğer yayılım 1'den büyükse yani varyans ortalamadan büyükse yayılım çok fazla, 1'den küçükse yani varyans ortalamadan küçükse yayılım azdır. Yayılımın fazlalığı ve azlığı Poisson modelin yetersiz olduğunu gösterir. Yayılımın fazla olması durumunda veriler Negatif Binom dağılımına uygun olduğu için Negatif Binom regresyon kullanılması daha uygundur. Bu nedenle varyansın ve ortalamanın eşitliği olan Poisson dağılımı yerine Negatif Binom dağılımı kullanılır.

Negatif Binom dağılımında;

Beklenen olgu sayısı; Y

Bağımsız değişkenler; x_1, x_2, \dots, x_q

olasılık yoğunluk fonksiyonu;

$$\Pr(Y_i = y_i; \alpha, d) = \frac{(y_i + d - 1)!}{y_i! (d - 1)!} \frac{\alpha^{y_i}}{(1 - \alpha)^{y_i + d}} ; y_i = 0, 1, 2, \dots$$

α : birim zamanda bir olgu ortaya çıkma olasılığı [$P(Y=1)$]

d : yayılım parametresi

Negatif Binom dağılımında;

Ortalama = $E(Y) = \mu_i = d\alpha$

Varyans = $V(Y) = d\alpha + d\alpha^2 = \mu_i + \mu_i^2/d$ ($\alpha \geq 0$)

Negatif Binom Regresyon çözümlemesinde parametre kestirimleri Newton-Raphson algoritması yardımıyla En Çok Olabilirlik yöntemi kullanılır. Negatif Binom regresyonun model eşitliği;

$$E(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_q X_{iq}$$

$E(Y_i)$; negatif binom dağılımına sahip herhangi Y değişkeni için birim zamanda ortaya çıkma sayısı

En çok olabilirlik tahminlerinin asimptotik dağılımı parametrelerin anlamlılık testinde ve modelin uyum iyiliğinde kullanılır.

Negatif Binom modelin anlamlılığı için iki tür test kullanılır.

1. alfa (α) nın anlamlılık testidir. Eğer α anlamlı değilse, durum Poisson dağılımına uygundur. Yani;

$H_0 : \alpha = 0$ (Aşırı yaygınlık yok)

2. Poisson ve Negatif Binom için olabilirlik oran testlerini karşılaştırmak.

$LL_{(poisson)}$ = Poisson regresyon modeli için log-olabilirlik

$LL_{(negatif\ binom)}$ = Negatif Binom regresyon modeli için log-olabilirlik

$H_0: d=0$

$H_1: d > 0$

$LR = -2[LL_{(poisson)} - LL_{(NB)}]$

LR istatistiğinin asimptotik dağılımı serbestlik derecesi 1 olan Ki-kare dağılımıdır. $LR \sim \chi^2_{(s, d=1)}$ Eğer $LR > \chi^2_{(1-2\alpha, sd=1)}$ olduğunda yokluk hipotezi reddedilir.

Daha önce yapılan çalışmalar; belli bir kavşakta meydana gelen kaza sayısı, bir okuldaki öğrencilerin okula gelmediği gün sayısı, bebek ölüm oranı, hastaneye belli mesafelerde yaşayan hastaların belli bir zaman periyodundaki hastaneye geliş sayıları gibi örnekler üzerine yapılmış ve çalışmalarda veriler sayılabilir olduğu için belli kriterlere bakılarak Negatif Binom regresyon uygulanmış.

Gereç ve Yöntem

1988 Türkiye Nüfus ve Sağlık Araştırmasından elde edilen yeni doğan çocukların ölümlerine ilişkin verilerin analizinde; çocuk ölümlerini açıklayan anne yaşı yedi düzeyle nitelendirilirken, bölgeler beş düzeyli sınıflanabilir nitel veri özelliğindedir. Bağımlı değişken yeni doğan çocukların ölümleridir.

1988 Türkiye Nüfus ve Sağlık araştırmasından elde edilen evlenmiş kadınların canlı ve ölü doğum yapmasına ilişkin verilere ilişkin olarak;

y_{ki} : yaş değişkeninin k. düzeyi ile bölge değişkenin l. düzeyindeki ölü doğan çocukların sayısı (parantezin dışındakiler)

n_{kl} : yaş değişkeninin k. düzeyi ile bölge değişkenin l. düzeyindeki canlı doğan çocukların sayısı (parantezin içindekiler)

Verilerin analizi Stata 8.0 programında incelenmiştir. Verilerin ilk olarak tanımlayıcı istatistiklerine bakıldıktan sonra hem Poisson regresyon hem de Negatif Binom Regresyon uygulanmıştır. Modelin anlamlılık testi sonuçları incelendikten sonra, tanımlayıcı istatistikleri ve anlamlılık testi sonuçlarına göre verilerin Negatif Binoma uygun olduğu görülmüştür. Sonuçlar verildikten sonra da Akaike Information Criteria (AKI) ile birlikte verilere negatif Binom regresyon uygulanmasının daha doğru olduğu desteklenmiştir.

Bulgular

Bağımlı değişken olan ölü doğan çocukların sayısı, bağımsız değişkenler bölgelere ve annelerin yaşına Tablo 1'de verilmiştir.

Tablo 1.

Anne Yaşı	Batı Anadolu	Güney Anadolu	Orta Anadolu	Kuzey Anadolu	Doğu Anadolu
15-19	1(32)	0(20)	5(40)	0(9)	11(40)
20-24	19(347)	28(232)	25(296)	2(130)	21(268)
25-29	75(777)	54(405)	75(617)	34(337)	67(527)
30-34	95(960)	78(499)	129(788)	42(335)	130(797)
35-39	127(1023)	89(576)	152(774)	41(270)	168(857)
40-44	80(784)	90(481)	181(868)	45(292)	138(661)
45-49	131(834)	82(385)	153(631)	61(314)	113(531)

İlk olarak bağımlı değişkenimizin dağılımı incelenmiştir. Poisson dağılım göstermediği bulunmuştur. Tanımlayıcı istatistiklerden ortalaması 72.62857 ve varyansı 2857.005'dir. Yayılım çok fazla olduğu için verilere Negatif binom regresyon uygulanmıştır. Stata 8.0 istatistik paket programında verilere Negatif Binom Regresyon uygulandığında $\alpha = 0.09$ değeri sıfırdan farklıdır. Veriler Negatif Binom modeline uygundur. Verilere aynı zamanda Poisson regresyon uygulanarak log-olabilirlik değerleri karşılaştırılır. Bu durumda; Negatif Binom regresyonu için log-olabilirlik değeri -129.99014, Poisson regresyonu için log-olabilirlik değeri -

132.7984799 bulunmuştur. Buradan da veriler için Negatif Binom model uygundur. AKI kriteri sonucu Negatif Binom regresyonu modeli, Poisson regresyonu modeline göre 1.0835 kere daha iyi bir modeldir.

Tablo 2. Ölü doğan çocuk sayısı

	Çeyreklikler	En küçük		
1%	0	0		
5%	0	0		
10%	2	1	Gözlem sayısı	35
25%	25	2		
50%	75		Ortalama	72.62857
		En büyük	Standart sapma	53.45096
75%	127	152		
90%	152	153	Varyans	2857.005
95%	168	168	Çarpıklık	.3104833
99%	181	181	Basıklık	1.988981

Tablo 3. Poisson ve Negatif Binom Regresyon modelleri

Ö.D.Ç	Katsayı (NB)	Katsayı (P)	S E.	z	Odds Ratio	P> z	P> z (NB)	[95% GA] (P)	
15-19	-3.477082	-3.458356	.2544096	-13.67	.0308974	0.000	0.000	-3.975716	-2.978449
20-24	-1.751448	-1.737692	.1278556	-13.70	.1735224	0.000	0.000	-2.002041	-1.500856
25-29	-5.704967	-5.712574	.0949519	-6.01	.5652446	0.000	0.000	-.7565991	-.384394
30-34	-.144198	-.1303618	.0886706	-1.63	.8657163	0.104	0.038	-.3179891	.0295931
35-39	.0431999	.0662731	.0868082	0.50	1.044147	0.619	0.268	-.1269409	.2133408
40-44	-.0349791	-.0111733	.0875971	-0.40	.9656256	0.690	0.855	-.2066661	.136708
45-49
Batı A.	-2.085493	-.2047944	.082279	-2.53	.811761	0.011	0.000	-.3698132	-.0472854
Güney A.	-4.203989	-.4312579	.0851806	-4.94	.6567848	0.000	0.000	-.5873498	-.2534479
Orta A.	.0963986	.1053605	.0790888	1.22	1.101198	0.223	0.052	-.0586126	.2514098
Kuzey A.	-1.067058	-1.05779	.0969028	-11.01	.3440193	0.000	0.000	-1.256984	-.877131
Doğu A.
sabit	4.940319	4.924753	.0804136	61.44		0.000	0.000	4.782711	5.097926
lnalfa	-4.712223	.6632792				-6.012226	-3.4122		
alfa	.0089848	.0059594				.0024486	.0329679		

Bu sonuçlara göre 15-19 yaş grubundaki annelerin ölü doğum yapma riski 45-49 yaş grubundaki annelere göre % 97 daha azdır. 20-24 yaş

grubundaki annelerin ölü doğum yapma riski % 83, 25-29 yaş grubundaki annelerin ise % 43 daha azdır. Batı Anadolu' da ölü doğum yapma riski Doğu Anadolu'ya göre % 19 daha az iken, Güney Anadolu' da % 34, Kuzey Anadolu'da ise % 66 daha azdır.

Tartışma

Sayılabilir veriler Normal dağılım göstermediğinden Doğrusal Regresyon uygulaması doğru değildir. Bu durumda uygulanan regresyon çözümlemesi verilerin yayılımına bağlı olarak farklılık gösterir. Verilere ilişkin olarak hem Poisson hem de Negatif Binom Regresyon uyguladık. Regresyon çözümlemesine ilişkin bulgular Tablo 3. belirtilmiştir. Hangi modelin daha iyi sonuç verdiği üzerine bazı testler yapılmıştır. Tablo 3 belirtildiği gibi katsayıların anlamlılığı her iki regresyon tipinde benzerlik göstermesine rağmen, alfa katsayısının değeri ve olabilirlik oran testlerine göre Negatif Binom regresyonun daha iyi doğru olduğu görülmüştür.

Kaynaklar

1. Byers AL, Allore H, Gill TM, Peduzzi PN. Application of negative binomial modeling for discrete outcomes A case study in aging research. Journal of Clinical Epidemiology 2003;56:559-564.
2. Hardin J, Hilbe J . Generalized Linear Models and Extensions. America: Stata Press; 2001. 139-157
3. Negative Binomial; Testing for over dispersion in Poisson regression. http://www.uky.edu/ComputingCenter/SSTARS/P_NB_3.htm
4. Poisson regression Overview; http://www.uky.edu/ComputingCenter/SSTARS/P_NB_1.htm
5. Poisson and negative Binomial Regression; UCLA Academic Technology services <http://www.ats.ucla.edu/stat/stata/stat130/count2.htm>
6. Negative Binomial Regression Models; U:S Department of Transportation, Federal Highway Administration <http://www.tfhr.gov/safety/pubs/02089/appb.htm>

POPULASYON BÜYÜME ORANININ DOĞRUSAL OLMAYAN BİR REGRESYON MODELİ İLE TAHMİNİ

Gürol Zırhloğlu¹, Kazım Kara²

Yüzüncü Yıl Üniversitesi, Van Meslek Yüksekokulu, Bilgisayar Tek. ve Prog. Bölümü¹

Adnan Menderes Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü²

e-mail: gurolzirhli@yyu.edu.tr

Özet

Populasyon büyüme oranı zaman ve populasyon miktarındaki değişimlerden etkilenen kavramlardır. Populasyon büyüklüğünün belirlenmesi için kullanılan modeller doğum, ölüm ve geçerli populasyon büyüklüğü arasındaki ilişki ile belirlenir. Bu modeller doğrusal model değildir. Bu modellere ait regresyon katsayılarını hesaplamak için uygun dönüşüm yöntemleri kullanılmaktadır. Bu çalışmada doğrusal olmayan bir regresyon modeli kullanılarak yoğunluğa bağlı olmayan populasyon büyüme oranı ve t-1 zamanındaki populasyon büyüklüğü tahminlenmiştir.

Anahtar kelimeler: Populasyon büyüme oranı, dorusal olmayan regresyon modeli, üssel model.

Giriş

Populasyon büyümesinin incelenmesi demografik çalışmaların ana konularından birisidir. Çevre koşullarına bağlı olarak populasyon miktarında değişimler meydana gelebilir. Bu durum uygun koşullarda populasyon miktarında artış şeklinde ortaya çıkarken uygun olmayan koşullarda ise azalma şeklinde ortaya çıkabilmektedir. Uygun çevre koşulu durumu ise canlı türlerine göre farklılık gösterebilen bir kavramdır. Populasyon büyümesi doğum, ölüm ve göç olaylarının etkisi altındadır. Populasyon büyüme oranı zaman ve populasyon miktarındaki değişimden etkilenebilen bir değerdir (Krebs,1994; Akçakaya ve ark., 1999; Krivan and Havelka, 2000).

Populasyon çalışmalarında veri ve teori arasında ilişki kurmak için kantitatif modeller kullanılır. Bu modellerin bazıları matematiksel olup, çoğunlukla basitleştirilerek populasyon davranışlarını açıklamak için kulla-

nılır. Bazı modeller ise istatistiksel ve genellikle populasyon miktarının tahmini, yaşam oranı veya farklı modellerin karşılaştırılması gibi işlemleri gerçekleştirmek için kullanılmaktadırlar (Mills ve ark., 1999).

Populasyon büyüme oranının belirlenebilmesi için kullanılan modeller doğum, ölüm ve geçerli populasyon büyüklüğü arasındaki ilişki ile belirlenir. Populasyon büyümesi yoğunluğa bağlı büyüme ve yoğunluğa bağlı olmayan büyüme olmak üzere 2 grupta incelenebilir. Yoğunluğa bağlı olmayan büyüme modelleri sürekli zaman modelleri ve kesikli zaman modelleri şeklinde gruplandırılabilir. Üssel büyüme ve logisitit bu zamanlara ait temel modellerdir (Rikfles, 1993; Krebs, 1994; Shuterland, 1996; Akçakaya ve ark., 1999).

Populasyon büyüme oranı, R , çeşitli ekolojik ve demografik parametrelerin hesaplanmasında kullanılan önemli bir değerdir. Bu değer herhangi bir t ve $t+1$ zamanları arasında meydana gelen değişimi ifade eder. Eğer $R < 1$ ise popülasyonda azalma olduğu, $R = 1$ ise popülasyonun dengede olduğu ve $R > 1$ ise popülasyonda artış olduğu sonuçları elde edilir (Krebs, 1994; Gotelli, 2001).

Populasyon büyüme oranının belirlenmesinde t zamanındaki birey sayısı, $t+1$ zamanındaki birey sayısı ve populasyon büyüme oranının belirleneceği zaman aralığı değerleri kullanılmaktadır. Herhangi bir t zamanındaki populasyon büyüklüğü bağımlı değişken, zaman bağımsız değişken, populasyon büyüme oranı ve populasyon başlangıç değeri ise, parametrelerdir (Akçakaya ve ark., 1999). İki değişken arasındaki ilişki her zaman doğrusal olmayabilir. Doğrusal modellerde her parametrenin üssünün 1 olması, parametrenin modellerde üs olarak yer almaması ve parametrelerin diğer parametreler ile çarpılıp bölünmesi söz konusudur (Alpar, 1997).

Doğrusal olmayan modellerde bağımlı değişken bağımsız değişkenlerin doğrusal bir fonksiyonu değildir. Regresyon analizindeki standart varsayımlardan biri tanımlanan verilere ait modelin doğrusal olmasıdır. Ancak bağımlı değişken ile bağımsız değişken arasındaki ilişki her zaman doğrusal olamadığı için, doğrusal olmayan modeller uygun dönüşümler ile doğrusal yapılabilirler (Weisberg, 1985; Düzgüneş ve ark., 1987; Chatterje, 1991; Alpar, 1997).

Yapılan çalışmada, yoğunluğa bağlı olmayan populasyon büyüme oranı ve $t-1$ zamanındaki populasyon büyüklüğü doğrusal olmayan bir regresyon modeli kullanılarak tahminlenmiştir.

Materyal ve Yöntem

Çalışmada kullanılan yöntemlerin daha iyi anlaşılabilmesi amacıyla Devlet İstatistik Enstitüsü kaynaklarından elde edilen ve 1955-1990 yıllarına

ait nüfus bilgileri kullanılmıştır (DİE, 2000). Bu veriler Çizelge 1'de belirtildiği gibidir.

Çizelge 1. 1955-1990 yıllarına ait nüfus verileri

Zaman (1955-1990)	Birey Sayısı (N _x)
1	24064763
2	27754820
3	31391421
4	35605176
5	40347719
6	44736957
7	50664458
8	56473035

Populasyon büyüme oranı

Doğum ve ölüm olayları birey sayısı olarak belirtilmekten ziyade oran olarak belirtilirler. Artan populasyon büyüme oranı R ile gösterilir ve bir populasyona ait sınırlı büyüme oranı olarak ifade edilir. Yoğunluğa bağlı olmayan üssel büyüme kesikli zamanlar için,

$$N(t) = N(0)R^t \quad (1)$$

eşitliği ile elde edilir (Gotelli, 2001). Burada $N(t)$, t zamanındaki populasyon büyüklüğünü, $N(0)$ ise, başlangıç zamanındaki populasyon büyüklüğünü ifade etmektedir. Bu eşitlik dikkate alınarak populasyon büyüme oranının belirlenebilmesi için

$$R = \left(\frac{N(t)}{N(0)} \right)^{\frac{1}{t}} \quad (2)$$

ifadesi kullanılır. Sürekli zamanlar için yoğunluğa bağlı olmayan üssel büyüme ise,

$$N(t) = N(0)e^{rt} \quad (3)$$

ifadesi ile belirtilir (Gotelli, 2001). Burada r anlık büyüme oranı olup aşağıdaki şekilde elde edilir.

$$r = \frac{\ln N(t) - \ln N(0)}{t} \quad (4)$$

İki değerln logaritmaları arasındaki fark, bu iki değerin oranlarının logaritmasına eşit olduğuna göre yukarıdaki eşitlik,

$$rt = \ln \left(\frac{N(t)}{N(0)} \right) \quad (5)$$

olarak yazılabilir. Elde edilen bu ifade üssel olarak yeniden yazıldıktan sonra her iki tarafının t inci dereceden karekökü alındığında,

$$e^r = \left(\frac{N(t)}{N(0)} \right)^{\frac{1}{t}} \quad (6)$$

sonucu elde edilir. Bu denklem populasyon büyüme oranı R 'ye eşittir (Gotelli, 2001).

Doğrusal olmayan regresyon modeli

Populasyon büyüme modelinde yer alan $N(t)$ zamana bağımlı bir değişkendir. $N(0)$ ve R ise parametrelerdir. Bu model üssel bir ifade olduğu için doğrusal olmayan bir modeldir.

$$Y = ab^x \quad (7)$$

şeklinde belirtilen bir regresyon modelinde katsayıların hesaplanabilmesi için denklem logaritmik bir dönüşüme tabi tutulur (Düzgüneş ve ark., 1987; Chatterje, 1991; Lewis-Beck, 1993; Alpar, 1997; Gotelli, 2001; Gotelli ve Ellison, 2004). Bu durumda denklem,

$$\log Y = \log a + x \log b \quad (8)$$

şeklinde yeniden yazılabilir. Burada $\log Y=Z$; $\log a=A$ ve $\log b=B$ şeklinde düşünülürse,

$$Z = A + Bx \quad (9)$$

eşitliği yazılabilir. Elde edilen bu eşitlik basit bir doğrusal denklemdir. Buna göre B katsayısı,

$$B = \frac{\sum xZ - \frac{\sum x \sum Z}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} \quad (10)$$

eşitliği ile elde edilebilir (Düzgüneş ve ark., 1987; Alpar, 1997). B katsayısı elde edildikten sonra denklemde yer alan A katsayısının hesaplanması için,

$$A = \bar{Z} - B\bar{x} \quad (11)$$

eşitliği kullanılır. Regresyon katsayılarının elde edilmesinden sonra bu katsayıların anti logaritmaları hesaplanarak esas denklem elde edilir.

Yapılan çalışmada populasyon büyüme modeli olan,

$$N(t) = N(0)R^t$$

eşitliği logaritmik dönüşüme tabi tutularak,

$$\log N(t) = \log N(0) + t \log R \quad (12)$$

denklemi elde edilmiştir (Akçakaya ve ark., 1999; Gotelli, 2001). Daha sonra $\log N(t)=Z$; $\log N(0)=A$ ve $\log R=B$ şeklinde ifade edilerek 9 numaralı denklem elde edilmiştir. Elde edilen doğrusal modeldeki katsayılar hesaplandıktan sonra, B değerinin anti logaritması yaklaşık populasyon büyüme oranını, A değerinin anti logaritması ise $t-1$ zamanındaki populasyon büyüklüğüne ait tahmini vermiştir.

Bulgular

Çizelge 1'de verilen değerlere göre populasyon büyüme oranı 2 numaralı eşitlik ile hesaplanarak $R = 1.1125$ olarak hesaplanmıştır. Populasyon büyüme oranı modeli 1 numaralı denkleme göre yazılarak

$$N(1990) = N(1955)R^8$$

eşitliği elde edilmiştir. Bu eşitlik doğrusal olmayan bir regresyon modeli olup bu modele göre çözüm yapabilmek için hesaplanan değerler Çizelge 2'de belirtilmiştir.

Çizelge 2. Doğrusal olmayan regresyon modeli katsayıların hesaplanmasında kullanılan değerler

Zaman (1955-90) (x)	Birey Sayısı (Y)	log(Y)=Z	xZ	x ²
1	24064763	7.381	7.381	1
2	27754820	7.443	14.886	4
3	31391421	7.497	22.491	9
4	35605176	7.552	30.208	16
5	40347719	7.606	38.030	25
6	44736957	7.651	45.906	36
7	50664458	7.705	53.935	49
8	56473035	7.752	62.016	64
Toplam	36	60.587	274.853	204
Ortalama	4.5	7.573		

Çizelge 2'de belirtilen değerlere göre regresyon denklemine ait B katsayısı 0.053 olarak elde edilmiştir. A katsayısı ise 7.3345 olarak hesaplanmıştır. Elde edilen bu değerlere göre regresyon denklemi,

$$Z = 7.3345 + 0.053 x$$

olarak düzenlenmiştir. Esas denklemin elde edilebilmesi için belirtilen bu katsayıların anti logaritmaları hesaplanarak denklem üssel formda yazılmıştır. Buna göre,

$$Z = (21602300)(1.129)^x$$

eşitliği elde edilmiştir. Bu denklem populasyon büyüme modeline göre yeniden yazılarak,

$$N(t) = (21602300)(1.129)^t$$

ifadesi elde edilmiştir. Buna göre, 21602300 değeri t değeri sıfır olduğunda $N(t)$ 'nin alacağı değerdir. Bu da yaklaşık olarak $t-1$ zaman adımında yer alan yaklaşık popülasyon büyüklüğünü göstermektedir. Modelin diğer katsayısı olan 1.129 değeri ise yaklaşık popülasyon büyüme oranı (R) değeridir.

Tartışma

Popülasyon büyüme oranı değerinin elde edilebilmesi için üssel büyüme modeli popülasyon çalışmalarında sıklıkla kullanılmaktadır (Krebs, 1994; Akçakaya ve ark, 1999). Çalışmada üssel büyüme modeli yerine doğrusal olmayan bir regresyon modeli kullanılarak yaklaşık popülasyon büyüme oranı değeri ve $t-1$ zaman adımındaki yaklaşık popülasyon büyüklüğü değerleri elde edilmiştir. Devlet İstatistik Enstitüsü tarafından belirtilen 1955-1990 yılları arasındaki nüfus değerlerine göre üssel büyüme modeli ve doğrusal olmayan regresyon modeli ile elde edilen büyüme oranı değerleri birbirlerine yakın sonuçlar vermiştir. Büyüme oranı değerlerinin 1'den büyük olması ($R>1$) popülasyon miktarında artış olduğunu göstermiştir (Gotelli, 2001).

Üssel büyüme modelinde ilk sayım yılındaki popülasyon büyüklüğü ile son sayım yılındaki popülasyon büyüklüğü değerleri kullanılarak popülasyon büyüme oranı değeri elde edilirken, doğrusal olmayan regresyon modeli kullanılarak yapılan hesaplamalarda bütün sayım yıllarına ait veriler kullanılmaktadır. Aynı şekilde tüm sayım yıllarındaki değerler kullanılarak $t-1$ zamanındaki yaklaşık popülasyon büyüklüğü değeri de tahminlenmiştir.

Elde edilen regresyon denkleminde 21602300 değeri başlangıç değeri olarak dikkate alınırsa t 'nin alabileceği herhangi bir değere göre istenilen zaman adımındaki nüfus miktarı tahminlenebilir. Eğer $t>0$ ise gelecek zaman adımlarındaki popülasyon büyüklüğü, $t<0$ ise başlangıç değerinin yer aldığı zaman adımından önceki zamanlara ait popülasyon büyüklüğü değerleri tahminlenebilmektedir. Ayrıca çalışmada kullanılan zaman değerleri için yapılan hesaplamalarda doğrusal olmayan modelle elde edilen nüfus büyüklüğü değerlerinin gerçek değerlere daha yakın sonuçlar verdiği gözlenmiştir.

Kaynaklar

7. Akçakaya HR, Burgman MA, Ginzburg LR. Applied Population Ecology. Sunderland; 1999.
8. Alpar R. Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş – I. Kültür Ofset, Ankara; 1997.

9. Chatterje S. Regression Analysis By Example. New York; 1991.
10. DİE. İstatistik Göstergeler, 1923-1998. T.C. Başbakanlık Devlet İstatistik Enstitüsü. Ankara; 2000.
11. Düzgüneş O, Kesici T, Kavuncu O, Gürbüz F. Araştırma ve Deneme Metodları (İstatistik Metodları II). Ankara Üniversitesi Basımevi, Ankara; 1987.
12. Gotelli NJ. A Primer of Ecology. 3rd Edition. Sinauer Associates, Inc. USA; 2001.
13. Gotelli NJ, Ellison AM. A Primer of Ecological Statistics. Sinauer Associates, Inc. USA; 2004.
14. Krebs CJ. Ecology. HarperCollins, New York; 1994.
15. Krivan V, Havelka J. Leslie Model for Predatory Gall-midge Population. Ecological Modelling 2000; 126:73-77.
16. Lewis-Beck MS. Regression Analysis (International Handbooks of Quantitative Applications in the Social Science, Vol.2). Sage, London; 1993.
17. Mills LS, Doak DF, Wisdom MJ. Reliability of Conservation Actions Based on Elasticity Analysis of Matrix Models. Conservation Biology 1999; 13:815-829.
18. Rikfles RE. The Economy of Nature: A Textbook in Basic Ecology. WH Freeman, England; 1993.
19. Shuterland WJ. Ecological Census Techniques: A Handbook. Cambridge, UK; 1996.
20. Weisberg S. Applied Linear Regression. John Wiley & Sons, Minnesota; 1985.

MERSİN' DE YAŞAYAN ÇOCUKLARDA İNTİHAR GİRİŞİM SIKLIĞI İÇİN POISSON REGRESYON MODELİNİN KULLANIMI

Handan Ankaralı (Çamdeviren)¹, Mehmet Mendes², Zeki Akkuş³,
Sema Erden¹, Semra Erdoğan¹, Siddik Keskin⁴, Fevziye Toros⁵

Mersin Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, MERSİN¹
Çanakkale Onsekiz Mart Üniversitesi, Ziraat Fakültesi, Biyomeri-Genetik AD,
ÇANAKKALE²

Dicle Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, DİYARBAKIR³
Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi, Biyomeri-Genetik AD, VAN⁴

Mersin Üniversitesi, Tıp Fakültesi, Çocuk Psikiyatrisi AD, MERSİN⁵
e-mail: hcamdeviren@yahoo.com

Özet

Bu çalışmanın amacı, okulda fiziksel ve/veya sözel şiddete maruz kalan çocuklarda intihar girişimini etkileyen bazı etkenlerin Poisson regresyon modeli yardımıyla belirlenmesidir. Çalışmada, Poisson regresyon modelinin teorik özellikleri açıklanmış ve uygulama amacıyla Mersin' de okuyan 12-20 yaş grubundaki 2143 çocuktan elde edilen veriler kullanılmıştır. Örneklemin seçiminde çok aşamalı tabakalı kümeleme örnekleme yöntemi kullanılmıştır. Çocuğun yaşı, Beck depresyon durumu, ailedeki doğum sırası ve ailesi ile problem yaşayıp yaşamadığı dikkate alınarak intihar girişim sayıları kayıt edilmiştir. Bu 4 risk faktöründen ikisinin intihar girişimini anlamlı düzeyde etkilediği gözlenmiştir. Bunlar çocuğun depresyonlu olması ve ailesi ile problem yaşamasıdır. Bu etkiler incelendiğinde çocuğun depresyonlu olmasının intihar girişimini % 19 artırdığı (P=0.009) ve ailesi ile problem yaşamasının ise intihar girişimini % 57 artırdığı (P=0.0001) gözlenmiştir. Sonuç olarak, intihar girişimine ait verilerin analizinde Poisson regresyon modelinin uygun bir model olduğu açıklanmıştır.

Anahtar Kelimeler: Ergin çocuk, sayımla elde edilen veriler, poisson regresyon modeli, hız oranı, intihar girişimi.

Giriş

Bazı gözlemsel veya deneysel çalışmalarda bağımlı değişken sayılarak elde edilen verilerden oluşur ve bu değişken Poisson dağılımı gösterir. Poisson dağılımı gösteren değişkenlere şöyle örnekler verilebilir: Doktorların ziyaret sayısı, nadir gözlenen hastalıkların görülme sıklığı, suç oranı, çocuğun okula gitmediği gün sayısı, bakteri koloni sayısı. Genellikle bu tip bağımlı bir değişken ile açıklayıcı değişkenler arasındaki ilişki, bilinmeyen parametrelerin bir lineer olmayan fonksiyonu şeklinde tanımlanır. Bu fonksiyon Poisson regresyon modeli (PRM) olarak bilinir ve bu model Genelleştirilmiş Lineer Modeller (GLM) ailesinden olup pratikte kullanım sıklığı yine bu aileden olan lojistik regresyondan sonra gelir (1, 2, 3).

Poisson dağılımı aşağıdaki özelliklere sahip olduğu için model parametrelerinin tahmininde en küçük kareler yöntemi kullanılmaz.

1. Poisson dağılımı eğri bir dağılımdır buna karşılık lineer regresyonda hataların dağılımının simetrik olması gerekir.
2. Poisson dağılımı negatif olmayan bir dağılımdır, buna karşılık lineer regresyonda bazen tahmin değerleri negatif değer olabilir.
3. Poisson dağılımında ortalama ve varyans eşit olduğu için ortalamanın artması varyansın da artması demektir ancak lineer regresyonda hataların varyansının sabit olduğu varsayılır (1, 4).

Poisson regresyon modeli, logaritmik transformasyon kullanarak dağılımın eğikliğini düzeltir ve negatif değerlere karşı korur.

PRM'de, Deviance (G^2 istatistiği) ve Pearson Chi-Square istatistikleri modelin overdispersion (varyans ortalamadan daha büyük ise) veya underdispersion durumunu ölçmek için kullanılır. Underdispersion veya overdispersion, PRM'nin tahminlerinin uygun olup olmadığını gösterir. Eğer sayılarak elde edilen verilerin dağılımında bir overdispersion var ise PRM'den elde edilecek tahminler tutarlı ancak yetersiz olacaktır (5, 4, 3). Overdispersion durumunda negatif binomial regresyon modeli tercih edilmektedir (6).

Toplumda intihar girişim sıklığı nadir gözlenen olaylar kapsamında olduğu için Poisson dağılımı gösterir (7, 8). Çocuk ve ergenlerdeki intihar girişimi, önemli bir sosyal problem olmasından dolayı sık çalışılan konular arasında yer almaktadır (9, 10). 1990'lı yıllarda Amerika'da çocuk ve ergenlerde kaza ile ortaya çıkan ölümlerin ikinci nedeninin intihar olduğu açıklanmıştır (11) ve Kore' de ise sıralamada üçüncülüğü almıştır (12).

Bu çalışmanın amacı, Poisson regresyon modeli yardımıyla, Mersin'de okuyan ve okulda fiziksel veya sözel şiddete maruz kalmış çocuk ve ergenlerdeki intihar girişim sıklığını etkileyen bazı risk faktörlerinin incelenmesidir.

Gereç ve Yöntem

Çalışmada kullanılan veriler, Mersin ilinde okuyan 10-20 yaş grubundaki çocuk ve ergenlerden seçilen bir örneklemden elde edilen bazı sosyo-demografik bilgileri içermektedir. Çalışma kesitsel (cross-sectional) olup veriler 2002 yılında elde edilmiştir. Bu yaş grubunda toplam 81676 birey bulunmaktaydı ve bunlardan 4256'sı örneklem olarak seçilmiştir. Bu örnek % 99 güvenle toplumun % 5.5'ini temsil etmektedir. Ayrıca cinsiyet ve sosyo-ekonomik duruma göre tabakalama yapılmıştır. Veri kontrol aşamasında 113 deneye ait veri uygunsuz cevap olarak görüldüğü için örneklemden çıkarılmıştır. Geriye kalan 4143 çocuktan 2143'ü ise okulda fiziksel veya sözel şiddete uğramıştır. Bu çalışmada, bu 2143 çocuktan elde edilen veriler kullanılarak bazı risk faktörleri incelenmiştir.

PRM' nin genel gösterimi (1) nolu eşitlikteki gibidir.

$$c/N = e^{\alpha + \beta_i x_i + \varepsilon_p} \quad (1)$$

Bu eşitlikte; c intihar girişim sayısını, N populasyon büyüklüğünü, α sabit terimi, β_i i . açıklayıcı değişkenin regresyon katsayısını, x_i , i . açıklayıcı değişkeni, ε rasgele hatayı veya örnekleme hatasını ve $e=2.718...$ doğal logaritma tabanını göstermektedir. En genel kullanılan bağlantı (link) fonksiyonu logaritmdir ve bu **canonical link** olarak, $\log(N)$ ise offset olarak adlandırılır.

Regresyon katsayılarının exponential değeri açıklayıcı değişkenlerin Hız Oranını (RR) verir ($RR_i = \exp(\beta_i)$). Açıklayıcı değişken veya risk faktörlerinin regresyon katsayılarının yorumlanmasında RR değerleri oldukça kullanışlıdır. Bu değerler, risk faktörünün risk taşıyan seviyesinin referans seviyesine göre intihar girişiminin insidansını verir.

Bir veri setinde npq (p : intihar girişiminde bulunanların oranı, q : intihar girişiminde bulunmayanların oranı ve n : örneklem genişliği) 5'ten küçük ise dağılım Poisson dağılımı olur.

Poisson regresyon modelinde, intihar girişiminde bulunanların ortalama sayısı aynı zamanda varyansını verir. Bununla beraber gerçek verilerde genellikle varyans ortalamadan daha büyük çıkmaktadır. Veri içinde çok sayıda sıfır değeri yer aldığı zaman böyle bir sonuçla karşılaşılmaktadır (13). Daha öncede değinildiği gibi böyle durumlarda ya negative binomial regresyon modeli kullanılır ya da katsayıların standart hataları Dispersion Phi katsayısı ile düzeltilir.

Pearson χ^2 ve Deviance G^2 testleri PRM' nin tahminlerinin uyum iyiliğini belirler. G^2 veya Pearson chi-square istatistik değeri (n-p) (n: gözlem sayısı, p: modeldeki parametre sayısı) serbestlik dereceli tablo χ^2 değerinden küçük ise PRM'nin tahminlerinin uygun olduğuna karar verilir (14).

Çalışmada kullanılan veriler, yaşa göre gruplandırılmış ve çocuk/ergenlerin depresyon sıklığı ise Child Beck Depression Inventory (CBDI) göre belirlenmiş ve depresyon puanı 19'un altında olanlar depresyonsuz, üstünde olanlar ise depresyonlu olarak kodlanmıştır (15). Ayrıca çocukların ailedeki doğum sırası ve ailesiyle problem yaşayıp yaşamadığı da belirlenmiştir. Bu durumda PRM' ye 4 risk faktörü alınmış bunlardan yaş sürekli diğerleri ise binary değişken olarak tanımlanmıştır. Bu risk faktörlerinden toplam 56 farklı kombinasyon oluşturulmuş ve alt populasyonlar olarak adlandırılmıştır. Ayrıca çalışılan örnekleme intihar girişiminde bulunan ve bulunmayan bireylerde tespit edilmiştir. Tablo 1'de Poisson modelinde kullanılan değişkenlere ait özet tanımlamalar yapılmıştır.

Tablo 1. Modeldeki bağımlı değişken ve risk faktörlerine ait tanımlayıcı bilgiler

Risk Faktörleri				Bağımlı değişken
Yaş aralığı	Beck Depresyon	Ailedeki doğum sırası	Ebeveynleri ile problem yaşama durumu	Çocuklardaki intihar girişim sıklığı
12 – 18	Var =1 Yok = 0	Birinci çocuk =1 Diğer = 0	Var =1 Yok = 0	Sayı

Hesaplamalarda EP16 INFO ve NCSS 2002 (Deneme Sürümü) istatistik paket programları kullanılmıştır Buna ilaveten EK bölümünde SAS paket programında, hem Poisson regresyon hem de negatif binomial regresyon analizlerinin program algoritmaları verilmiştir.

Bulgular

Çalışmada kullanılan verilerde okulda fiziksel veya sözel şiddete maruz kalan çocuk ve ergenlerde intihar girişiminin insidansı % 3.1 olarak hesaplanmıştır. Çalışmadaki gözlem sayısı, kombinasyon sayısına eşittir (n = 56). Sonuç olarak dağılımın ortalaması 1.68 olarak bulunur ($=npq$). Bulunan değer 5' ten küçük olduğu için intihar girişim sayılarının Poisson dağılımı gösterdiğini söyleyebiliriz.

Kurulan modele ilişkin ayrıntılı sonuçlar ise şöyledir. Çalışmada dikkate alınan yaş aralığının ve doğum sırasının intihar girişimi üzerine istatistik olarak anlamlı bir etkisinin olmadığı gözlenmiştir. Buna karşılık depresyon durumu ve ebeveynleri ile problemlili olmasının intihar girişimini etkilediği belirlenmiştir ($P<0.01$).

İki risk faktörü anlamlı olmadığı için bunlar sırasıyla modelden çıkarılarak model yeniden test edilmiştir. İlk olarak doğum sırası atılmış ve geriye kalan 3 risk faktörün etkilerinde anlamlı bir değişiklik gözlenmemiş ve log-likelihood istatistiğindeki azalma istatistik olarak anlamlı bulunmamıştır. Daha sonra modelden yaş atılmış ve modelde kalan 2 değişkenin etkilerinde önemli bir değişiklik gözlenmemiştir. Bu iki değişkenin ana etkilerini içeren model final model olarak adlandırılmıştır. Final modelin katsayıları Newton-Raphson algoritması yardımıyla tahmin edilmiştir. Bu modelin katsayıları Tablo 2'de verilmiştir. Pearson χ^2 ve Deviance (G^2) istatistiklerinin değeri hipotez kontrolü sonucunda anlamlı bulunmadığı için PMR'nin uygun bir model olduğuna karar verilir. Dispersion Phi değeri, model katsayılarının standart hataları ile çarpıldığı zaman varyans ortalamaya yaklaşacaktır.

Tablo 2. Poisson regresyon analiz sonuçları

Risk Faktörleri	B	SE	P	RR	RR için % 95 CI	
					Alt sınır	Üst sınır
Sabit terim	-4.49	0.258	0.000	0.011	0.007	0.018
Beck Depresyon (referans="Yok")	0.78	0.304	0.009	2.191	1.208	3.974
Ebevenyler ile Problem (referans="Yok")	1.72	0.321	0.000	5.572	2.970	10.452

Pearson $\chi^2 = 69.46$, SD=53 ve P=0.070
Deviance $G^2 = 59.04$, SD=53 ve P=0.265
Dispersion Phi=1.31

Elde edilen final model değerlendirildiğinde, depresyonlu bireylerin intihar girişim sıklığı % 19 oranında daha yüksek bulunmuştur ($P=0.009$). Ayrıca ebevenyleri ile sorun yaşayanların intihar girişim sıklığının ise % 57 kat daha fazla olduğu gözlenmiştir ($P=0.000$). Gözlenen ve final model yardımıyla tahmin edilen intihar girişim sayıları Tablo 3'te verilmiştir.

Tablo 3. Gözlenen ve beklenen intihar girişim sayılarına ait tanımlayıcı istatistikler

	Ortalama	Ortalamanın SE	Varyans	Varyansın SE	Varyans için % 95 CI	
					Alt sınır	Üst sınır
Gözlenen intihar sayıları	1.143	0.181	1.833	0.349	1.203	2.771
Beklenen intihar sayıları	1.143	0.108	0.755	0.100	0.466	1.172

* SE : Standard Error

Yukarıdaki açıklamalarda, kurulan PRM'nin uygun olduğunu yani önemli düzeyde overdispersion içermediğini vurgulamıştık. Bununla beraber Tablo 3 incelendiği zaman gözlenen intihar girişim sayısına ait varyans ortalamadan biraz büyük çıktığı için az da olsa overdispersion varlığını göstermektedir. Bu tabloda varyansa ait % 95 ihtimalli güven aralığı içine ortalama değer girmediği için overdispersion varlığı istatistik olarak anlamlı bulunmuştur. Bu durumda ya negatif binomial regresyon modeli kurulmalı veya regresyon katsayılarının standart hataları Dispersion Phi katsayısı ile düzeltilmelidir. Burada düzeltme yapılmış ve bu durumda tahmin edilen ortalamanın, varyansa ait % 95 ihtimalli güven sınırları içine düştüğü gözlenmiştir. Bu durumda overdispersion etkisi ortadan kaldırılmıştır.

Bu sonuçlara göre Poisson modeli aşağıdaki gibi oluşturulur.

(Tahmini intihar girişim sayısı / Populasyon büyüklüğü) =

$$\text{Exp}(-4.49+0.78 \times \text{Depresyon}+1.72 \times \text{Problem})$$

veya

Log (Tahmini intihar girişim sayısı) - log (Populasyon büyüklüğü) =

$$- 4.49+0.78 \times \text{Depresyon}+1.72 \times \text{Problem}$$

Tablo 2' de verilen regresyon katsayıları, depresyon durumunu ifade eden (var veya yok) her bir kategori için ortalama intihar girişim sayısını tahmin etmede kullanılır. Bu değerler aşağıda tanımlanmıştır.

- 1) Depresyonlu ve ebeveyni ile problemlili olan çocuklar (1 ve 1)
- 2) Depresyonlu ve ebeveyni ile problem yaşamayan çocuklar (1 ve 0)
- 3) Depresyonsuz ve ebeveyni ile problemlili olan çocuklar (0 ve 1)
- 4) Depresyonsuz ve ebeveyni ile problem yaşamayan çocuklar (0 ve 0)

Yukarıda tanımlanan PRM kullanılarak bu 4 kategoriye ait ortalama intihar girişim sayıları hesaplanabilir. Mesela “1” nolu kategorinin ortalaması;

(Tahmini intihar girişim sayısı / Populasyon büyüklüğü) =

$$\text{Exp} (-4.49+0.78x1+1.72x1) = 0.1366\text{'dir.}$$

Bulunan bu değer “1” nolu kategoriye ait intihar girişim insidansı olup c / N değerine eşittir. N ise bu kategoriye giren toplam çocuk sayısıdır. Bu sonuca göre “1” nolu kategorinin ortalama intihar girişim sayısı 12.28 ve tahmin edilen intihar girişim sayısı (c) ise $12.28 \times 0.1366 = 1.68 \approx 2$ dir. Diğer kategorilere ait hesaplamalarda benzer şekilde yapılır.

Tartışma

Toplumda çok sık gözlenmeyen bir olay olmasından dolayı, çocuk ve ergenlerdeki intihar girişim sıklığı ile 4 risk faktörü arasındaki ilişkinin incelenmesinde PRM' yi kullandık. Bu risk faktöründen, çocuğun depresyon ve ailesiyle problem yaşama durumunun intihar girişim sıklığını artırdığı gözlenmiştir. Bu faktörlerin etkisi başka araştırmalarda da vurgulanmıştır. Özellikle depresyon intihar girişiminin en önemli etkenlerinden biri olarak açıklanmaktadır (15, 16, 17, 18, 19, 20, 21, 22). Yapılan son epidemiyolojik çalışmalar, intihar girişiminde bulunan çocukların yaşam oranının 1% ile 15% arasında olduğunu göstermektedir (23, 24, 25). Ayrıca bireylerin karşılaştıkları problemleri çözme yeteneklerinde eksikliklerin olması da intihar girişimini artıran bir risk faktörü olarak açıklanmıştır (26).

Poisson regresyon modeli tıbbi çalışmalarda daha sık karşılaşılan bir model çeşididir. Çünkü nadir olaylar veya nadir görülen hastalıklarla ilgili risk çalışmalarında bu modelin kullanılması gerekir. Bu çalışma ile hem Poisson regresyon modelinin işleyiş adımları açıklanmış hem de uygun model seçimi üzerinde durulmuştur. Benzer materyali kullanan ve benzer amaçlar içeren birçok araştırmada Poisson modelinden farklı modeller kullanılarak risk araştırması yapılmıştır (15, 16, 17, 18, 19, 20, 21). Bu çalışma ile hem metodolojik hem de konu içeriği bakımından benzerlik gösteren çalışmalarda mevcuttur (10, 11). Bunun dışında epidemiyolojistler tarafından ölümlerin diğer bazı nedenleri ile ilgili çalışmalarda da PRM'nin kullanıldığını görmekteyiz (14, 27).

Sonuç olarak veri yapısı bu çalışmada tanımlandığı gibi ise Poisson regresyon modelinin kullanımı tercih edilmelidir.

Kaynaklar

1. Cameron A C, Trivedi PK. Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics* 1986; 1: 29-53.
2. Cameron AC, Trivedi PK. Regression analysis of count data. Cambridge, U.K. Cambridge University Press, 1998.
3. Poston DL. Modeling of the fertility of Chinese Woman. *Journal of Modern Applied Statistical Methods*, 2002; 1(2): 387-396.
4. Hendrickx J. Review of regression models for categorical dependent variables using stata by long and freese. *The stata Journal* 2002; 2(1): 103-105.
5. Long JS. Regression models for categorical and limited dependent variables. Thousand Oaks, California: Sage Publications, 1997.
6. Land KC, McCall PL, Nagin DS. A Comparison of Poisson, Negative binomial, and semiparametric Mixed Poisson regression models with empirical applications to criminal careers data. *Social Methods and Research* 1996; 24: 387-442.
7. Birckmayer J, Hemenway D. Minimum-Age drinking laws and youth suicide, 1970-1990. *Am J Public Health* 1999; 89: 1365-1368.
8. Agbayewa MO, Marion SA, Wiggins S. Socioeconomic factors associated with suicide in elderly populations in British Columbia: An 11-Year Review. *Can J Psychiatry* 1998; 43: 829-836.
9. Brent DA, Perper JA, Alman CJ. Alcohol, firearms, and suicide among youth. Temporal trends in Allegheny County, Pennsylvania, 1960 to 1983. *The journal of American Medical Association* 1987; 257(24).
10. Ray LY, Johnson N. Adolescent suicide. *Personel and Guidance Journal* 1983; 62(3): 131-135.
11. Gutstein, S. E. and Rudd, M. D. (1990). An outpatient treatment alternative for suicidal youth. *Journal of Adolescence* 13, 265-277.
12. Juon, H. S., Nam, J. J. and Ensminger, M. E. (1994). Epidemiology of suicidal behavior among Korean adolescents. *J Child Psychol Psychiatry* 35(4), 663-676.
13. Kelvin KWY, Andy HL. Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme. *Statistics in Medicine* 2001; 20: 2907-2920.
14. Frome EL. The analysis of rates using Poisson regression models. *Biometrics* 1983; 39: 665-674.

15. Kovacs. Rating scale to assess depression in school aged children. *Acta Paedopsychiat* 1983; 46: 305-315.
16. Garrison CZ, Jackson KL, Addy CL, McKeown RE, Waller JI. Suicidal behaviour young adolescents. *Am J Epidemiology* 1992; 133: 1005-1014.
17. Brinkman-Sull DC, Overholser JC, Silverman E. *Suicide life threat. Behav* 2000; 30(4): 327-340.
18. Lewinsohn PM, Rohde P, Seeley JR. Psychosocial characteristics of adolescents with a history of suicide attempts. *J Am Acad Child and Adolesc Psychiat* 1993; 32: 60-68.
19. Öy B. Çocuklar için depresyon ölçeği. Geçerlilik ve güvenilirlik çalışması. *Türk Psikiyatri Dergisi* 1991; 2: 132-136.
20. Toros F, Bilgin GN, Buğdaycı R, Şaşmaz T, Çamdeviren H. Çocuk ve ergenlerde intihar girişimi ve risk faktörleri. *Adli Bilimler Dergisi* 2003; 2(2): 11-18.
21. Murphy TR. Woful childe of parents rage: suicide of children and adolescents in early modern England, 1507-1710. *Sixt Century J* 1986; 17(3): 259-270.
22. Klerman GL. Clinical epidemiology of suicide. *Jour. Clin. Psychiatry* 1987; 48: 33-38.
23. Lewinsohn PM, Rohde P, Seeley JR. Psychosocial risk factors for future adolescent suicide attempts. *J Consult Clin Psychol* 1994; 62: 297-305.
24. Lewinsohn PM, Rohde P, Seeley JR, Baldwin CL. Gender differences in suicide attempts from adolescence to young adulthood. *J Am Acad Child Adolesc Psychiatry* 2001; 40(4): 427-434.
25. Winchstrom L. Predictors of adolescent suicide attempts: A nationally representative longitudinal study of Norwegian adolescents, *J Am Acad Child Adolesc Psychiatry* 2000; 39(5): 444-450.
26. Goldston DB, Daniel SS, Reboussin BA, Reboussin DM, Frazier PH, Haris AE. Cognitive risk factors and suicide attempts among formerly hospitalised adolescents: A prospective naturalistics study, *J Am Acad Child Adolesc Psychiatry* 2001; 40(1): 91-99.
27. Gerry H. Horse kicks, anthrax and the Poisson model for deaths. *Chronic Diseases* 2002; 23(2):

APPENDIX
SAS ALGORITMS

The following SAS program runs Poisson regression.

```
*****  
* Poisson regression *  
*****  
data suicide;  
  input age problem depression order suicide population;  
  l_population = log(population);  
  cards;  
  12  0 0 0 0 145  
  13  0 0 0 3 210  
  14  0 0 0 3 195  
  .  
  .  
  .  
  ;  
run;  
proc genmod data=suicide;  
  class problem depression order;  
  model suicide = age problem depression order / dist = poisson  
    link = log  
    offset = l_population type3 wald;  
  estimate 'problem 1 depression 1' int 1 problem 1 0 depression 1 0 /exp;  
  estimate 'problem 1 depression 0' int 1 problem 1 0 depression 0 1/exp;  
run;  
  
*****  
* negative binomial regression *  
*****
```

```
proc genmod data=suicide;
  class problem depression order;
  model suicide = age problem depression order / dist = nb
    link = log
    offset = l_population type3 wald;
  estimate 'problem 1 depression 1' int 1 problem 1 0 depression 1 0 /exp;
  estimate 'problem 1 depression 0' int 1 problem 1 0 depression 0 1 /exp;
run;
```

The SAS syntax for running negative binomial regression is almost the same as for Poisson regression. The only change is the **dist** option in the **MODEL** statement. Instead of **dist = poisson**, **dist = nb** should be used.

GÜÇ DOĞUM NEDENİ İLE ÖLEN BUZAĞILARA MEVSİM VE YIL FAKTÖRLERİNİN ETKİSİNİN POISSON REGRESYON ANALİZİ KULLANARAK İNCELENMESİ

I.Safa Gürcan¹, İlknur Özmen², Aytaç Akçay¹, Mehmet N. Orman¹

*Ankara Üniversitesi Veteriner Fakültesi Biyoistatistik Anabilim Dalı¹
Başkent Üniversitesi Fen-Edebiyat Fakültesi İstatistik ve Bilgisayar Bilimleri Bölümü²*

Özet

Genel anlamda regresyon analizi, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi matematiksel olarak ortaya koyan bir modeller topluluğudur. Klasik regresyon analizinde bağımlı değişkenin sürekli olması ve hata terimleri üzerinde bir takım normallik varsayımlarının sağlanması zorunluluğu vardır. Bağımlı değişkenin süreklilik göstermediği durumlar için geliştirilmiş modeller bulunmaktadır. Bağımlı değişkenin kesikli olduğu Poisson Regresyon modelinde normallik varsayımlarına ihtiyaç duyulmaması nedeni ile klasik doğrusal regresyon modellerine bir alternatif olarak görülmektedir. Poisson regresyon modelinde amaç, bağımlı değişkenin belirli bir zaman aralığında tekrar etme sayılarının modellenmesidir. Bu modelde, açıklayıcı değişkenleri yanıt değişkenin beklenen değerine bağlayan fonksiyon logaritmik olup hatalar Poisson dağılımlıdır.

Güç doğum (Dystocia) hayvancılıkta daha çok ilk defa buzağılama yapan sığırlarda görülmektedir. Güç doğum, doğum sırasında sığırın hem yavrusunu hem de kendisinin ölmesi ile sonuçlanabilen bir vakadır. Yapılan araştırmalar, annenin yaşının, doğum mevsiminin, yavrunun doğum ağırlığı ve cinsiyeti ile annenin pelvik alanının güç doğuma etkisi olduğunu göstermiştir.

Güç doğum sayısında görülen artış, işletmelerde hayvan ölümlerinde olduğu gibi veterinerlik maliyetlerinde de bir artışa neden olur.

Bu çalışma ile, 2000 – 2002 yılları arasında özel bir işletmede güç doğum vakası nedeni ile ölen buzağı sayıları, Poisson regresyon analizi ile yıl ve mevsim etkisi gözönünde bulundurularak modellenmiştir.

Pearson Ki-Kare ve sapma (Deviance) istatistiklerinin serbestlik derecelerine bölünerek elde edilen değer, modelin aşırı yayılıma veya az yayılıma sahip olmasının belirlenmesinde kullanılır. Poisson dağılımında örneklem ortalaması ve örneklem varyansı birbirine eşit olmalıdır. Bu

durumda Sapma ve Pearson istatistikleri 1'e yakın değer almalıdır. Değerlerin >1 olması aşırı yayılıma (overdispersion), dolayısı ile varyansın ortalamadan büyük değer aldığına işaret eder. Değerlerin 1'den az olması ise az yayılıma (underdispersion), varyansın ortalamadan az değer aldığını gösterir. Kullanılan modeldeki Sapma ve Pearson Ki- Kare değerleri 0'a yakın değerler aldığından dolayı model az yayılım (underdispersion) göstermektedir.

İKİ POISSON ORTALAMASININ KARŞILAŞTIRILMASI

S. Karahan, J. Karakaya, O. Saraçbaşı

Hacettepe Üniversitesi Tıp Fakültesi, Biyoistatistik AD, Ankara

e-mail: sevilaykarahan@gmail.com

Amaç

Poisson dağılımı; nadir görülen olaylarda olasılık hesaplamak için kullanılan bir dağılımdır. Biyolojik, epidemiyolojik ve sağlık alanında yapılan pek çok çalışmada; iki bağımsız kitleden elde edilmiş poisson ortalamasının karşılaştırılması ilgi konusu olmuştur.

Bu çalışmada iki bağımsız kitleden elde edilen poisson ortalamalarının farklı test istatistikleri ile karşılaştırılması amaçlanmaktadır.

Anahtar Kelimeler: Poisson Dağılımı, Örneklem tahmini, Kısıtlı en çok olabilirlik tahmini, Logaritmik dönüşüm.

Gereç ve Yöntem

Poisson dağılımında olayın görülüş olasılığı (p) küçük, kitle ise (n) çok büyüktür. Poisson dağılımı;

$$P(X, \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x=0,1,2,\dots$$
$$\lambda > 0$$

şeklinde gösterilir.

t_0 ve t_1 gibi iken belirli zaman aralığında, λ_0 ve λ_1 parametreleri ile Poisson dağılımı gösteren iki kitle olsun. X_0 ve X_1 ise bu kitlelerde gözlenen olay sayısını göstere.

$$X_i \sim \text{Poisson}(\lambda_i)$$

$$\lambda_i = t_i \gamma_i, i = 0,1$$

İki poisson ortalaması bu çalışmada;

1- Logaritmik dönüşüm olmaksızın

a- Örneklem tahmini yöntemi ile (W_1)

b- Kısıtlı en çok olabilirlik tahmin (CMLE) yöntemi ile (W_2)

2- Logaritmik dönüşüm ile

a- Örneklem tahmini yöntemi ile (W_3)

b- Kısıtlı en çok olabilirlik tahmin (CMLE) yöntemi ile (W_4)

hesaplanacaktır.

$$H_0 : \gamma_1 / \gamma_0 = 1$$

$$H_A : \gamma_1 / \gamma_0 > 1$$

$$\gamma_i \text{ 'nin örneklem tahmini} \Rightarrow \hat{\gamma}_i = X_i / t_i, i = 0, 1$$

$$\gamma_i \text{ 'nin CMLE'si (} H_0 \text{ altında)} \Rightarrow \hat{\gamma} = (X_0 + X_1) / (t_0 + t_1)$$

Logaritmik Dönüşüm Olmaksızın Hesaplanan İstatistikler:

H_0 'ın testinde T/s_T wald istatistiği kullanılabilir.

$$H_0 \Rightarrow \gamma_1 - \gamma_0 = 0$$

$$T = \hat{\gamma}_1 - \hat{\gamma}_0 = X_1 / t_1 - X_0 / t_0$$

$$V(T) = \sigma_T^2 = \gamma_1 / t_1 + \gamma_0 / t_0$$

$$s_T^2 = \hat{\gamma}_1 / t_1 + \hat{\gamma}_0 / t_0$$

Burada $\hat{\gamma}_i$, γ_i 'nin tahmini ($i=0,1$)

Örneklem tahmini:

$$\hat{\gamma}_i = \hat{\gamma}_i \quad (i=0,1)$$

$$W_1 = \frac{X_1 - dX_0}{(X_1 + d^2X_0)^{1/2}}$$

Burada $d=t_1/t_0$

CMLE tahmini:

$$\hat{\gamma}_i = \hat{\gamma} \quad (i=0,1)$$

$$W_2 = \frac{X_1 - dX_0}{[d(X_0 + X_1)]^{1/2}}$$

Logaritmik dönüşüm ile hesaplanan istatistikler:

$$U = \ln(\hat{\gamma}_1 / \hat{\gamma}_0) = \ln(X_1 / t_1) - \ln(X_0 / t_0)$$

$$\text{Delta metodu ile: } \text{Var}(U) = \sigma_U^2 = \frac{1}{(t_0 \gamma_0)} + \frac{1}{(t_1 \gamma_1)}$$

$$s_U^2 = \frac{1}{(t_0 \hat{\gamma}_0)} + \frac{1}{(t_1 \hat{\gamma}_1)}$$

U/s_U H_0 'ın testinde kullanılabilir.

Örnekleme tahmini:

$$\hat{\gamma}_i = \hat{\gamma}_i \quad (i=0,1)$$

$$W_3 = \frac{[\ln(X_1 / X_0) - \ln(d)]}{\sqrt{(1/X_0) + (1/X_1)}}$$

CMLE tahmini:

$$\hat{\gamma}_i = \hat{\gamma} \quad (i=0,1)$$

$$W_4 = \frac{[\ln(X_1 / X_0) - \ln(d)]}{\sqrt{(2 + (1/d) + d)/(X_0 + X_1)}}$$

Yokluk hipotezi altında; W_j ($j=1, \dots, 4$) asimptotik olarak standart normal dağılım gösterir. Yani; $W_j > z_{1-\alpha}$ olduğunda H_0 hipotezi reddedilebilir.

Sonuç

Bu çalışmada Poisson dağılımı gösteren örnek bir veri kümesi üzerinde bu dört istatistik hesaplanacak ve bunların performansları karşılaştırılacaktır.

KONTENJANS TABLOLARINDA ASİMPOTİK P, EXACT P VE MONTE CARLO P DEĞERLERİNİN KARŞILAŞTIRILMASI

Ferhan Elmalı, Canan Demirüstü, Ertuğrul Çolak, Kazım Özdamar

Osmangazi Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı

Özet

Bu çalışmanın amacı 3x3, 4x4, 5x5 ve 10x10 düzeni kontenjans tablolarında sıklıkla kullanılan Pearson Ki-kare test istatistiğinin Asimptotik p değeri ile bu tablolardan elde edilen Exact p ve Monte Carlo p değerlerinin karşılaştırılmasıdır.

3x3 düzeni kontenjans tablolarında Asimptotik p, Exact p ve Monte Carlo p değerleri karşılaştırılırken diğer kontenjans tablolarında ise sadece Asimptotik p ve Monte Carlo p değerleri karşılaştırıldı. Çalışmada veriler integer dağılımdan rasgele olarak türetildi ve tüm düzenler için 100 tablo üzerinde çalışıldı. Tablolardaki 5'ten küçük beklenen değerlerin gözelerdeki dağılımı dikkate alınarak tablolar gruplandırıldı. Her bir tablo için p değerleri hesaplandı. Monte Carlo p değerleri her bir tablo için 10.000, 100.000 ve 250.000 örnek alınarak elde edildi. 5'ten küçük beklenen değerlerin bulunmadığı 10x10 düzeni kontenjans tabloları dışındaki tüm tablolarda Asimptotik p değerleri ile Monte Carlo p tahmin değerleri arasında önemli düzeyde farklılık bulundu. Ayrıca 3x3 düzeni kontenjans tablolarında Asimptotik p değerleri ile hem Exact p değerleri hem de Monte Carlo p değerleri arasında farklılık bulunurken, Exact p değerleri ile Monte Carlo p değerleri arasında farklılık saptanmadı. Yapılan karşılaştırmalar sonucunda bulunan farklılıklar, Exact p ve Monte Carlo p değerlerinin Asimptotik p değerlerinden yüksek olduğunu gösterdi.

Sonuç olarak 3x3, 4x4, 5x5 ve 10x10 düzeni kontenjans tablolarında daha doğru karar verebilmek için Pearson Ki-kare test istatistiğinin Asimptotik p değerinin kullanılması yerine Exact p değeri ile Monte Carlo p tahmin değerinin kullanılması önerilmektedir.

Anahtar Kelimeler: Pearson Ki-Kare test istatistiği, Asimptotik p değeri, Exact p değeri, Monte Carlo Simülasyonu.

BURUN YAPISININ İSTATİSTİKSEL ŞEKİL ANALİZİ İLE İNCELENMESİ

İlker Ercan¹, İbrahim Güney², Abdullah Etöz³, Gökhan Ocakoğlu¹,
İsmet Kan¹, Mesut Özcan³

Uludağ Üni. Tıp Fak. Biyoistatistik AD.¹

Uludağ Üni. Fen Edebiyat Fakültesi Matematik Bölümü²

Uludağ Üni. Tıp Fak. Plastik ve Rekonstrüktif Cerrahi AD.³

Özet

Çalışmamızda 18-25 yaş arasındaki kadın ve erkeklerin burun yapısının önden, yandan ve alttan görünüşü istatistiksel şekil analizi ile incelendi. Araştırmada 32 kadın ve 28 erkek denek yer almıştır.

Deneklerin üç konumdan çekilen dijital fotoğrafları anatomik landmarklar kullanılarak işaretlenmiştir. İki boyutlu olarak yapılan çalışmada, önden görünüşte 11, yandan görünüşte 12 ve alttan görünüşte 14 anatomik landmarkla işaretleme yapılmıştır. İstatistiksel şekil analizinde Genelleştirilmiş Procrustes Analizi (GPA) kullanıldı. Kız ve erkeklerin burun şekillerinin Procrustes şekil ortalamaları, şekiller arası uzaklık için Riemannian şekil uzaklığı (ρ), genel bir şekil değişkenliğinin ölçüsü hakkında bilgi sahibi olabilmek için her bir konfigürasyondan tam procrustes ortalamasına kadar ölçülen tam procrustes uzaklığının kareli ortalamasının karekökü olan istatistik ($RMS(d_F)$) hesaplandı. Şekil ortalamaları arasındaki farklılıklar Goodall'ın F testi ile karşılaştırıldı. Kadın ve erkeklerde burun şeklinin önden görünüşü bakımından istatistiksel olarak anlamlı farklılık bulunurken, yandan ve alttan görünüş bakımından anlamlı bir farklılık bulunamadı ($p = 0.003$, $p = 0.092$, $p = 0.361$). Kadın ve erkekler arasındaki şekil uzaklıkları (ρ) incelendiğinde önden 0.043, yandan 0.039, alttan 0.048 olarak bulundu. $RMS(d_F)$ değerleri ise önden görünüşte erkeklerde 0.109, kadınlarda 0.112, yandan görünüşte erkeklerde 0.122, kadınlarda 0.126, alttan görünüşte erkeklerde 0.220, kadınlarda 0.100 olarak bulundu.

İstatistiksel şekil analizi sonucunda kadın ve erkeklerde burunun önden görünüş şekli bakımından bir farklılık olduğuna karar verildi.

KATEGORİK VERİ ÇÖZÜMLEMESİNDE LOGARİTMİK DOĞRUSAL MODELLER VE BİR UYGULAMA

Beyza Doğanay, Kenan Köse, Umut Özbek, Yıldır Atakurt

Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı

e-mail: bdoğanay@excite.com

Özet

Kategorik veri çözümlemesinde sıkça kullanılan ki-kare testi, tablo boyutu arttıkça elde edilen sonuçların yorumlanmasında yetersiz kalmaktadır. Bu durumda çok boyutlu çapraz tablonun iki boyutlu alt tablolara ayrılarak; her bir alt tablo için ki-kare istatistiğinin yorumlanması yerine, logaritmik doğrusal modeller yardımıyla değişkenler arasındaki ilişkinin yapısının güvenilir ve düzgün modeller kurularak tanımlanması mümkündür. Logaritmik doğrusal modellerin amacı ilişki ve etkileşim yapılarının analiz edilmesidir [1].

Logaritmik doğrusal modeller aslında, genelleştirilmiş doğrusal modellerin özel bir durumu olup, poisson ya da çokterimli dağılımdan gelen ve $R \times C$ biçiminde düzenlenmiş kategorik değişkenler arasındaki birlikteliğin ve etkileşimin yapısal özelliğini ortaya koymaya çalışırlar. Gerçekte bu analiz, iki boyutlu çapraz tablolardan yola çıkılarak, iki ya da daha fazla kategoriye sahip değişkenler arasındaki ilişkiyi, gözelerdeki beklenen sıklıkların logaritmasını kullanarak ortaya koyar. Bununla birlikte logaritmik doğrusal model analizinin üç ya da daha çok boyutlu tablolarda kullanımı daha yaygındır. Bu modelde kullanılan değişkenler arasında bağımlı ya da bağımsız değişken ayrımı yapılmaz. Çapraz tablodaki değişkenlerin her biri açıklayıcı değişken, tablo gözelerindeki sıklıklar ise cevap değişkeni olarak alınır.

Logaritmik doğrusal modellerde, bütün olası etkileşimlerin ve ana etkilerin olduğu model "doğgun model" olarak tanımlanır ve doğgun model bir hiyerarşik modeldir. Çapraz tablodaki sıklıkların dağılımına en uygun model, hiyerarşik modeller kullanılarak belirlenir. Hiyerarşik modellerde en karmaşık etkileşim yapısından en az karmaşık olan etkilere kadar olan parametreler vardır. Örneğin; $2 \times 2 \times 2$ 'lik bir çapraz tablo için üçlü etkileşimin olduğu bir hiyerarşik modelde tüm ikili etkileşimler ve ana etkiler de mevcuttur. Model seçim işlemi regresyonda olduğu gibi ileriye ya da geriye yönelik yapılabilir.

Bu çalışmada, 8-18 yaş arası 2776 hipertansiyon hastası üzerinde yapılan bir araştırmanın verileri kullanılarak, hipertansiyona sahip olmaları üzerinde, bu bireylerin anne ve babalarında “hipertansiyon”, “obezite” ve “lipit kolesterol” değişkenlerinin tek başına ve birlikte etkili olup olmadığı incelenmiştir.

Anahtar Kelimeler: Loglinear Modeller, Kategorik veri analizi, Çapraz tablo çözümlemesi.

Kaynaklar

1. Agresti, A. (1996). *An Introduction to Categorical Data Analysis*, John Wiley and Sons Inc.
2. <http://www2.chass.ncsu.edu/garson/pa765/logit.htm>
3. <http://www.uky.edu/ComputingCenter/SSTARS/LoglinearModelAnalysisinSASandSPSS.htm>

ALT GRUP ANALİZİ

Beyza Akdağ, Vildan Sümbüloğlu, Kadir Sümbüloğlu

Pamukkale Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, DENİZLİ¹

Karaelmas Üniversitesi, Zonguldak²

Biyoistatistik Derneği Başkanı, Ankara³

[e-mail:bakdag@pamukkale.edu.tr](mailto:bakdag@pamukkale.edu.tr)

Özet

Klinik çalışmalarda verilen yanıtlar demografik özellikler, hastalık evresi, tedavi ve diğer birçok faktöre göre değişiklik gösterir. Bir popülasyonun herhangi bir alt grubunda belirli bir tedavi grubu daha etkiliyse bu durumu bilmek ve neden olduğunu anlamak önemlidir. Bu uygulama, klinik çalışmalarda alt grup analizleri olarak isimlendirilir. Alt grupların incelenmesi çok önemlidir. Bu nedenle, elde edilen verinin ayrıntılı değerlendirilmesi yerinde bir düşüncedir. Klinisyenler, klinik çalışmalarda farklı özellikteki hastalara tedaviyi uygular ve tedavinin hepsinde işleyip işlemeyeceğini araştırır. Elbetteki, amaç tedavinin en iyi sonuç verdiği alt grupları bulmaktır. Diğer taraftan ise, çok sayıda alt grup analizi yapılması sakıncalı bir durumdur. Çünkü yapılan çok fazla sayıda alt karşılaştırma sonucunda herhangi bir alt grupta tedavinin farklı etkinlikte olduğu sonucu ortaya çıkabilir. Alt grup analizinden beklenti hangi grubun entervasyondan (uygulanan tedavi) daha iyi yararlanacağını ya da hangi gruba entervasyon önerileceğini açıklığa kavuşturmadır. Ancak, alt grup analizi yapılacaksa çalışma çok dikkatli biçimde planlanmalı ve yorumlanmalıdır. Aksi takdirde deneklerin herhangi bir çıktı (sonuç) değişkenine göre kategorilere ayrılması, örneğin; uyum durumu ya da yanıt verenler-yanıt vermeyenler gibi yanlı karara sürüklenmesine neden olabilir. Alt grup analizi yapılacaksa, çalışmanın başında hangi grup ya da grupların inceleneceği, örneklem büyüklüğünün ne olacağı belirlenmelidir ve çalışma protokolünde açık biçimde belirtilmelidir. Çalışma sonuçları bu durumda güvenilir olacaktır.

Alt grup analizi tıp literatüründe karışıklık ortaya çıkarmaktadır. Çoğu klinisyenler ve bilim adamları, alt grup analizlerinin uygulanmaması gerektiğini, bazıları da alt grup analizinin uygulanması gerektiğini düşünmektedir. Kimileri de alt grup analizlerinin sadece deneyimli araştırmacılar tarafından uygulanması gerektiğini savunmakta ve başkalarının alt grup analizlerine şüphe ile bakmaktadırlar. Bu karmaşaya birkaç örnek verelim.

- Kimi bilim adamlarına göre yapılsa bir türlü yapılmasa bir türdür
- Alt grup analizi, hem bilgilendirici hem de kuvvetli biçimde yanıtıcıdır (Oxman ve Guyatt, 1992)
- Alt grup analizi önemlidir. Ancak dikkatli yapılmalı ve dikkatli biçimde yorumlanmalıdır (Friedman, Furberg ve De Mets, 1998)
- Araştırmacılar alt grup analizini kullanabilirler ancak veri eşelemeden (data dredging) kaçınmalıdırlar (Meinert, 1998)
- Alt grup analizlerine klinik denemelerde eğlendirici olarak bakılmalıdır ancak güvenilmemelidir (Sleight, 2000)

Alt gruplar arası karşılaştırmalarda varılan kararın düşük güce sahip olma olasılığı her zaman vardır. Bu nedenle, alt grup karşılaştırmalarında araştırmacı önemlilik testlerine birincil önem (genel değerlendirme kadar) vermemelidir. İnteraksiyonların incelenmesini ön plana çıkarmalıdır.

Anahtar Kelimeler: Alt grup analizi, veri eşeleme, keşifsel veri analizi.

POPULASYON GENETİĞİ PRENSİPLERİ VE HARDY-WEINBERG KURALI

Mesut Akyol¹, S.Yavuz Sanisoğlu², İlker Etikan³, Zeki Akkuş⁴,
Reha Alpar⁵

GATA¹

GATA Biyoistatistik BD²

GOP Üniversitesi Biyoistatistik AD³

Dicle Üniversitesi Biyoistatistik AD⁴

Hacettepe Üniversitesi Biyoistatistik AD⁵

e-mail:makyol@gata.edu.tr

Özet

Hardy-Weinberg Kuralı, belli koşullar altında rasgele eşleştirmeden bir kuşak sonra; tek bir gen bölgesinde, gen tipi frekanslarının belli bir denge değerinde sabitlendiğini öne sürer. Aynı zamanda bu denge frekanslarının aynı gen bölgesinde alel frekanslarının basit bir fonksiyonu olarak gösterilebileceğini belirtir.

Hardy-Weinberg Kuralı "Genetik Denge" deki populasyonun gösteriminin bir tanımı ve Populasyon Genetiğinin temel prensibidir. En basit hali ile frekansları sırasıyla p ve q olan B ve b gibi iki alelli tekil bir bölgede BB homozigot, Bb heterozigot ve bb homozigotlarının p^2 , $2pq$ ve q^2 frekanslarını kestirir.

Bu çalışmada Populasyon Genetiği istatistiğinin temel yaklaşımı olarak kabul edilen Hardy-Weinberg Kuralı'nı tanıtmaya çalıştık.

Anahtar Kelimeler: Hardy-Weinberg Kuralı, HWE, Populasyon Genetiği.

Giriş

Bir özellik için heterozigot iki birey eşleştirildiğinde, bunlardan elde edilen çıktıların (dölün) % 25'inin baskın alel (BB), % 25'inin ebeveynlerine benzemeyen şekilde çekinik alel (bb) ve % 50'sinin de ebeveynleri gibi heterozigot (Bb) oldukları görülür. Bu Mendel'in monohibritleri çaprazlaması ile elde edilmiş bir sonuçtur. Bu olay her bir heterozigot ebeveynin iki alelinin

ayrıştırılmasıyla, gametlerin % 50'sinin bir alel, diğer % 50'sinin ise diğer aleli taşımasından kaynaklanır. Bu durum matematiksel olarak

$$(p + q)^2 = p^2 + 2pq + q^2 \quad (1)$$

şeklinde binomiyal açılım ile gösterilir.

Populasyondaki genlerin toplam sayısı gen havuzu olarak adlandırılır. Burada p ile havuzdaki bir genin frekansı, q ile bunun tek alelinin frekansı gösterildiğinde $p+q = 1$ olup, p^2 , p için homozigot populasyonun oranı, q^2 , q için homozigot oranı ve $2pq$ de heterozigot oranıdır.¹

Gereç ve Yöntem

Hardy-Weinberg kuralı (HWE)

- Sonsuz bir popülasyonda,
- Rasgele eşleştirme,
- Çakışmayan kuşaklar,
- Özel seçim ve göç olmaması

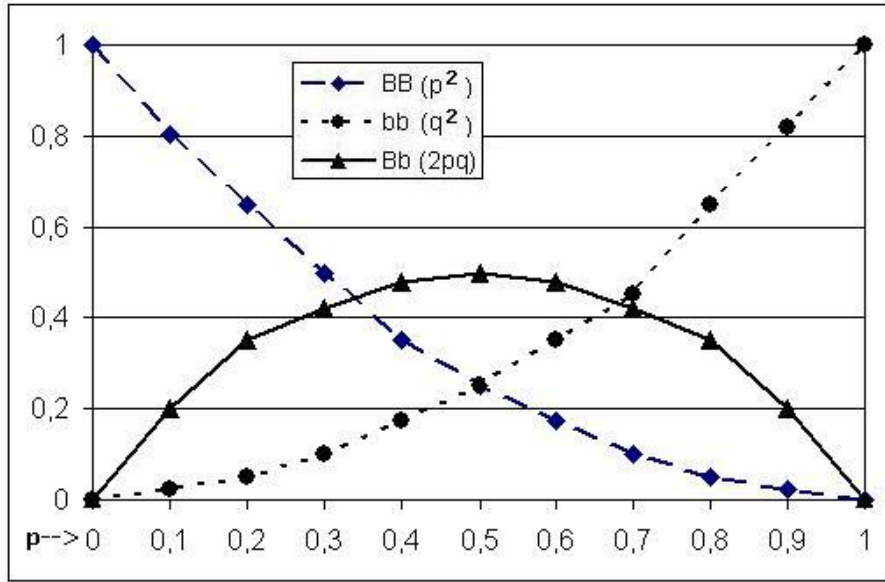
durumunda, gen frekansları olan p ve q değerleri yardımı ile sonsuz popülasyon gen tipi frekanslarının (1) bağıntısı ile elde edilebileceğini öne sürer.² Yukarıdaki varsayımlardan herhangi birisi bozulmadığı sürece tekil bir popülasyonda HWE her zaman sağlanır. HWE'nin sağlanıp sağlanmadığı gözlenen ve beklenen frekans değerleri karşılaştırılarak test edilir.³

İki alel için HWE dağılım grafiği Şekil 1'de gösterilmiştir. Şekilde yatay eksen alellerin p ve q frekanslarını, düşey eksen ise gen tipi frekanslarını göstermektedir. Şekildeki üç ayrı çizgi üç ayrı gen tipini göstermektedir.⁴

HWE'ye göre; belli koşullar altında, rasgele eşleştirmeden bir kuşak sonra tek bir gen bölgesinde (Locus), gen tipi frekansları belli bir denge değerinde sabitlenir. HWE'ye göre aynı zamanda denge frekansları bu bölgede alel frekanslarının basit bir fonksiyonu olarak gösterilebilir.

HWE varsayımlarının bozulması, beklenen değerlerden sapmalara neden olur. Sapma miktarı, hangi varsayımın bozulduğuna bağlıdır. Bu sapmaların bir kısmı istatistiksel olarak saptanabilirken bir kısmının saptanması olanaksızdır. Sapmalar, Wahlund etkisi, birbiriyle yakın bireylerin eşleştirilmesi (inbreeding), birbirine bazı özellikleri açısından benzeyen bireylerin eşleştirilmesi (assortative mating), seçim yöntemi ya da genetik kayma gibi nedenler yüzünden oluşabilir. Bunlardan bazı özellikleri açısından benzeyen bireylerin eşleştirilmesi; sadece arzu edilen genlerin,

gen tipi frekanslarını değiştirirken, genetik kayma özellikle küçük populasyonlar için geçerlidir. Seçim şeklinden kaynaklanan sapmalar, seçim için Hardy-Weinberg oranlarını test etmek için yetersiz olduğu düşünülen sapmaları belirleyebilmek amacıyla anlamlı bir seçim katsayısına gerek duyar.



Şekil 1. İki alel için HWE dağılımı

HWE ikiden çok alel için de genelleştirilebilir. Bunun için p , q frekansları yanında frekansı r olan bir başka alel olduğu varsayalım. Bu durumda (1) bağıntısı

$$(p + q + r)^2 = p^2 + r^2 + q^2 + 2pq + 2pr + 2qr$$

şeklini alır. Genelleştirilecek olursa homozigot ve heterozigot için sırasıyla

$$f(B_i B_j) = p_i^2 \text{ ve } f(B_i B_j) = 2p_i p_j \text{ dir.}^5$$

HWE kuramına göre değişik eşleştirme tipleri sonucunda çıktı (döl) ve çıktı frekansları Tablo 1'de gösterilmiştir.

Tablo 1'deki BB ile Bb gen tiplerinin çaprazlama sonuçları incelendiğinde ilk gen tipi sadece B gametini, ikinci gen tipi ise B ve b gametlerini eşit oranda oluşturur. Bu çaprazlamada BB gen tipi tarafından üretilen gamet Bb gen tipinden gelen gamet ile eşit olarak bir araya gelir. Bu yüzden $BB \times Bb$ çaprazlamasında çıktı frekansı $\frac{1}{2} BB$ ve $\frac{1}{2} Bb$ dir. Benzer şekilde $BB \times bb$ çaprazlaması sadece Bb çıktısı verir. Bu şekilde oluşturulan değişik olası çaprazlamaların çıktı oranları, ayırım oranları (segregation ratios) olarak adlandırılır.

Tablo 1. HEW'ye göre eşleştirme çıktı ve frekansları

EŞLEŞTİRME	ÇIKTI	FREKANS
$BB \times BB$	BB	u^2
$BB \times Bb$	$\frac{1}{2} BB + \frac{1}{2} Bb$	$2uv$
$BB \times bb$	Bb	$2uw$
$Bb \times Bb$	$\frac{1}{4} BB + \frac{1}{2} Bb + \frac{1}{4} bb$	v^2
$Bb \times bb$	$\frac{1}{2} Bb + \frac{1}{2} bb$	$2vw$
$bb \times bb$	bb	w^2

BB için u , Bb için v , bb için w ile gen tiplerinin başlangıç oranları gösterilmiş olsun. Bu durumda yukarıda sıralanan varsayımlar sağlanıyorsa bir sonraki kuşak Tablo 1'deki gibi oluşacaktır. Tablo 1'deki bilgilere göre BB , Bb ve bb gen tipleri için yeni frekanslar sırasıyla;

$$u^2 + uv + \frac{1}{4}v^2 = \left(u + \frac{1}{2}v\right)^2$$

$$uv + 2uw + \frac{1}{2}v^2 = 2(u + \frac{1}{2}v)(\frac{1}{2}v + w)$$

$$\frac{1}{4}v^2 + vw + w^2 = (\frac{1}{2}v + w)^2$$

şeklinde oluşur.

B ve b alellerinin frekansları $p_1 = u + \frac{v}{2}$ ve $p_2 = \frac{v}{2} + w$ şeklinde tanımlandığında BB 'nin frekansı p_1^2 , Bb 'nin frekansı $2p_1p_2$ ve bb 'nin frekansı da p_2^2 olur. İkinci tur rasgele eşleştirmeden sonra BB , Bb ve bb gen tiplerinin frekansları sırasıyla;

$$\begin{aligned} (p_1^2 + \frac{1}{2}2p_1p_2)^2 &= [p_1(p_1 + p_2)]^2 \\ &= p_1^2 \end{aligned}$$

$$\begin{aligned} 2(p_1^2 + \frac{1}{2}p_1p_2)(\frac{1}{2}p_1p_2 + p_2^2) &= 2p_1(p_1 + p_2)p_2(p_1 + p_2) \\ &= 2p_1p_2 \end{aligned}$$

$$\begin{aligned} (\frac{1}{2}p_1p_2 + p_2^2)^2 &= [p_2(p_1 + p_2)]^2 \\ &= p_2^2 \end{aligned}$$

şeklinde olur.

Görüldüğü gibi tek bir rasgele eşleştirmeden sonra gen tiplerinin frekansları Hardy-Weinberg oranlarında durağan hale gelir.

Aynı sonuçlar gamet popülasyonu için de elde edilebilir. B gametlerinin frekansı p_1 ile b gametlerinin frekansı ise p_2 ile gösterilsin. Gametlerin rasgele birlikteliği, rasgele eşleştirmeye eşdeğer olduğundan bir sonraki kuşakta BB gametleri p_1^2 , Bb gametleri $2p_1p_2$ ve bb gametleri de p_2^2 frekansları ile görünür. Yeni kuşaktaki gamet havuzunda B gameti

$p_1^2 + p_1 p_2 = p_1(p_1 + p_2) = p_1$ ve b gameti de p_2 frekansı ile görülür. Bir başka deyişle durağanlığa tek bir kuşakta ulaşılır. Bu durum ikiden çok alel durumuna kolaylıkla genelleştirilebilir.

HWE X -bağlantılı yer için biraz daha kesindir. X kromozomlarında bir bölge ve bu bölgede bir alel olmak üzere, n inci kuşakta kadın aleli q_n , erkek aleli ise r_n ile gösterilsin. HWE varsayımları altında q_n ve r_n hızlı bir biçimde

$$p = \frac{2}{3} q_o + \frac{1}{3} r_o \text{ değerine dönüşür. Burada başlangıç kadın frekansının erkek}$$

frekansının iki katı olması; kadınların iki, erkeklerin ise sadece bir tane X kromozomuna sahip olmalarından kaynaklanmaktadır.

Erkeklerin X kromozomlarını annelerinden almalarından ve annelerinin bir kuşak önde olmalarından dolayı

$$r_n = q_{n-1} \quad (2)$$

benzer şekilde kadınlardaki frekans bir kuşaktan iki cinsiyetin ortalama frekansı olduğundan

$$q_n = \frac{1}{2} q_{n-1} + \frac{1}{2} r_{n-1} \quad (3)$$

dir. (2) ve (3) eşitlikleri bir araya getirildiğinde

$$\begin{aligned} \frac{2}{3} q_n + \frac{1}{3} r_n &= \frac{2}{3} \left(\frac{1}{2} q_{n-1} + \frac{1}{2} r_{n-1} \right) + \frac{1}{3} q_{n-1} \\ &= \frac{2}{3} q_{n-1} + \frac{1}{3} r_{n-1} \end{aligned}$$

olup, ağırlıklandırılmış ortalama tüm n 'ler için $p = \frac{2}{3} q_n + \frac{1}{3} r_n$ dir. (2) ve (3) eşitliklerinden

$$\begin{aligned} q_n - p &= q_n - \frac{3}{2} p + \frac{1}{2} p \\ &= \frac{1}{2} q_{n-1} + \frac{1}{2} r_{n-1} - \frac{1}{3} \left(\frac{2}{3} q_{n-1} + \frac{1}{3} r_{n-1} \right) + \frac{1}{2} p \end{aligned}$$

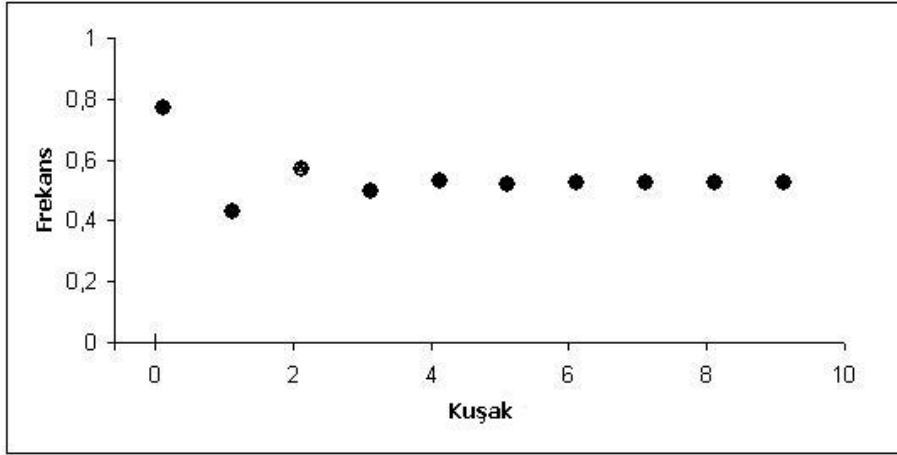
$$\begin{aligned} &= -\frac{1}{2}q_{n-1} + \frac{1}{2}p \\ &= -\frac{1}{2}(q_{n-1} - p) \end{aligned}$$

olarak bulunur. Bu şekilde devam ederek

$$q_n - p = \left(-\frac{1}{2}\right)^n (q_0 - p) \quad (4)$$

eşitliği elde edilir.

Bağıntı (4)'de de görüldüğü gibi q_n ve p arasındaki fark her bir kuşakta yarısı kadar azalmakta ve q_n zikzak bir yapı izleyerek p ye yaklaşmaktadır. r_n erkek frekansı benzer bir yapı izler ancak q_n 'i bir kuşak geriden takip eder. Yukarıdakilerden farklı olarak, cinsiyetsiz kromozom (autosomal) durumunda dengeye ulaşma bir kuşak sonraya sarkar. Bununla birlikte dengeye bağlı olarak daha hızlı biçimde ulaşılır. En aşırı durumda, yani $q_0 = 0.75$ ve $r_0 = 0.12$ durumunda birkaç kuşak sonrası için q_n grafiği Şekil.2'de gösterilmiştir.



Şekil 2. n'in fonksiyonu olarak q_n 'in dengeye yaklaşması.

Değişik gen tiplerinin frekansları denge durumunda şu şekilde hesaplanır:

Örnek olarak B ve b genleri ve bunların denge frekansları p_1, p_2 olan iki alel alınmış olsun. BB, Bb ve bb kadın genlerinin frekansları sırasıyla $p_1^2, 2p_1p_2$ ve p_2^2 şeklinde olur. Erkeklerde benzerleri bastırılmış B ve b gen tipleri p_1 ve p_2 değerlerine sahip olur.

Örnek:

Kırmızı hücre antijeni $Xg(a)$ frekansı yaklaşık olarak $p = 0.65$ olan, Kafkaslara özgü X -bağlantılı baskın bir antijendir. Bu yüzden tüm Kafkas kökenli erkeklerin % 65'i ve tüm Kafkas kökenli kadınların $p^2 + 2p(1-p) = 0.8775$ (yaklaşık % 88)'i bu antijeni taşır.

Tartışma ve Sonuç

Yirminci yüzyılın en önemli gelişmelerinden birisi, canlıların özelliklerinin nasıl meydana geldiği ve bu özelliklerin devam eden kuşaktaki bireylere nasıl aktarıldıklarının aydınlatılması olmuştur. Canlılar ve bir canlı topluluğunu meydana getiren bireyler arasındaki benzerlikler, bu canlıların ya da bireylerin biyolojik akrabalık dereceleri ile yakından ilişkilidir. Bir bireyin morfolojik ve anatomik özelliklerini nasıl kazandığını, neden kendi türüne ait bireylere diğerlerinden daha çok benzediğini, bu özelliklerin yavru bireylere nasıl aktarıldığını, bu özelliklerin aktarılışındaki kuralları, doğadaki biyolojik çeşitlenmenin nasıl meydana geldiğini inceleyen bilim dalı genetik'tir.

Populasyon ve alt-populasyonlar düzeyinde yapılan genetik çalışmalarda alel çokluklarının hesaplanması ve Hardy-Weinberg Kuralı uzantısında homozigot ve heterozigot çokluklarının, doğal seçim, mutasyon, göç, genetik sürüklenme ve rasgele olmayan çiftleşme durumları altında incelenmesi, ilişki (association), bağlantı (linkage), ayırım (segregation), haplotip çözümlemesi gibi konular olasılık ve istatistik modellerinin yardımıyla gerçekleştirilmektedir.

Hardy-Weinberg Kuralı'na göre; belli koşullar altında, rasgele eşleştirmeden bir kuşak sonra tek bir gen bölgesinde (Locus) gen tipi frekansları belli bir denge değerinde sabitlenir. Hardy-Weinberg Kuralı aynı zamanda denge frekanslarının bu bölgede alel frekanslarının basit bir fonksiyonu olarak gösterilebileceğini belirtir.

Kaynaklar

1. The Hardy-Weinberg Equilibrium, http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/H/Hardy_Weinberg.html, 16.04.2005, 14:17

2. Ensenauer R E, Reinke S S, Ackerman M J, Tester D J, Whiteman D A H, Tefferi A. Essentials of Medical Genetics for the Practicing Physician, Mayo Clin Proc. 2003;78:846-857
3. http://www.hi.is/~johannaj/ranns_verk/skyrsla_drog.doc, 12.04.2005, 19:33.
4. Hardy-Weinberg Principle, http://en.wikipedia.org/wiki/Hardy-Weinberg_principle, 17.06.2005, 10:22
5. Lange K., Mathematical and Statistical Methods for Genetics Analysis, 2nd Edition, USA, Springer-Verlag New York Inc.; 2002, 4-8

VAKA-KONTROL ÇALIŞMALARINDA HASTALIK-GEN İLİŞKİSİNİN İSTATİSTİKSEL ANALİZİ

Aysun Çetinyürek¹, Atilla Halil Elhan², Osman Tolga Kaskatı²

İstatistik ve Bilgisayar Bilimleri Bölümü, Başkent Üniversitesi¹

Biyoistatistik Anabilim Dalı, Tıp Fakültesi Ankara Üniversitesi²

1. Hastalık İlişki Çalışmaları

Bu çalışmada ilişki çalışmalarında sık kullanılan istatistiksel yöntemler tanıtılacaktır. İlişki çalışmaları genetik epidemiyolojide çokça kullanılan çalışmalardan biridir. Genetik epidemioloji, hastalık insidanslarında genetik materyalin kalıtsal etkisinin epidemiolojik açıdan değerlendirilmesidir. Bu alanda yapılan çalışmalar, herhangi bir hastalığın kalıtımı, hastalıkla ilgili olabilecek genlerin veya allellerin belirlenmesi yönünde yapılmaktadır¹.

Herhangi bir hastalık ve bir gen arasındaki ilişkiyi bulmak amacıyla yapılan çalışmalara "ilişki çalışmaları" denir. İlişki çalışmaları, hastalıkla ilgili genlerin yerini bulmak amacıyla yapılan bağlantı çalışmalarını tamamlayıcı nitelikte olan çalışmalardır². Dolayısıyla ilişki çalışmaları yapılırken hastalıkla ilgili genin yeri tespit edilmiş olduğu varsayılarak hastalıkla gen arasında ilişki olup olmadığı araştırılır³.

İlişki çalışmaları, popülasyon tabanlı (vaka kontrol çalışmaları) veya aile tabanlı olmak üzere iki türde yapılabilmektedir. Vaka kontrol çalışmaları da, kendi içinde allelik ve genotipik olmak üzere iki seviyede gerçekleştirilir. Burada kullanılan istatistiksel analizler gruplarına göre aşağıda verilmiştir.

- Allelik Analizler: Allel Testi ve Genotip Testi⁴
- Genotipik Analizler: Uyum İyiliği Testi ve Armitage Eğilim Testi⁵

Aile tabanlı çalışmalarda ise, Kalıtımda Dengesiz Aktarım Testi^{5,6} (TDT) kullanılır.

Bu çalışmada, hastalıkla genler arasında ilişki olup olmadığını test eden ilişki çalışmalarında kullanılan testler incelenecektir. Ayrıca farklı durumlarda testlerin uygunluğu ve varsayımlar açısından geçerlilikleri de incelenecektir. Bunun yanında durumlara göre bu testler arasındaki farklılıklara değinilecektir.

Anahtar Kelimeler: Vaka kontrol, İlişki çalışması, Armitage eğilim testi, Kalıtımda Dengesiz Aktarım testi.

Kaynaklar

1. Akarsu NA, Çakır B. Psikiyatrik genetik araştırmalarda kullanılacak istatistiksel yöntemler: IV-A. Hastalık geni haritalanması. 3P Dergisi 2004;12: 31-49.
2. Lecture notes on genetic epidemiology
http://www.gmap.net/topics/notes_02.htm
3. Cheng KF, Chen JH. Bayesian Models for Population-Based Case Control Studies When the Population is in Hardy Weinberg Equilibrium. Genetic Epidemiology 2005; 28: 183-192.
4. Nielsen DM, Weir BS. Association Studies under General Disease Models. Theoretical Population Biology 2001; 60: 253-263
5. Human Population Genetic Data Analysis, Summer Institute in Statistical Genetics Ders Notları, 2005.
6. Schaid DJ. Likelihoods and TDT for the case-parents design. Genet Epidemiol 1999;16(3):250-60.

GENELLEŞTİRİLMİŞ SIRALI LOJİSTİK REGRESYON MODELİ

N. Anıl Barak¹, Osman Saraçbaşı¹, Meltem Halil²,
Servet Arıoğul²

Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, ANKARA¹

Hacettepe Üniversitesi Tıp Fakültesi İç Hastalıkları Anabilim Dalı, Geriatri
Ünitesi, ANKARA²

e-mail: anilbarak@yahoo.com

Özet

Sıralı Lojistik Regresyon modelleri, çok kategorili ve sıralı yapıdaki bağımlı değişken olasılıklarını bağımsız değişkenler yardımı ile tahmin etmek için kullanılmaktadır. Sıralı Lojistik Regresyon modellerinde, bağımlı değişken kategorileri karşılaştırmaları için kullanılan farklı *Logit* oluşturma biçimleri bulunmaktadır. Bunlar içerisinde yorumlanması en kolay olan ve en sık kullanılanı Yığılımlı *Logit*'lerdir. Yığılımlı *Logit*'ler; belirlenen kategori ve daha düşük kategorilere düşme olasılığını birlikte ele almakta ve dolayısı ile sıralı bağımlı değişkenin yapısındaki tüm bilgileri kullanmaktadır. Yığılımlı *Logit* modelleri, Orantısal Risk varsayımının (Proportional Odds Assumption) sağlandığı ve sağlanmadığı durumlar için iki ana başlık altında incelenmektedir. Bunlar Orantısal Risk modeli (Proportional Odds Model: McCullagh 1980) ve Orantısal Olmayan Risk modelidir (Non- Proportional Odds Model: Fu 1998). Orantısal Risk modelinde oluşturulan Yığılımlı *Logit*'lerin paralelligi varsayımı, literatürde Paralel Eğimler varsayımı (Parallel Slopes Assumption) ya da Orantısal Risk varsayımı (Proportional Odds Assumption) olarak da bilinmektedir. Ancak çoğu zaman bu varsayım sağlanamamaktadır. Bu gibi durumlarda sıklıkla yapılan hata, bağımlı değişken sıralı bir yapıda olmasına rağmen Multinomial Lojistik Regresyon modelini kullanmaktır. Bu çalışmanın amacı, Sıralı Lojistik Regresyon modellerinde Orantısal Risk varsayımının sağlanmadığı durumlarda kullanılan Orantısal Olmayan Risk modeli diğer adıyla Genelleştirilmiş Sıralı Lojistik Regresyon modelini tanıtmak ve bu modelin sağlık alanı ile ilgili bir uygulamasını yapmaktır.

Anahtar Kelimeler: Sıralı Lojistik Regresyon; Genelleştirilmiş Sıralı Lojistik Regresyon; Orantısal Risk Modeli; Orantısal Olmayan Risk Modeli; Orantısal Risk Varsayımı.

Giriş

Lojistik Regresyon modelleri açıklayıcı değişken kümesi bilindiğinde, kategorik bağımlı değişken olasılıklarını elde etmede kullanılan bir yöntemdir. Lojistik Regresyon modelleri bağımlı değişkenin yapısına göre üçe ayrılmaktadır. Bu modellerden İki durumlu Lojistik Regresyon modeli, kategorik bağımlı değişkenin iki durumlu (Örn: var-yok) olduğu durumda kullanılmaktadır. *Multinomial* Lojistik Regresyon modeli kategorik bağımlı değişkenin çok kategorili ve nominal (Örn: Medeni durum, Evli- Bekar-Boşanmış) olduğu durumlarda kullanılırken; çok kategorili ve sıralı bir yapı söz konusu ise (Örn: Likert tipi ölçekler, az-orta-çok) Sıralı Lojistik Regresyon modelleri kullanılmaktadır

Gereç ve Yöntem

1. Sıralı Lojistik Regresyon Modelleri (Ordinal Logistic Regression Models)

Bağımlı değişken kategorisinin ikiden fazla olması durumunda kullanılacak bir çok model bulunmaktadır. Bunlardan *Multinomial* Lojistik Regresyon modeli, çok kategorili nominal yapıdaki bağımlı değişken olasılıklarını tahmin etmede kullanılan bir yöntemdir. Bazı durumlarda ise çok kategorili bağımlı değişken sıralı bir yapıda olabilir. Bu tip yapılara en çok anket çalışmalarında likert tipi ölçeklerde ya da sağlık alanında hastalık şiddetini ölçmek için yapılan çalışmalarda rastlanmaktadır. Bu tip bağımlı değişken kategorileri genellikle 0,1,2,..vb. ya da 1,2,3,..vb. şeklinde kodlanabilmektedir. Bağımlı değişken kategorileri arasında açık sıralanabilir bir yapı vardır. Fakat ardışık kategoriler arasındaki uzaklıklar birbirlerine eşit değildir. Bağımlı değişken kategorileri aralık ölçek ile ölçülmediği ve ardışık kategoriler arasındaki uzaklıklar eşit olmadığından, bu tip veriler klasik regresyon ile kolayca modellenememektedir. Diğer yandan; bu tip verilere *Multinomial* Lojistik Regresyon analizini kullanmak ise sıralı yapıdaki bağımlı değişkenin bu yapısını gözardı edecek ve bağımlı değişkendeki tüm bilgileri kullanmakta yetersiz kalacaktır. Bu sebeplerden dolayı Sıralı Lojistik Regresyon modelleri bu tip verilerin çözümlenmesinde geniş bir kullanım alanı bulmuştur.

1.1. Yığılımlı *Logit* Modelleri (Cumulative Logit Models)

Sıralı Lojistik Regresyon modellerinde, bağımlı değişken kategori karşılaştırmalarında kullanılan farklı *Logit* oluşturma biçimleri bulunmaktadır. Bunlardan yorumlanması en kolay olan ve en sık kullanılan Yığılımlı *Logit*lerdir. Sıralı Lojistik Regresyon modellerinde Yığılımlı *Logit* dışında kullanılan diğer *Logit* modelleri ise Ardışık Kategori *Logit* (Adjacent- Category Logit) ve

Sürekli Oran *Logit* (Continuation Ratio Logit) modelleridir [1]. Ancak bu çalışma dahilinde bu modellerden bahsedilmeyecektir.

Yığılımlı *Logit* modelleri ise Orantısal Risk varsayımının (Proportional Odds Assumption) sağlandığı ve sağlanmadığı durumlar için iki ana başlık altında incelenebilmektedir. Bunlar Orantısal Risk modeli (Proportional Odds Model: McCullagh 1980) ve Orantısal Olmayan Risk modelidir (Non Proportional Odds Model: Fu 1998).

Bağımlı değişken kategorisinin sıralı olduğu durumda, kategori olasılıkları çoğu zaman yığılımlı olasılıklar cinsinden ifade edilmektedir. Yığılımlı olasılık; belirlenen kategori ve daha düşük kategorilere düşme olasılığını birlikte ele almakta ve dolayısı ile sıralı bağımlı değişkenin yapısındaki tüm bilgileri kullanmaktadır. Genel olarak j kategorisi için yığılımlı olasılık aşağıdaki şekilde tanımlanır:

$$P(y \leq j) = P(y = 1) + P(y = 2) + \dots + P(y = j) \quad (2.1)$$

Yığılımlı *Logit* modeli, *Logit* dönüşümünde kategori olasılığı ($P(y = j)$) yerine yığılımlı olasılığın ($P(y \leq j)$) kullanılmasıyla oluşturulmaktadır. *Logit* bağ fonksiyonunu kullanarak Yığılımlı *Logit* model aşağıdaki gibi yazılmaktadır [1]:

$$\text{logit}[P(y \leq j)] = \log \left[\frac{P(y \leq j)}{1 - P(y \leq j)} \right] \quad j = 1, 2, \dots, J - 1 \quad (2.2)$$

Formül 2.2.'deki ($P(y \leq j)/1 - P(y \leq j)$) ifadesi, j 'inci bağımlı değişken kategorisi için Yığılımlı *Odds* (Cumulative Odds) olarak tanımlanmaktadır. Dolayısı ile Yığılımlı *Logit* modellerinde *Odds* j 'inci bağımlı değişken kategorisi için daha düşük kategorilere düşme olasılığının, daha yüksek kategoriye düşme olasılığı ile oranlanmasıyla elde edilir.

1.1.1. Orantısal Risk Modeli (Proportional Odds Model) ve Orantısal Risk Varsayımı (Proportional Odds Assumption)

Yığılımlı *Logit* modelinin özel bir durumu olan Orantısal Risk modeli aşağıdaki eşitlik ile verilmektedir.

$$\text{logit}[P(y \leq j)] = \tau_j - \sum_{k=1}^K \beta_k x_k \quad j = 1, 2, \dots, J - 1 \quad (2.3)$$

Orantısal Risk modelinde j 'inci kategori için yığılımlı olasılık değerleri, Lojistik dağılımı yardımı ile aşağıdaki şekilde elde edilir:

$$P(y \leq j) = F(\tau_j - \sum_{k=1}^K \beta_k x_k) = \frac{\exp\left(\tau_j - \sum_{k=1}^K \beta_k x_k\right)}{1 + \exp\left(\tau_j - \sum_{k=1}^K \beta_k x_k\right)} \quad j = 1, 2, \dots, J-1 \quad (2.4)$$

Formül 2.3. ile verilen Orantısal Risk modelinde; her bir yığılımlı *Logit*'in τ_j ile gösterilen kendine ait bir eşik (threshold) değeri bulunmaktadır. Bağımlı değişken kategorileri $j = 1, 2, \dots, J-1$ ile gösterildiğinde formüldeki β_k katsayılarının bağımlı değişken kategorilerinden bağımsız olduğu görülmektedir. Diğer bir ifade ile; bağımsız değişkenlerin bağımlı değişken *odds*'una (yığılımlı *odds*), hangi kategoride olduğuna bakılmaksızın sabit bir etkisi vardır ve tüm yığılımlı *Logit*'lerde k 'inci bağımsız değişkene ait β_k katsayıları birbirine eşittir. Bu da sağlanması zor olan ve öncelikle doğruluğunun test edilmesi gerekli bir varsayımdır. Orantısal Risk modelinde oluşturulan Yığılımlı *Logit*'lerin paralelliği varsayımı, literatürde Paralel Eğimler varsayımı (Parallel Slopes Assumption) ya da Orantısal Risk varsayımı (Proportional Odds Assumption) olarak da bilinmektedir.

1.1.2. Orantısal Olmayan Risk Modeli (Genelleştirilmiş Sıralı Lojistik Regresyon Modeli) (Non-Proportional Odds Model)

Orantısal Risk modelinde, k 'inci bağımsız değişkenin bağımlı değişkene etkisi (β_k) oluşturulan her bir yığılımlı *Logit*'de sabit olduğundan daha önce bahsedilmişti. Orantısal Risk varsayımı adı verilen bu varsayımın kimi durumlarda sağlanması zor olabilir. Orantısal Risk varsayımının sağlanmadığı durumlarda sıklıkla yapılan hata; bağımlı değişken sıralı yapıda olmasına rağmen *Multinomial* Lojistik Regresyon çözümlenmesine gitmektir. Fu (1998) tarafından önerilen Orantısal Olmayan Risk modelinde, *Multinomial* Lojistik Regresyon modelinin aksine *Logit* oluşturma esnasında Yığılımlı *Logit*'ler kullanılmaktadır ancak Orantısal Risk varsayımı sağlanmamaktadır. Diğer bir ifade ile; bağımsız değişkenlerin bağımlı değişken *odds*'una (yığılımlı *odds*) etkisi artık sabit değildir ve bağımlı değişken kategorisi $j = 1, 2, \dots, J-1$ ile gösterildiğinde β_{jk} katsayıları her bir bağımlı değişken kategorisi için farklıdır.

Dolayısı ile Yığılımlı *Logit* modeli yardımıyla oluşturulan Genelleştirilmiş Sıralı Lojistik Regresyon modeli, aşağıdaki gibi ifade edilir:

$$\text{logit}[P(y \leq j)] = \tau_j - \sum_{k=1}^K \beta_{jk} x_k \quad j = 1, 2, \dots, J-1 \quad (2.5)$$

Orantısız Olmayan Risk modelinde (diğer adıyla Genelleştirilmiş Sıralı Lojistik Regresyon modeli), j 'inci kategori için yığılımlı olasılık değeri ise aşağıdaki şekilde tanımlanır [2]:

$$P(y \leq j) = F\left(\tau_j - \sum_{k=1}^K \beta_{jk} x_k\right) = \frac{\exp\left(\tau_j - \sum_{k=1}^K \beta_{jk} x_k\right)}{1 + \exp\left(\tau_j - \sum_{k=1}^K \beta_{jk} x_k\right)} \quad j = 1, 2, \dots, J-1 \quad (2.6)$$

Orantısız ve Orantısız Olmayan Risk Modellerinde β_k ve β_{jk} katsayılarının önünde (-) işareti bulunmaktadır. Bu negatif işareti anlamı; pozitif bir β katsayısı ile daha düşük kategoriye düşme olasılığının ($P(y \leq j)$) ters orantılı olduğudur. Diğer bir ifade ile; pozitif bir β katsayısı, daha düşük kategoriye düşme olasılığının azaldığını, daha yüksek bir kategoriye düşme olasılığının ($P(y \geq j)$) ise arttığını ifade etmektedir [1], [3], [4]. Tam tersi bir durum için; negatif bir β katsayısı, daha düşük kategoriye düşme olasılığının ($P(y \leq j)$) arttığını, daha yüksek kategoriye düşme olasılığının ($P(y \geq j)$) ise azaldığını ifade etmektedir.

2. Orantısız Risk Varsayımının Geçerliliğinin Testi (Test of Proportional Odds Assumption)

Orantısız Risk varsayımının sağlandığı durumda Yığılımlı *Logit*'lerde k 'inci bağımsız değişkene ait β_k katsayıları birbirine eşit olacağından Orantısız Risk varsayımının testi için bu katsayıların birbirine eşit olup olmadığı aşağıdaki hipotez ile test edilir:

$$H_0 : \beta_{1k} = \beta_{2k} = \dots = \beta_{(J-1)k} \quad j = 1, 2, \dots, J-1 \quad (2.7)$$

Orantısal Risk modeli, Genelleştirilmiş Sıralı Lojistik Regresyon modelinin kısıtlanmış bir halidir. Dolayısı ile bu iki model içiçe geçmiş (nested) modellerdir. Orantısal Risk modelinden elde edilen Log- Olabilirlik L_1 , Genelleştirilmiş Sıralı Lojistik Regresyon elde edilen Log- Olabilirlik L_2 ile gösterildiğinde Orantısal Risk varsayımının yaklaşık olarak geçerliliğinin testi, Olabilirlik Oranı test istatistiği ile aşağıdaki gibi yapılmaktadır.

$$G^2 = -2(L_1 - L_2) \quad G^2 \sim \chi^2_{[K(J-2)]} \quad (2.8)$$

G^2 Olabilirlik oranı test istatistiği $\chi^2_{[K(J-2)]}$ değeri ile karşılaştırılarak test edilir. Eğer H_0 red edilirse, Orantısal Risk varsayımı geçerli olmayacağından Genelleştirilmiş Sıralı Lojistik Regresyon modeli kullanılmalıdır. Ancak bu testin yaklaşık bir test olduğu unutulmamalıdır. Ayrıca bu test ile modeldeki hangi bağımsız değişken veya değişkenlerin orantısallığı bozulduğu hakkında bilgi edinilememektedir.

Orantısal Risk varsayımının geçerliliğinin testi için kullanılan bir diğer test ise Brant (1990) tarafından önerilen Wald test istatistiğidir. Bu test hem genel olarak orantısallığın test etmekte, hem de orantısallığın hangi değişken veya değişkenler yüzünden bozulduğu hakkında bilgi vermektedir. Ancak test işlemleri yoğun matris işlemleri gerektirdiğinden el ile hesaplamalar oldukça güçtür. İstatistiksel paket programları (STATA) Brant testi adı altında bu testi yapmaktadır [2].

Uygulama

1. Uygulama Verisi ve Değişkenlerin Tanımı:

Sıralı Lojistik Regresyon modellerinden Genelleştirilmiş Sıralı Lojistik Regresyon modelinin uygulaması için Hacettepe Üniversitesi İç Hastalıkları Anabilim Dalı Geriatri ünitesine Şubat 2002 ile Haziran 2005 tarihleri arasında başvuran 1417 hastanın verileri kullanılmıştır. Çalışmada sıralı yapıdaki bağımlı değişken Framingham Risk Skorudur (FRS). Framingham Risk Skoru üç gruba aşağıdaki şekilde ayrılmıştır:

$$FRS = \begin{cases} 0 : \text{Düşük Risk} & \text{Eğer } 0 \leq FRS \leq 9 \\ 1 : \text{Orta Risk} & \text{Eğer } 10 \leq FRS \leq 19 \\ 2 : \text{Yüksek Risk} & \text{Eğer } 20 \leq FRS \end{cases}$$

Framingham Risk Skorunu (FRS) etkilediği düşünülen bağımsız değişkenler olarak ise hastanın Homosistein, Ferritin, Trigliserit, Ürikasit ve Kreatin değerleri sürekli değişken olarak alınmıştır.

2. Bulgular

2.1. Orantısal Risk Modeli

Aşağıdaki tabloda Orantısal Risk modeli için elde edilen katsayılar, standart hataları, *Odds* oranları ve katsayıların önemlilik testleri verilmektedir:

Tablo 1. Orantısal Risk Modelinin Regresyon Katsayı, Standart Hata ve *Odds* Oranı Tahminleri

Framingham Risk Skor Grubu Karşılaştırmaları	Değişkenler	Beta	Standart Hata	<i>Odds</i> Oranı	p değeri
Düşük Riske göre Orta ve Yüksek Risk karşılaştırılması	Eşik 1	2,5916	0,24665		
Düşük ve Orta Riske göre Yüksek Risk karşılaştırılması	Eşik 2	4,3264	0,26472		
	Homosistein	0,0123	0,00602	1,012	0,040
	Ferritin	0,0017	0,00061	1,001	0,004
	Trigliserit	0,0025	0,00068	1,002	0,000
	Ürikasit	0,0666	0,03323	1,068	0,045
	Kreatin	2,2181	0,24448	9,190	0,000

Orantısal Risk modeli istatistiksel olarak anlamlı bulunmuştur (Log-Olabilirlik= -1501,2948 sd= 5 p= 0,0000 Pseudo R²= 0,0617). Ancak bu model kullanılmadan önce mutlaka Orantısal Risk varsayımı test edilmelidir.

2.2. Orantısal Olmayan Risk Modeli (Genelleştirilmiş Sıralı Lojistik Regresyon Modeli)

Aşağıdaki tabloda Orantısal Olmayan Risk modeli için elde edilen katsayılar, standart hataları, *Odds* oranları ve katsayıların önemlilik testleri verilmektedir:

Tablo 2. Orantısız Olmayan Risk Modelinin Regresyon Katsayı, Standart Hata ve Odds Oranı Tahminleri

Framingham Risk Skor Grubu Karşılaştırmaları	Değişkenler	Beta	Standart Hata	Odds Oranı	p değeri
Düşük Riske göre Orta ve Yüksek Risk karşılaştırılması	Eşik 1	-3,0882	0,2995		
	Homosistein	0,0144	0,0073	1,014	0,048
	Ferritin	0,0025	0,0008	1,002	0,001
	Trigliserit	0,0019	0,0008	1,0019	0,014
	Ürikasit	0,0944	0,0402	1,099	0,018
	Kreatin	2,6369	0,3147	13,97	0,000
Düşük ve Orta Riske göre Yüksek Risk karşılaştırılması	Eşik 2	-3,8071	0,3123		
	Homosistein	0,0095	0,0074	1,0095	0,198
	Ferritin	0,0011	0,0007	1,001	0,119
	Trigliserit	0,0032	0,0008	1,003	0,000
	Ürikasit	0,0202	0,0437	1,02	0,644
	Kreatin	1,9774	0,2821	7,224	0,000

Orantısız Olmayan Risk modeli istatistiksel olarak anlamlı bulunmuştur (Log- Olabilirlik= -1492,1328 sd= 10 p= 0,0000 Pseudo R²= 0,0674).

Orantısız Risk modelinden elde edilen Log- Olabilirlik L₁, Genelleştirilmiş Sıralı Lojistik Regresyon elde edilen Log- Olabilirlik L₂ ile gösterildiğinde Orantısız Risk varsayımının yaklaşık olarak testi, Olabilirlik Oranı test istatistiği ile aşağıdaki gibi yapılmaktadır.

$$G^2 = -2(L_1 - L_2) = -2[(-1501,2948) - (-1492,1328)] = -2(-9,162) = 18,324$$

G² = 18,324 değeri 5 serbestlik derecesindeki Ki-kare değerinden büyük olduğu için Orantısız Risk varsayımı sağlanmamaktadır. Orantısız Risk varsayımının geçerliliğinin test eden diğer bir test olan Brant testini STATA paket programında uyguladığımızda aşağıdaki tablodaki sonuçlara ulaşırız:

Tablo 3. Orantısal Risk varsayımının Geçerliliği Testi (Brant Testi)

Değişken	Ki-Kare değeri	p değeri	Sd
Tümü	22,31	0,000	5
Homosistein	0,29	0,592	1
Ferritin	4,18	0,041	1
Tg	1,78	0,182	1
Ürikasit	1,86	0,172	1
Kreatin	6,92	0,009	1

Tablo 3'e göre de Orantısal Risk varsayımı sağlanmamaktadır. Bu testin orantısallığın hangi değişken veya değişkenler yüzünden bozulduğu hakkında bilgi verdiği için daha önce bahsedilmiştir. Tabloya göre orantısallığı bozan bağımsız değişkenler Ferritin ve Kreatin olmaktadır.

Sonuç ve Tartışma

Orantısal Risk varsayımı sağlanmadığından dolayı, Genelleştirilmiş Sıralı Lojistik Regresyon modeli verimiz için geçerli olan modeldir ve Tablo 2'den yararlanılarak Odds oranları yorumlamaları yapılır.

Hastanın Homosistein değerindeki bir birimlik artış, düşük ($0 \leq \text{FRS} \leq 9$) riske göre orta ve yüksek risk ($10 \leq \text{FRS}$) grubunda olma olasılığını 1,014 kat, Ferritin değerindeki bir birimlik artış aynı olasılığı 1,002 kat, Trigliserit değerindeki bir birimlik artış 1,0019 kat, Ürikasit değerindeki bir birimlik artış 1,099 kat ve Kreatin değerindeki bir birimlik artış 13,97 kat arttırmaktadır. Bütün bu artışlar istatistiksel olarak anlamlıdır.

Hastanın Homosistein değerindeki bir birimlik artış, düşük ve orta ($\text{FRS} \leq 19$) riske göre daha yüksek ($20 \leq \text{FRS}$) risk grubunda olma olasılığını 1,0095 kat, Ferritin değerindeki bir birimlik artış aynı olasılığı 1,001 kat, Trigliserit değerindeki bir birimlik artış 1,003 kat, Ürikasit değerindeki bir birimlik artış 1,02 kat ve Kreatin değerindeki bir birimlik artış 7,224 kat arttırmaktadır. Ancak bu artışlardan yalnızca Trigliseritin ve Kreatinin yarattığı risk artışları istatistiksel açıdan anlamlıdır.

Genelleştirilmiş Sıralı Lojistik Regresyon modeli, Orantısal Risk modelinde sağlanması zor bir varsayım olan Orantısal Risk varsayımının sağlanmadığı durumlarda kullanılması açısından bizlere avantaj sağlamaktadır. Multinomial Lojistik Regresyon çözümlemesinin aksine model

oluşturmada kullandığı Yığılımlı *Logit*'ler sayesinde, sıralı bağımlı değişkenin yapısındaki tüm bilgiyi kullanmaktadır. Bu açıdan Orantısız Risk varsayımının sağlanmadığı durumlarda Genelleştirilmiş Sıralı Lojistik Regresyon modeli, Multinomial Lojistik Regresyon modeline tercih edilmelidir.

Kaynaklar

1. Agresti, A. (2002), *Categorical Data Analysis*. New York: Wiley.
2. Hardin J., Hilbe J. (2001), *Generalized Linear Models and Extensions*. College Station, Texas: Stata Press.
3. Hosmer, D. W., Lemeshow, S. (2000), *Applied Logistic Regression*. 2nd ed. New York: Wiley.
4. Liao, T. F. (1994), *Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models*. Thousand Oaks, CA: Sage Publications.
5. McCullagh, P. (1980), Regression models for ordinal data. *J. Roy. Statist. Soc. Ser. B* 42: 109-142.
6. Fox, J. (1997), *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage Publications.
7. *Statistical Computing Seminars: Beyond Binary: Ordinal Logistic Regression in Stata*.
http://www.ats.ucla.edu/stat/stata/seminars/stata_BeyondBinaryLogistic/Stata_ordinal_seminar.htm

EŞ GRUPLARIN ANALİZİNDE ŞARTLI LOJİSTİK REGRESYON MODELİNİN KULLANIMI

Sema Erden, Handan Çamdeviren, Semra Erdoğan

Mersin Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Mersin

e-mail: semaerden@mersin.edu.tr

Özet

Bu çalışmada şartlı lojistik regresyon modelinin önemi ve kullanım alanları üzerinde durulmuş ayrıca 12-19 yaş arasında olan bireylerdeki intihar girişimi üzerine etkisi olan etkenlerin bu yöntem aracılığı ile belirlenmesi hedeflenmiştir.

Bazı durumlarda bir veya daha fazla değişken lojistik regresyon modeline etki karıştırıcı değişken olarak girebilir. Pratikte genellikle yaş ve cinsiyet etki karıştırıcı değişken olarak dikkate alınır. Böyle durumlarda söz konusu etki karıştırıcı değişkenin seviyelerine göre eş gruplar oluşturulur. Veri parçalandığı için oluşan bu gruplarda az sayıda gözlem ve çok sayıda parametre yer alır. Bu durumda model parametrelerinin tahmininde, şartlı olmayan tahmin yöntemleri yanlış sonuçlar verir. Bu nedenle şartlı tahmin yöntemleri kullanmak gerekir.

Çalışmada kullanılan örnek veri setinin sonuçları incelendiğinde ise yaşın etki karıştırıcı bir değişken olduğu görülmüş ve hem standart hem de şartlı lojistik regresyon analizleri aynı değişkenleri anlamlı bulmasına karşılık, şartlı lojistik regresyon modelin katsayıları, standart lojistik model katsayılarına göre daha büyük değer almıştır.

Giriş

Biyolojik bilimlerde, birkaç tane açıklayıcı değişkenin fonksiyonu olarak binary bağımlı bir değişkenin görülme olasılığını belirlemede en yaygın kullanılan yöntem lojistik regresyon modelidir. Genellikle açıklayıcı değişkenlerin bir kısmı ilgi alanına girer diğerleri de ilgilenilen açıklayıcı değişkenin veya sonuç değişkeninin biri veya her ikisi üzerindeki etkisini kontrol altına almak için modele dahil edilir (1, 2, 3).

Bazı durumlarda bir veya daha fazla değişken lojistik regresyon modeline etki karıştırıcı (confounding) değişken olarak girebilir. Pratikte

genellikle yaş ve cinsiyet etki karıştırıcı değişken olarak dikkate alınır. Böyle durumlarda söz konusu etki karıştırıcı değişkenin (lerin) seviyelerine göre gruplama yapma ihtiyacı doğar. Bu gruplar tabaka (stratum) veya eş gruplar (matched) olarak adlandırılır. Bu gruplarda bir vakaya karşılık bir de kontrol bireyi alınabileceği gibi (1:1 eş gruplar), bir vakaya karşılık birden fazla kontrol bireyi de (1:M eş gruplar) alınabilir. Pratikte 1:1 eş gruplar sıkça kullanılmaktadır. Bu durumda her bir tabakada bir vaka birde kontrol bireyi yer alır ancak mesela yaşa göre eşleme yapılıyorsa aynı yaşta olan birden fazla tabaka oluşturulabilir ve bu tabakalar ayrı birer kovaryat yapısı içerdiği için her biri ayrı bir tabaka olarak düşünülerek analizler yapılır. Bazı durumlar da oluşturulan tabakaların bir kısmı aynı yaşta bireyler olmayabilir. Bu gibi durumlarda vaka ve kontrol bireylerinin seçiminde en yakın yaş dikkate alınır. Eğer araştırma planlama aşamasında iken bir eşleme yapılması gerektiği biliniyorsa oluşturulacak tabakalara etki karıştırıcı değişkenin aynı değerlerine sahip bireyleri alma şansı artar. Ancak planlanıp veri elde edilmiş bir araştırmada, tabakalama işi yapılacaksa bu durumda her bir tabakaya giren bireyler etki karıştırıcı değişkenin aynı değerine sahip olamayabilir. Analizlerde bu durumların hangisi geçerli olduğu önceden belirlenmelidir. Çünkü birinci koşulda etki karıştırıcı değişken bir kovaryat değişkeni gibi modele alınmaz iken, ikinci koşulda etki karıştırıcı değişken modele alınmalıdır. Ayrıca denemede etki karıştırıcı değişken dışında, sonuç değişkeni üzerindeki etkisi araştırılan primer kovaryat değişkenleri de yer alır. Amaç bu değişkenlerin sonuç üzerine etkilerini belirlemektir. Ancak burada yapılan işlem etki karıştırıcının etkisini kontrol altına almak şeklinde özetlenebilir. Bu tip denemeler genellikle vaka-kontrol denemeleri veya kesitsel denemelerdir. Ancak kohort denemelerinde de böyle durumlarla karşılaşmaktadır (4).

Her bir tabakadaki denek sayısı, dikkate alınan kovaryat sayısına göre az olduğu için modeldeki katsayıların tahminin de şartlı olmayan (unconditional) likelihood fonksiyonu yani standart lojistik regresyon modeli kullanılamaz. Aksi takdirde katsayılar yanlı tahmin edilmiş olur. Bu gibi durumlarda en uygun yaklaşım şartlı (conditional) likelihood tahmin yöntemini kullanmaktır. Conditional lojistik regresyon modeli (CLR) fixed effect lojistik regresyon modeli olarak ta bilinir.

Materyal ve Metot

Bu çalışmada kullanılan veriler 2002 yılında Mersin ilinde yapılan kesitsel bir çalışmadan alınmıştır (5). Yapılan ankette ortaokul ve lisede okuyan çocukların depresyon durumları araştırılmış ve intihar girişimi olup olmadığı belirlenmiştir. Çalışmada toplam 4143 öğrenciye ait kayıt tutulmuş ve bunların 80'inde en az bir defa olmak üzere intihar girişimi olduğu tespit

edilmiştir. Modelde incelenmek üzere seçilen kovaryatlar ise sırasıyla şöyledir. Yaş, çocukların depresyon durumu, cinsiyet, okula devam durumları, sınıfta kalma durumları, anne babasıyla sorun yaşayıp yaşamadığı, somatik bir bulgusunun olup olmadığı, notlarında düşmenin olup olmadığı, arkadaşları ile sorun yaşayıp yaşamadığı, evde ceza verilip verilmediği, okulda sözel şiddete maruz kalıp kalmadığı, okulda dayak olup olmadığı ve sigara içip içmediğidir. Bu değişkenlerden sadece yaş sürekli yapıda olup diğerleri var veya yok şeklinde ifade edilen binary yapıdadır. İntihar girişiminin olup olmaması ise sonuç değişkeni olarak belirlenmiştir.

İntihar girişimi üzerine etkisi araştırılan bu etkenlerden hangisinin anlamlı düzeyde etkili olduğunu belirlemek için önce standart veya şartsız (unconditional) lojistik regresyon modeli oluşturulmuştur. Burada söz konusu risk faktörleri modele teker teker alınmış ve en sonunda uygun lojistik regresyon modeli oluşturulmuştur. İkinci aşamada yaş değişkeninin birçok literatürde (2) etki karıştırıcı değişken olarak dikkate alınması nedeniyle bu çalışmada da yaşların intihar girişiminde veya intihar girişimini etkileyebilecek diğer değişkenler üzerinde etkili olabileceği düşünülerek yaşa göre tabakalama yapıp eş gruplar oluşturulması planlanmıştır. Bu amaçla yaşları 12 ile 19 arasında olan ve intihar girişiminde bulunan bütün bireyler ($n_c=80$) analize alınmış ve aynı yaşlarda birer tanede intihar girişiminde bulunmayan öğrenci rasgele belirlenerek ($n_k=80$), 1:1 eş gruplar oluşturulmuştur. Daha sonra standart lojistik regresyon modelinden elde edilen final modelden yararlanarak etkisi anlamlı bulunan veya confounder ya da interaksiyon etkisi anlamlı olan kovaryat değişkenleri dikkate alınarak conditional lojistik regresyon modeli oluşturulmuştur. Model kurulumunda önce her bir eş gruba giren vaka (intihar girişiminde bulunan) ve kontrol (intihar girişiminde bulunmayan) bireylerinin kovaryat ölçümleri yan yana girilir. Daha sonra bu grupları gösteren bir indikatör değişken tanımlanır. Bu değişkenin değerleri kaç tane eş grup olduğunu gösterir ve bu tanımlanan değişken modele alınarak hangi değişkene göre eşleme yapıldıysa o değişkenin etkisi giderilmiş olur. Daha sonra her bir kovaryat değişkeni için her bir gruptaki vaka-kontrol farkları bulunur ve sonuç değişkeni "1" olarak kodlanır (vaka=1 ve kontrol=0 olarak kabul edildiğinde bu iki değer farkı=1 olur).

Hesaplama adımları şöyle özetlenebilir. Çalışmada m tane ($i=1,2,\dots,m$) eş grup olduğunu, her bir grup içinde n_i tane denek olduğunu kabul edelim. Bu gruplar içindeki j . deneğin sonuç değişkeni Y_{ij} ve kovaryat ölçümleri ise $x_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})^t$ olarak gösterilir. $Y_{ij}=1$ olma olasılığını π_{ij} ile gösterecek olursak bu olasılık, şartlı lojistik regresyon modelinde α_i ile gösterilen eş grup değişkeninin etkisi ve kovaryat değişkenlerinin bir lineer fonksiyonu olarak aşağıdaki gibi modellenabilir.

$$\pi_{ij} = \frac{\exp(\alpha_i + x'_{ij}\beta)}{1 + \exp(\alpha_i + x'_{ij}\beta)}$$

Ayrıca $(1-\pi_{ij}) = 1 - \frac{\exp(\alpha_i + x'_{ij}\beta)}{1 + \exp(\alpha_i + x'_{ij}\beta)}$ dir. Bu iki olasılığın bir birine oranı

rölatif (multiplicative) risk olup i. tabakadaki j. gözlemin 1 olma olasılığı ($Y_{ij}=1$) olarak tanımlanır. Bu oranın logaritması model parametreleri ile lineer bir ilişki içinde olup logit fonksiyon olarak tanımlanır. Bu fonksiyon aşağıdaki gibi sembolize edilir.

$$\text{logit}(\pi_{ij}) = \log(\pi_{ij} / (1 - \pi_{ij})) = \alpha_i + x_{ij}\beta$$

Bu model standart logistik regresyon modeline çok benzemekle beraber buradaki en önemli fark nuisance parametresi olarak adlandırılan α_i etkisinin modelden çıkarılmasıdır. Böylece likelihood fonksiyonu sadece bilinmeyen β parametrelerinin bir fonksiyonu olarak elde edilir.

k. tabakada yer alan vakalara ait kovaryat vektörü, x_{1k} , kontrol grubuna ait kovaryat vektörü ise x_{0k} olmak üzere bu tabaka için şartlı likelihood aşağıdaki gibi formülize edilir.

$$l_k(\beta) = \frac{\exp(x_{1k}\beta')}{1 + \exp(x_{1k}\beta')} \text{ bu eşitliği sadeleştirmek için pay ve payda}$$

$\exp(x_{0k}\beta')$ terimi ile bölünürse,

$l_k(\beta) = \exp(\beta'(x_{1k}-x_{0k})) / (1 + \exp(\beta'(x_{1k}-x_{0k})))$ elde edilir. Standart lojistik modelindeki sabit terimin sifıra eşit olması durumundaki model ile şartlı lojistik regresyon modelinin aynı olduğu görülmektedir. Yeni kovaryat vektörü ise vaka ve kontrol gruplarındaki deneklerin kovaryat değerlerinin farklarına eşittir. Gerçek kovaryat değişkenlerinin tipine bakmadan farkları modele sürekli bir değişken olarak alınır. Denemedeki örnek genişliği ise tabaka sayısına eşit kabul edilir (1, 2).

Bu yöntemde Newton-Rapson iterasyonu kullanılarak şartlı likelihood fonksiyonunun doğal logaritmasını maximum yapan değerler model parametrelerinin tahminlerini verir (6).

Hesaplamalarda StatDirect (ver. 2,4,4) istatistik paket programının deneme sürümü kullanılmıştır (7).

Sonuçlar

Hesaplamaların ilk adımında standart lojistik regresyon modeli kurulmuş ve değişkenlerin intihar girişimi üzerine etkilerine bakılmıştır. Kayıp gözlemler de dikkate alınınca toplam 3155 birey üzerinden hesaplamalar yürütülmüştür. Modelde kullanılan değişkenlere ait tanımlayıcı istatistikler Tablo 1’ de verilmiştir.

Önce yaşta dahil olmak üzere materyal bölümünde adı geçen bütün değişkenler modele alınmış ve bu modelde etkisi anlamlı bulunmayan değişkenler teker teker modelden çıkarılarak modelin yeni değerlendirmesi yapılmıştır. Kurulan full modeldeki katsayı tahminleri ve P değerleri Tablo 2’ de topluca verilmiştir.

Tablo 1. Modelde kullanılan değişkenlere ait tanımlayıcı istatistikler

DEĞİŞKENLER	KATEGORI	İNTİHAR GİRİŞİMİ OLAN	İNTİHAR GİRİŞİMİ OLMAYAN
BECK DEPRESYON DURUMU	<19	44 (%55)	2725 (%88,6)
	>=19	36 (%45)	350 (%11,4)
CİNSİYET	Erkek	40 (%50)	1606 (%52,2)
	Kız	40 (%50)	1469 (47.8)
OKULA DEVAM DURUMU	Yapan	65 (%81,3)	2156 (%70,1)
	Yapmayan	15 (%18,8)	919 (%29,9)
SİNİFTA KALMA	Evet	10 (%12,5)	282 (%9,2)
	Hayır	70 (%87,5)	2793 (%90,8)
ANNE BABA İLE SORUN	Evet	52 (%65)	522 (%17)
	Hayır	28 (%35)	2553 (%83)
SOMATİK HASTALIK	Var	62 (%77,5)	2278 (%74,1)
	Yok	18 (%22,3)	797 (%25,9)
NOTLARDA DÜŞME	Evet	50 (%62,5)	1291 (%42)
	Hayır	30 (%37,5)	1784 (%58)
SINIF ARK. SORUN	Evet	59 (%73,8)	1477 (%48)
	Hayır	21 (%26,3)	1598 (%52)
EVDE CEZA VAR MI?	Evet	38 (%47,5)	945 (%30,7)
	Hayır	42 (%52,5)	2130 (%69,3)
OKULDA KÜÇÜK DÜŞÜRÜLME VE ALAY VAR MI?	Evet	64 (%80)	1569 (%51)
	Hayır	16 (%20)	1506 (%49)
OKULDA DAYAK VAR MI?	Evet	52 (%65)	1183 (%38,5)
	Hayır	28 (%35)	1892 (%61,5)
HIÇ SİGARA İÇTİN Mİ?	Evet	41 (%51,3)	751 (%24,4)
	Hayır	39 (%48,7)	2324 (%75,6)
YAŞ	—	Ort±SD=15.13±1.66	Ort±SD=14.54±1.86

Tablo 2. Full standart lojistik regresyon modelindeki katsayılar ve anlamlılık düzeyleri

RİSK FAKTÖRLERİ	b	b' nin SE	P	OR	OR için %95 ihtimalli güven aralığı	
					ALT SINIR	ÜST SINIR
BEC_DEP	0.993	0.257	0.000	2.699	1.630	4.471
YAŞ	0.125	0.074	0.092	1.133	0.980	1.310
AB_SORUN	1.538	0.270	0.000	4.655	2.742	7.905
SÖZEL ŞİDDET	0.553	0.404	0.171	1.738	0.787	3.839
FİZİKSEL ŞİDDET	0.403	0.349	0.249	1.496	0.754	2.967
CİNSİYET	-0.057	0.241	0.812	0.944	0.589	1.513
OKUL DEVAM DUR.	-0.235	0.307	0.444	0.790	0.433	1.443
SNF_KALMA	-0.118	0.364	0.746	0.889	0.435	1.815
SOMAT	-0.294	0.283	0.298	0.745	0.428	1.297
NOT DÜŞMESİ	0.429	0.246	0.081	1.536	0.948	2.489
ARKADAŞLARIYLA SORUN	0.321	0.280	0.251	1.378	0.797	2.384
CEZA	-0.109	0.246	0.659	0.897	0.554	1.453

Modelden çıkarma sırasının belirlenmesinde değişkenlere ait “P” değerlerinin büyüklüğü dikkate alınmıştır. Standart lojistik regresyon analizi sonucunda final modelde yaş, beck depresyon olup olmaması, anne babası ile sorun yaşayıp yaşamaması, okulda sözel şiddete maruz kalıp kalmadığı ve yine okulda fiziksel şiddete maruz kalma durumu yer almaktadır. İntihar girişimi üzerine yaşların etkisi istatistik olarak anlamlı bulunmasa da etki karıştırıcı bir değişken olduğu yani modelden çıkarılınca diğer bazı kovaryatların katsayılarındaki değişim % 20'nin üzerinde olduğu için modelde bırakılmıştır. Bu etkilere ait katsayı tahminleri Tablo 3'te verilmiştir.

Sözel ve fiziksel şiddetler modelde bir arada buldukları zaman intihar girişimini anlamlı düzeyde etkilemedikleri gözlenmiş ancak fiziksel şiddet modelden atılınca sözel şiddete ait P değeri 0.182'den 0.005'e düşmüştür. Diğer katsayılarda önemli bir değişme gözlenmemiştir. Bu durumda fiziksel ve sözel şiddet arasında bir interaksyon olabileceği şüphe edildiği için bu değişkenlere ait esas etkiler modelden atılıp interaksyon etkileri incelenmiştir. Bu durumda elde edilen final model ise Tablo 4'te verilmiştir. Bu durumda bir arada modelde buldukları zaman esas etkileri anlamlı olmasa bile birbirleriyle interaksyon etkileri anlamlı olabileceği için modele bunların interaksyon etkileri de ilave edilmiş ve bu etkinin anlamlı

olduğu görülmüştür. İntihar girişimi üzerine fiziksel ve sözel şiddetin birlikte etkisi Tablo 5' te açıkça görülmektedir.

Tablo 3. Esas etkileri içeren final model

RİSK FAKTÖRLERİ	b	b' nin SE	P	OR	OR için % 95 ihtimalli güven aralığı	
					ALT SINIR	ÜST SINIR
BEC_DEP	1.130	0.251	0.000	3.094	1.891	5.065
YAŞ	0.107	0.070	0.125	1.133	0.971	1.276
AB_SORUN	1.667	0.260	0.000	5.296	3.181	8.817
SÖZEL ŞİDDET	0.620	0.401	0.182	1.859	0.847	4.080
FİZİKSEL ŞİDDET	0.381	0.346	0.271	1.464	0.742	2.886

Tablo 4. İnteraksiyon etkilerini içeren final modele ait katsayı tahminleri

RİSK FAKTÖRLERİ	b	b' nin SE	P	OR	OR için %95 ihtimalli güven aralığı	
					ALT SINIR	ÜST SINIR
BEC_DEP	1.159	0.251	0.000	3.188	1.949	5.213
YAŞ	0.122	0.069	0.076	1.130	0.988	1.292
AB_SORUN	1.682	0.260	0.000	5.374	3.227	8.948
FİZİKSEL X SÖZEL ŞİDDET	0.786	0.251	0.002	2.194	1.342	3.586

Tablo 5. İntihar girişimi üzerine fiziksel ve sözel şiddetin birlikte etkisi

Okulda dayak var mı?	Çocukta intihar girişimi	Okulda küçük düşürülme var mı?	
		Hayır	Evet
hayır	olmadı	1506 (% 79.6)	386 (% 20.4)
	oldu	16 (% 57.1)	12 (% 42.9)
evet	olmadı	-	1183 (% 100)
	oldu	-	52 (% 100)

Final modelde yer alan değişkenler dikkate alınarak yaşa göre eş gruplar oluşturulmuş ve şartlı lojistik regresyon modeli kurulmuştur. Yaşa göre oluşturulan eş gruplarda yer alan birey sayıları Tablo 6'te verilmiştir.

Tablo 6. Yaşa göre oluşturulan eş gruplarda yer alan birey sayıları

Yaş	İntihar girişiminde bulunan	İntihar girişiminde bulunmayan
12	3	3
13	15	15
14	11	11
15	15	15
16	18	18
17	13	13
18	4	4
19	1	1

Şartlı lojistik regresyon modelinin elde edilmesi esnasında fiziksel şiddete maruz kalan birey sayısı az olduğu için bu değişkene ait katsayı tahmini doğru yapılamamıştır. Bu nedenle fiziksel şiddet değişkeni modelden çıkarılmış ve diğer 3 kovaryata ait katsayılar, yaşa göre düzeltilmiş olarak elde edilmiştir. Kurulan modeldeki katsayılar ve hipotez testi sonuçları Tablo 7'de görünmektedir.

Tablo 7. Sonuç modele ait katsayı tahminleri

RİSK FAKTÖRLERİ	b	b' nin SE	P	OR	OR için % 95 ihtimalli güven aralığı	
					ALT SINIR	ÜST SINIR
Beck depresyon	2.44	1.06	0.02	11.5	1.44	92.1
Anne-baba ile sorun	3.32	1.53	0.02	27.7	1.38	553.6
Sözel şiddet	3.20	1.21	0.008	24.7	2.30	264.1

Hem şartlı hem de standart lojistik regresyon modellerinden elde edilen sonuçlara göre beck depresyon durumu, anne-baba ile sorun yaşama ve okulda sözel şiddete maruz kalma durumunun intihar girişimi riskini artırdığını söyleyebiliriz. Ancak şartlı lojistik regresyon modelindeki katsayıların yaşa göre düzeltilmiş katsayılar olduğu ve standart lojistik regresyon model katsayılarına göre anlamlılık düzeylerinin de bir miktar değiştiği gözlenmiştir.

Şartlı lojistik regresyon modelinin sonuçları özetlenecek olursa, depresyon durumu arttıkça intihar girişim sıklığı 11.5 ($p=0.02$) kat artmakta,

anne baba ile sorun yaşama durumunda bu risk 27.7 ($p=0.02$) kat artmakta ve okulda sözel şiddete maruz kalma durumunda ise risk 24.7 ($p=0.008$) kat artmaktadır.

Tartışma

Şartlı lojistik regresyon analizini yapmak amacıyla kullanılan Statsdirect istatistik paket programında interaksiyon terimlerini kurulan modele dahil etme seçeneği sunulmadığı için interaksiyon terimi incelenmemiştir.

Bu konuyla ilgili yapılmış araştırmalarda, özellikle epidemiyolojik çalışmalarda, bazı biyolojik önem taşıyan değişkenlerin, kullanılan diğer değişkenler üzerine etkilerini gidermek ve asıl ilgilenilen değişkenlere ait düzeltilmiş katsayıları tahmin etmek için, bu değişkenlere göre tabakalama işlemi yapılması gerekir. Böyle durumlarda gözlem sayısı azaldığı buna karşılık tahmin edilecek parametre sayısının arttığı gözlenebilir. Yansız tahmin ediciler elde etmek için şartlı tahmin yöntemi tercih edilmelidir (2).

Ayrıca denemelerde etkili değişkenlerin yer alması durumunda bu değişkenlere göre eşleme yapılırken eşleme oranının iyi belirlenmesi gerekir.

Kaynaklar

1. Breslow NE, Day NE. Statistical Methods in Cancer Research, Vol. 1. The Analysis of Case-Control Studies. Oxford University Press. 1980.
2. Hosmer DW, Lemeshow S. Applied Logistic Regression. John Wiley and Sons, Inc. 1989.
3. Armitage P, Berry G. Statistical Methods in Medical Research (3rd edition). Blackwell 1994.
4. Lindsey JK. Directly Modelling Matched Case-Control Data. Statist. Med. 2000; 19: 35-44.
5. Toros F, Bilgin NG, Şaşmaz T, Buğdaycı R, Çamdeviren H. Suicide Attempts and Risk Factors Among Children and Adolescents," Yonsei Medical Journal, 2004; 45(3): 367-374.
6. Krailo MD, Pike MC. Algorithm AS 196. Conditional multivariate logistic analysis of stratified case-control studies. Applied Statistics 1984; 33: 95-103.
7. StatDirect Statistical Software (ver. 2,4,4). (14-02-2005) Trial version. <http://www.statdirect.com>

SİMMENTAL X GÜNEY ANADOLU KIRMIZISI SIĞIRLARINA AİT CANLI AĞIRLIK ÖLÇÜMLERİ İÇİN LOJİSTİK BÜYÜME EĞRİSİ

Cemil Çolak¹, Mehmet N. Orman²

Başbakanlık Gümrük Müsteşarlığı Gümrükler Genel Müdürlüğü, Ankara¹

Ankara Üniversitesi Veteriner Fakültesi Biyoistatistik AD, Ankara²

Özet

Bu çalışmada, Simmental x Güney Anadolu Kırmızısı (GAK) sığır ırkı G_1 ve $F_1 \times G_1$ melezlerine ilişkin canlı ağırlık ölçümleri için doğrusal olmayan lojistik büyüme eğrileri oluşturulmuştur. Lojistik büyüme modeline ait artıklarda ortaya çıkabilecek özilişki sorunu incelenmiştir.

Modellerinin uyum iyiliği, hata kareler ortalaması, belirleme katsayısı ve iterasyon sayısı değerleri kullanılarak yapılmıştır.

Sonuç olarak, ağırlık-yaş sürecinin tanımlanmasında doğrusal olmayan lojistik modelinin başarılı olduğu uyum iyiliği ölçütleri ile belirlenmiştir.

Anahtar Kelimeler: Büyüme modeli, Sığır, Simmental, GAK.

GÖZLEME DAYALI ÇALIŞMALARDA PROPENSİTY SKOR: BİR UYGULAMA

Elif Çiğdem Altunok¹, Nural Bekiroğlu², Mehmet Genceli¹

¹Yıldız Teknik Üniversitesi, Fen-Edebiyat Fakültesi İstatistik Bölümü

²Marmara Üniversitesi, Tıp Fakültesi Biyoistatistik Anabilim Dalı

e-mail: caltunok@gmail.com

Özet

Gözleme dayalı çalışmalarda, araştırmacının tedavi ve kontrol gruplarındaki hastaları rastlantısal olarak gruplara atama işleminde kontrolü yoktur, bu nedenle tedavi ve kontrol gruplarının gözlenen ortak değişkenlerinde büyük farklar olabilir ve bu farklar tedavi etkisinin sistematik hatalı tahminlerine neden olabilir. Propensity skor, tedavinin gözlenen ortak değişkenlere göre koşullu olasılığı olarak tanımlanır. İki gruptaki ortak değişkenlerin dengelenmesinde ve böylelikle sistematik hatanın azaltılmasında kullanılır. Propensity skor genellikle lojistik regresyon yardımıyla tahmin edilir ve propensity skor bir kez tahmin edildikten sonra eşleştirme, tabakalara ayırma, regresyon düzeltilmesi veya bu üçünün bileşiminin kullanılması yöntemleriyle sistematik hatanın azaltılması amaçlanır. Çalışmamızda, Marmara Üniversitesi Hastanesi Göğüs Cerrahisi bölümünde 1996–2003 yılları arasında aynı doktor tarafından göğüs cerrahisi ameliyatı geçirmiş N=478 hasta kullanılmıştır. Ameliyat sonrası delirium tanısı alan (POD) 18 hasta ve almayan 460 hastaya ait ameliyat öncesi 10 risk faktörü ile ve ameliyat sonrası 14 risk faktörüne lojistik regresyon uygulanmış ve sonuçlar elde edilmiştir. Propensity skor hesaplanmış, tabakalara ayırma yöntemi kullanılarak benzer propensity skora sahip tedavi ve kontrol bireyle oluşan yeni örneğe istatistik analiz uygulanmış ve değerlendirilmiştir. Propensity skor ile tabaklara ayırma yöntemi kullanılarak birbiri ile benzer dağılıma sahip tedavi ve kontrol bireyleri seçilmiş böylece yeni örneklerdeki tedavi ve kontrol grupları hemen hemen aynı karakteristiklere sahip olmuş ve sistematik hata azaltılmıştır.

Sonuç olarak, propensity skor öncesi ile propensity skor sonrası tedavi ve kontrol bireylerine uygulanan analiz sonuçları risk faktörlerinin anlamlılıklarında değişiklik gözlemlenmiştir.

Anahtar Kelimeler: Propensity skor, gözleme dayalı çalışmalar, sistematik hatanın azaltılması.

1. Giriş

Rassal deneylere ait verilerle gerçekleştirilen çalışmalar literatürde genellikle pek fazla değildir. Bu nedenle (rassal olmayan) gözleme dayalı çalışmalara daha çok rastlanır. Propensity skor, gözleme dayalı bir veri kümesinden hareket ederek tedavi etkilerinin tahminindeki (varsa) sistematik hatayı azaltmak için Rosenbaum ve Rubin (1983) tarafından bir yöntem olarak önerilmiştir. Gittikçe artarak kullanılan bu yöntem, daha ziyade tıbbi denemelerde ve ekonomik politika müdahalelerinin değerlendirilmesinde popüler hale gelmiştir.

Rassal deneylerde, bireylerin farklı tedavilere rassal olarak atanmaları, farklı tedavilere atananlar arasındaki gözlenen ve gözlenmeyen ortak değişkenlerde sistematik hatanın mümkün olduğu kadar azalmasını sağlar. Buna karşın, gözleme dayalı çalışmalarda araştırmacının tedavi ve kontrol gruplarındaki hastaları gruplara atama işleminde kontrolü yoktur. Bunun nedeni örneğin, sigara içmenin veya radon gazının zararlı olduğu gösteren çalışmalarda insanlara bunların deney amacıyla verilememesinden kaynaklanır [1]. Gözleme dayalı çalışmalarda rastlantısal olarak atanmayan tedavi ve kontrol gruplarının gerek demografik özelliklerinde gerekse gözlenen ortak değişkenlerinde büyük farklar olabilir ve bu durumda tedavi gruplarında çıktılar doğrudan karşılaştırılması yanıltıcı olur. Bu farklar tedavi etkisinin sistematik hatalı tahminlerine sebep olur. Çalışma tasarımı gözlenen ortak değişkenlerdeki bilgilerin birleşmesi ile örneğin eşleştirilmiş örnekleme yaparak veya tedavi etkisinin tahmininde tabakalara ayırma veya kovaryans düzeltmesi yaparak bu farklardan kaçınılabilir. Geleneksel düzeltme metotları (eşleştirme, tabakalara ayırma, kovaryans düzeltmesi) genellikle sınırlıdır, çünkü düzeltme için sadece sınırlı sayıda ortak değişken kullanılabilirler. Propensity skorda ise böyle bir sınırlama yoktur. Ortak değişken sayısı arttığında, aynı veya benzer eşleştirilmiş çiftler bulmak zorlaşır. Propensity skor çok sayıda ortak değişken içerdiğinde eşleştirilmiş setlerin ve tabakaların oluşturulması için bir düzenektir. Ortak değişken bilgilerinin özet ölçüsünü içerir.

2. Propensity Skor Tanımı ve Propensity Skorun Yararları

Propensity skor, tamamlanmış bir veri setine sahip $(i=1,2,\dots,N)$ deneklerde x_i olarak tanımlanır ve x_i tedavinin ($Z_i = 1$), kontrole ($Z_i = 0$) göre gözlenen ortak değişkenler vektörünün koşullu olasılığı olarak;

$$e(x_i) = pr(Z_i = 1 | X_i = x_i)$$

ifade edilir. Burada verilen X ler için Z_i lerin bağımsız olduğu varsayılır;

$$pr(Z_1 = z_1, \dots, Z_N = z_N | X_1 = x_1, \dots, X_N = x_N) = \prod_{i=1}^N e(x_i)^{z_i} \{1 - e(x_i)\}^{1-z_i} .$$

Propensity score dengeleme skorudur. Çünkü propensity skorlar veri olduğunda tedavi öncesi karakteristiklerin koşullu dağılımları tedavi ve kontrol grupları için aynıdır [2]. Açıkça ifade etmek gerekirse, propensity skor bir kişinin sadece ortak değişken skoru kullanılarak tedavi edilmesinin bir olasılık ölçüsüdür. Bahsi geçen yöntem iki gruptaki ortak değişkenlerin dengelenmesinde kullanılabilir ve böylelikle sistematik hatayı azaltır. Propensity skor gözleme dayalı çalışmalarda başlıca sistematik hatanın azaltılmasında, kesinliğin artırılmasında ve belirli ortak değişkenlerin etkilerini anlamak için kullanılır.

Ortak değişkenlerde kayıp veri durumu söz konusu değilse, propensity skor diskriminant analizi veya lojistik regresyon yardımıyla tahmin edilir. Ortak değişkenlerin çok değişkenli normal dağılıma sahip olduğu varsayıldığında diskriminant analizi kullanılır, lojistik regresyon için ise bu varsayıma gerek yoktur [3]. Propensity skoru hesaplamasının amacı, tedavi alacak bireyin olasılığının hesaplanması yardımıyla (propensity skor) tedavi etkisinin tahmininin düzeltilerek hemen hemen rassal bir deneyin etkisini yaratmaktır. Aynı propensity skor ile biri tedavi grubunda biri de kontrol grubunda iki birey bulursak, bu durumda iki bireyin her bir gruba rassal olarak atandığını, bir anlamda eşit benzerlikte tedavi veya kontrolde olduğu düşünülür.

Propensity skor bir kez tahmin edildikten sonra eşleştirme, tabakalara ayırma (sınıflara ayırma), regresyon düzeltilmesi veya bu üçünün bazı birleşiminin kullanılması yöntemleriyle sistematik hata azaltılabilir [2]. Bu yöntemler, tedavinin gözlenen ortak değişkenlere göre koşullu olasılığı olan $e(x) = pr(Z = 1 | X)$, propensity skoru tanımları nedeniyle oldukça kullanışlıdır ve bu $e(x)$ göre Z ve X koşullu olarak bağımsızdır anlamına gelir. Böylelikle, tedavi ve kontrol gruplarındaki eşit (veya eşite yakın) propensity skora sahip bireyler, kendi geçmiş ortak değişkenlerine bağlı olarak aynı (veya neredeyse aynı) dağılıma sahip olma eğiliminde olacaktır [3].

2.1 Eşleştirme

Genellikle araştırmacılar çalışmalarda, sınırlı sayıda tedavi hastasına ve büyük sayıda kontrol ile karşı karşıya kalırlar. Eşleştirme,

araştırmacılar tarafından sıklıkla araştırmaya göre kontrol edilmesine inanılan geçmiş ortak değişkenlerine bağlı olarak, tedavi bireyleri ile eşleşmiş kontrol bireyleri seçmek için kullanılır. Rosenbaum ve Rubin propensity skor kullanarak eşleşmiş örnek oluşturmak için üç teknik önermişlerdir [4].

- i. **Tahmin edilen propensity skora bağlı olarak mümkün olan en yakın eşleştirme:** Burada tedavi ve kontrol bireyleri rassal şekilde dizilir. İlk tedavi bireyi ile propensity skoru mümkün olan en yakın kontrol bireyi ile eşleştirilir. Her iki bireyde eşleştirme yapılacak örneklemden çıkartılır. Bir sonraki tedavi bireyi seçilir. Eşleştirme yapmak için tahmini propensity skorun logitlerinin kullanılması tavsiye edilir [4]. $\hat{q}(X) = \log[(1 - \hat{e}(X)) / \hat{e}(X)]$ kullanılır çünkü $\hat{q}(X)$ in dağılımı genellikle yaklaşık olarak normaldir.
- ii. **Propensity skoru içeren Mahalanobis metrik eşleştirmesi:** Ortak değişkenler çok değişkenli normal dağılıyorsa ve tedavi ve kontrol gruplarının ortak kovaryans matrisleri var ise bu durumda Mahalanobis metrik eşleştirmesi eşit yüzdeli sistematik hata azaltıcı tekniğidir. Burada sistematik hata tedaviyi alanların ortalaması eksi kontrollerin ortalamasıdır [5].
- iii. **Propensity skorlar tarafından tanımlanan aralıklar arasında mümkün olan en yakın Mahalanobis metrik eşleştirmesi:** Bu yöntem yukarıda anlatılan iki yöntemi birleştirir. Tüm tedavi bireyleri rassal şekilde sıralanır ve ilk tedavi bireyi seçilir. Tedavi bireylerinin tahmini propensity skorlarının değer aralığındaki tüm kontrol bireyleri seçilir ve az sayıda ortak değişkene bağlı olarak bu bireyler ile tedavi bireyi arasındaki Mahalanobis mesafesi hesap edilir. En yakın kontrol bireyi ve tedavi bireyi örneklemden çıkarılır ve bu süreç tekrarlanır.

2.2. Tabakalara Ayırma

Tabakalara ayırma yöntemi de gözleme dayalı çalışmalarda tedavi ve kontrol grupları arasındaki sistematik farkları kontrol etmek için kullanılır. Bu teknik gözlenen geçmiş karakteristikleri tarafından bireyleri tabakalara gruplar. Propensity skora bağlı olarak tam tabakalara ayırma yapıldığında, tabakalar arasında ortalama tedavi etkisi, gerçek tedavi etkisinin yansız (sistematik hatasız) tahmincisini üretir [2]. Propensity skora bağlı olarak tabakalara ayırma, propensity skoru tahmin etmek için kullanılan k ortak değişkeni dengeler ve genellikle propensity skora bağlı beş tabaka, bu ortak değişkenlerin her birindeki sistematik hatanın % 90'nını ortadan kaldırır [6][7]. İlk önce propensity skor lojistik regresyon veya diskriminant analizi ile tahmin edilir. Araştırmacı, her iki grubun birlikte veya tedavi grubunda ve

kontrol grubunda ayrı olarak propensity skor değerlerine bağlı olarak tabaka sınırlarına karar vermelidir.

2.3. Regresyon Düzeltmesi

Regresyon düzeltmesinde, tedavi etkisi t şu şekilde tahmin edilir. Burada t tedaviyi, c kontrolü belirtir.

$$\hat{t} = (\bar{Y}_t - \bar{Y}_c) - b(\bar{X}_t - \bar{X}_c)$$

Geçmiş ortak değişkenlerin etkisi, yukarıdaki denklemin sağ tarafındaki ikinci terim çıkartılarak düzeltilir. Burada b geçmiş ortak değişkenlere bağlı olarak tedavi ve kontrol grupları için bir regresyon tahminidir. Bu nedenden ötürü, tedavi ve kontrol gruplarında propensity skora bağlı yanıtların regresyonunu bulmak ve bunu kullanarak tedavi etkisinin en son tahminini düzeltmek için propensity skor değişkeninin kullanımı oldukça kullanışlıdır.

Çalışmamızda, Marmara Üniversitesi Hastanesi göğüs cerrahisi bölümünde 1996–2003 yılları arasında aynı doktor tarafından göğüs cerrahisi ameliyatı geçirmiş N=478 hasta kullanılmıştır. Ameliyat sonrası delirium tanısı alan 18 hasta ile ve 460 delirium tanısı almayan hastaya ait ameliyat öncesi 10 risk faktörü ve ameliyat sonrası 14 risk faktörüne istatistik analiz uygulanmıştır.

3. Bulgular

Tablo 1a. Ameliyat öncesi risk faktörlerinin tek değişkenli analizleri

	Delirium tanısı almayan hasta (n=460)	Delirium tanısı alan hasta (n=18)	p değeri †
Ameliyat Öncesi			
Yaş (YASKOD) ((18–59=0) (59>=1)	% 33	% 66	0,005**
Cinsiyet (CINS) (erkek=1,bayan=0)	% 67	% 61	0,38
Kronik Hastalık ^a (K2) (var=1,yok=0)	% 3	% 0	0,55
Alkol Kullanımı (K3) (var=1,yok)	% 19	% 44	0,017**
Psikiyatri sorunu (K4) (var=1,yok=0)	% 12	% 22	0,171
Serebrovasküler hastalık (K5) (var=1,yok=0)	% 0,6	% 0	0,891
Diyabet (DM) (var=1,yok=0)	% 7	% 22	0,046**
Kemoterapi (KEMO) (var=1,yok=0)	% 3	% 0	0,536
Tümör türü (TTÜR) (var=1, yok=0)	% 53	% 55	0,514
Operasyon türü acil (OTÜR) (1–5)	2,76 ± 1,15	2,88 ± 1,13	0,667 ^d

† Ki-kare testi p değerleri, ** p<0,005 ;^a Herhangi bir akciğer, kalp, böbrek ve karaciğer hastalığının mevcudiyeti

^d . t istatistiği p değeri (sürekli veri, ortalama ± standart sapma şeklinde gösterilmiştir.)

Tablo 1b. Ameliyat sonrası risk faktörlerinin tek değişkenli analizleri

	Delirium tanısı almayan hasta (n=460)	Delirium tanısı alan hasta (n=18)	p değeri †
Ameliyat Sonrası			
Solunum yetmezliği ^b (RD) (var=1,yok=0)	% 18	% 27	0,237
Serumdaki kimyasal değerlerin			
anormalliği ^c (ELEKTR) (var=1,yok=0)	% 5	% 27	0,004**
Operasyon süresi (AMLSÜRE) (var=1,yok=0)	% 38	% 72	0,004**
Yoğun bakım süresi (YBSTAY)	2,56 ± 7,89	3,05 ± 1,92	0,79 ^d
Uyku bozukluğu (UYKU) (var=1,yok=0)	% 0	% 11	0,001**
Hipertansiyon (K10) (var=1,yok=0)	% 12	% 22	0,272
Enfeksiyon (K11) (var=1,yok=0)	% 19	% 22	0,760
Kan transfüzyonu (K12) (var=1,yok=0)	% 29	% 22	0,606
Aminofilin kullanımı (AFİLİN) (var=1,yok=0)	% 3	% 5	0,464
Antiaritmi kullanımı (ARİTMİ) (var=1,yok=0)	% 7	% 11	0,371
Antibiyotik kullanımı (K16) (var=1,yok=0)	% 24	% 22	0,564
Stteroid kullanımı (STREOİD) (var=1,yok=0)	% 5	% 11	0,285
Antihipertansif kullanımı (K17) (var=1,yok=0)	% 16	% 16	0,576
Hareket edememe (IMMOB) (var=1,yok=0)	% 10	% 16	0,288

† Ki-kare testi p değerleri; ** p<0,005 ; ^b Oda havasındaki P_{O_2} <55 mm Hg kandaki oksijen miktarının yetersizliği ve P_{aO_2} >44 mm Hg kandaki anormal karbondioksit seviyesi; ^c sodyum <130 veya >150 meq/L, potasyum <3.0 veya >6 meq/L, glukoz <60 veya >300 mg/dL;

^d .t istatistiği p değeri (sürekli veri, ortalama ± standart sapma şeklinde gösterilmiştir.)

Tablo.1a ve Tablo.1b her iki grup için ayrı ayrı 24 risk faktörünün tanımlayıcı istatistiklerini içerir. Tedavi ve kontrol grupları arasında ortak değişkenler bakımından fark olup olmadığını görmek için iki örnek t testi, Kikare testi ve Fisher's exact test uygulanmıştır. Buna göre ameliyat öncesi yaş, alkol kullanımı, diyabet, operasyon türü risk faktörleri ile ameliyat sonrası serumdaki kimyasal değerlerin anormalliği, operasyon süresi, uyku bozukluğu risk faktörlerinin p değerleri 0.05 altında çıkmış, yani iki grup arasında ortak değişkenler bakımından fark olduğu sonucuna varılmıştır. Propensity skorları tahmin etmek için lojistik regresyon uygulanmış ve delirium tanısı için önemli risk faktörleri belirlenmiştir.

Tablo 2. Propensity skor düzeltmesi öncesi lojistik regresyon çıktısı

		Variables in the Equation						95.0% C.I. for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step	TTÜR(1)	,515	,966	,284	1	,594	1,673	,252	11,112
1	YASKOD(1)	-1,907	,884	4,654	1	,031	,149	,026	,840
	K2(1)	20,568	7836,056	,000	1	,998	8,6E+08	,000	.
	K3(1)	-,787	,920	,733	1	,392	,455	,075	2,760
	K4(1)	-2,095	,926	5,118	1	,024	,123	,020	,756
	K5(1)	-1,360	22029,444	,000	1	1,000	,257	,000	.
	AMLSÜRE(1)	-2,732	,968	7,960	1	,005	,065	,010	,434
	IMMOB(1)	,183	1,269	,021	1	,886	1,200	,100	14,422
	RD(1)	-,300	,775	,149	1	,699	,741	,162	3,386
	UYKU(1)	-43,443	28943,294	,000	1	,999	,000	,000	.
	K10(1)	-,032	1,058	,001	1	,976	,968	,122	7,708
	K11(1)	-,309	,964	,102	1	,749	,735	,111	4,862
	K12(1)	1,256	,810	2,407	1	,121	3,512	,718	17,176
	DM(1)	-1,198	,955	1,574	1	,210	,302	,046	1,961
	AFILIN(1)	-1,987	1,508	1,736	1	,188	,137	,007	2,634
	ARITM(1)	18,993	5811,123	,000	1	,997	1,8E+08	,000	.
	K16(1)	,533	,977	,298	1	,585	1,704	,251	11,568
	K17(1)	2,078	1,437	2,090	1	,148	7,988	,477	133,623
	KEMO(1)	16,539	8692,541	,000	1	,998	1,5E+07	,000	.
	STEROID(1)	-1,688	1,366	1,527	1	,217	,185	,013	2,688
	ELEKTR(1)	-1,944	,939	4,290	1	,038	,143	,023	,901
	CINS(1)	1,149	,797	2,078	1	,149	3,156	,661	15,058
	OTÜR			1,432	4	,839			
	OTÜR(1)	-42,552	25344,458	,000	1	,999	,000	,000	.
	OTÜR(2)	-24,325	25075,102	,000	1	,999	,000	,000	.
	OTÜR(3)	-24,169	25075,102	,000	1	,999	,000	,000	.
	OTÜR(4)	-25,559	25075,102	,000	1	,999	,000	,000	.
	YBSTAY	,016	,055	,085	1	,771	1,016	,913	1,131
	Constant	17,271	45148,061	,000	1	1,000	3,2E+07		

a. Variable(s) entered on step 1: TTÜR, YASKOD, K2, K3, K4, K5, AMLSÜRE, IMMOB, RD, UYKU, K10, K11, K12, DM, AFILIN, ARITM, K16, K17, KEMO, STEROID, ELEKTR, CINS, OTÜR, YBSTAY.

Tablo 2. de lojistik regresyon sonucu yaş, psikiyatri sorunu, operasyon süresi, serumdaki kimyasal değerlerin anormalliği risk faktörlerinin p değerleri 0.05 altında çıkmış yani bu risk faktörlerinin parametre tahminleri anlamlı kabul edilmiştir. Modelimizin Nagelkerke R^2 değeri 0.528 dir. Hosmer-Lemeshow goodness of fit testi sonucu p değeri 0,998 çıkmış yani oluşturduğumuz modelin uygun bir model olduğu sonucuna varılmıştır.

Propensity skor genellikle lojistik regresyon kullanılarak aşağıda ifade edilen yöntem ile tahmin edilmiştir ve ilk 10 delirium tanısı konmayan hasta için propensity skorlar şöyledir: 0.01644, 0.00339, 0.01951, 0.02112, 0.00000, 0.00335, 0.00635, 0.00000, 0.01095, 0.00768

$$\ln\left(\frac{PS}{1-PS}\right) = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p \quad PS = \frac{\exp(b_0 + b_1X_1 + \dots + b_pX_p)}{1 + \exp(b_0 + b_1X_1 + \dots + b_pX_p)}$$

Odds ratio = exp(constant + (coeff. x var1) + ... + (coeff. x varN))
PPSkor = Oddsratio / (1+Oddsratio)

Tüm delirium tanısı alan ve almayan hastalar için propensity skorlar tahmin edildikten sonra her iki grubun birlikte bu değerlerine bağlı olarak eşit örneklem sayısına sahip tabaka sınırlarına karar verilmiştir. Tablo. 3 bu çalışmamız için yapılan tabakaları ve sınırlarını göstermektedir.

Tablo 3. 5'li tabakalara ayrıldığında propensity skor düzeltilmesi sonrası kontrol ve tedavi gruplarına düşen örneklem sayısı

Tabakalar	Sınırlar	Delirium tanısı alan	Delirium tanısı almayan
1	0.5–0.115	7	28
2	0.115–0.05	2	33
3	0.05–0.022	1	34
4	0.022–0.0121	1	34
5	0.0121–0.006	1	34

Tabakalara ayırdıktan sonra çalışmamızda delirium tanısı alan 12 hasta ile almayan 163 hasta kalmış ve böylece birbiri ile benzer dağılıma sahip tedavi ve kontrol bireyleri seçilmiştir. Tabakalara ayırmadan önceki ve sonraki F istatistik sonuçları Tablo.4a ve Tablo.4b de verilmiştir.

Tablo 4a. Tabakalara ayırmadan önce ve sonra iki grup arasındaki dengesizliği ölçmek amacıyla hesaplanan F-istatistiği sonuçları

	Tabakalara ayırmadan önce F istatistiği	Tabakalara ayırdıktan sonra † F istatistiği
Ameliyat Öncesi		
Yaş (YASKOD) ((18–59=0) (59>=1))	8,55 **	0,021
Cinsiyet (CINS) (erkek=1,bayan=0)	0,267	0,043
Kronik Hastalık (K2) (var=1,yok=0)	0,606	-
Alkol Kullanımı (K3) (var=1,yok)	6,578**	0,055
Psikiyatri sorunu (K4) (var=1,yok=0)	1,687	0,662
Serebrovasküler hastalık (K5) (var=1,yok=0)	0,118	-
Diyabet (DM) (var=1,yok=0)	5,207**	0,349
Kemoterapi(KEMO) (var=1,yok=0)	0,648	-
Tümör türü (TTÜR) (var=1, yok=0)	0,044	0,021
Operasyon türü acil (OTÜR) (1–5)	0,184 ^a	0,802 ^a

† F istatistiği: $C^2 = F_{(r,\infty),r}$ (r= serbestlik derecesi); ^a F istatistiği= t^2 ; ** p<0,05

Tablo 4b. Tabakalara ayırmadan önce ve sonra iki grup arasındaki dengesizliği ölçmek amacıyla hesaplanan F-istatistiği sonuçları

	Tabakalara ayırmadan önce F istatistiği †	Tabakalara ayırdıktan sonra F istatistiği
Ameliyat Sonrası		
Solunum yetmezliği (RD) (var=1,yok=0)	0,980	1,284
Serumdaki kimyasal değerlerin anormalliği (ELEKTR) (var=1,yok=0)	13,982**	1,077
Operasyon süresi (AMLSÜRE) (var=1,yok=0)	8,357**	0,744
Yoğun bakım süresi (YBSTAY)	0,068 ^a	0,082
Uyku bozukluğu (UYKU) (var=1,yok=0)	51,32**	-
Hipertansiyon (K10) (var=1,yok=0)	1,418	0,197
Enfeksiyon (K11) (var=1,yok=0)	0,065	0,459
Kan transfüzyonu (K12) (var=1,yok=0)	0,451	0,058
Aminofilin kullanımı (AFİLİN) (var=1,yok=0)	0,282	0,936
antiaritmi kullanımı (ARİTMİ) (var=1,yok=0)	0,453	-
Antibiyotik kullanımı (K16) (var=1,yok=0)	0,027	0,705
Streoid kullanımı (STREÖİD) (var=1,yok=0)	0,936	0,002
Antihipertansif kullanımı (K17) (var=1,yok=0)	0,004	0,010
Hareket edememe (İMMOB) (var=1,yok=0)	0,769	0,315

† F istatistiği: $c^2 = F_{(r,\infty),r}$ (r= serbestlik derecesi); ^a F istatistiği= t^2 ; ** p<0,05

Yukarıdaki tablolarda görülebileceği gibi tabakalara ayırdıktan sonra birbirleri ile benzer ortalamalara sahip tedavi ve kontrol gruplarının elde edildiğine F istatistiği sonucunda karar verilmiştir. Propensity skora bağlı olarak, tabakalara ayırmadan önce yaş, alkol kullanımı, diyabet, serumdaki kimyasal değerlerin anormalliği, operasyon süresi ve uyku bozukluğu değişkenlerinde delirium tanısı alanlar ile almayanlar arasında farklılık olduğu görülmüş fakat tabakalara ayırma yapıldıktan sonra bu farkların düzeldiğine F istatistiği sonucunda karar verilmiştir. Kronik hastalık, serebrovasküler hastalık, uyku bozukluğu, antiaritmi kullanımı, kemoterapi değişkenlerinin var veya yok şeklindeki iki yanıtında, tabakalara ayırma yapıldıktan sonra yanıt değerlerinde sabit değer gözlenildiğinden bu değişkenlere istatistik analiz uygulanamamıştır. Elde ettiğimiz yeni örneklem ile lojistik regresyon yardımıyla risk faktörlerinin anlamlılıkları test edilmiş ve sonuçlar Tablo 5 de verilmiştir.

Tablo 5. Propensity skor düzeltilmesi sonrası lojistik regresyon çıktısı

		Variables in the Equation						95,0% C.I. for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1	TTUR(1)	,662	1,012	,428	1	,513	1,939	,267	14,084
	YASKOD(1)	-1,354	,959	1,993	1	,158	,258	,039	1,692
	K3(1)	-,387	,999	,150	1	,699	,679	,096	4,810
	K4(1)	-1,734	,958	3,278	1	,070	,177	,027	1,154
	AMLSÜRE(1)	-2,091	1,069	3,824	1	,051	,124	,015	1,005
	IMMOB(1)	,226	1,281	,031	1	,860	1,254	,102	15,444
	RD(1)	-,223	,790	,080	1	,778	,800	,170	3,763
	K10(1)	-,170	1,067	,025	1	,873	,844	,104	6,826
	K11(1)	-,164	,960	,029	1	,864	,849	,129	5,566
	K12(1)	,902	,826	1,193	1	,275	2,465	,489	12,436
	DM(1)	-1,046	,946	1,222	1	,269	,351	,055	2,244
	AFILIN(1)	-1,777	1,479	1,443	1	,230	,169	,009	3,072
	K16(1)	,224	,977	,053	1	,819	1,251	,184	8,490
	K17(1)	1,519	1,485	1,046	1	,306	4,570	,249	84,001
	STEROID(1)	-1,374	1,400	,963	1	,326	,253	,016	3,937
	ELEKTR(1)	-1,346	1,061	1,611	1	,204	,260	,033	2,080
	CINS(1)	,974	,794	1,505	1	,220	2,649	,559	12,565
	OTÜR			1,200	3	,753			
	OTÜR(1)	-15,931	40192,970	,000	1	1,000	,000	,000	.
	OTÜR(2)	1,190	1,113	1,143	1	,285	3,287	,371	29,119
OTÜR(3)	1,316	1,388	,899	1	,343	3,727	,246	56,582	
YBSTAY	,012	,054	,050	1	,824	1,012	,911	1,125	
Constant	1,874	3,488	,288	1	,591	6,511			

a. Variable(s) entered on step 1: TTÜR, YASKOD, K3, K4, AMLSÜRE, IMMOB, RD, K10, K11, K12, DM, AFILIN, K16, K17, STEROID, ELEKTR, CINS, OTÜR, YBSTAY.

Tablo.5 deki çıktıya göre hiçbir risk faktörünün parametre tahmini p değerleri 0,05 altında çıkmamış yani anlamlı kabul edilmemiştir. Modelimizin Nagelkerke R^2 değeri 0.159 dir. Hosmer-Lemeshow goodness of fit testi sonucu p değeri 0,273 çıkmış yani oluşturduğumuz modelin uygun bir model olduğu sonucuna varılmıştır.

4. Sonuç ve Tartışma

Sonuç olarak, propensity skor öncesi örneklem ile propensity skor sonrası örnekleme uyguladığımız analiz sonuçları risk faktörlerinin anlamlılığı bakımından oldukça farklıdır. Tabakalara ayırma yöntemi sonucunda birbiri ile ortak değişkenleri bakımından benzer dağılıma sahip tedavi ve kontrol bireyleri seçilmiş böylece yeni örneklemdeki tedavi ve kontrol grupları hemen hemen aynı karakteristiklere sahip olmuş ve böylelikle sistematik hata azalmıştır. Delirium tanısı az rastlanır bir durum olduğundan çalışmamızdaki vaka sayısının oldukça düşük olması çalışma sonuçlarını olumsuz yönde etkilemiştir. Bu nedenle de propensity skoru tahmin ettikten sonra eşleştirme yöntemi yerine daha büyük örneklem ile çalışmak için tabakalara ayırma yöntemi kullanılmıştır. Eğer vaka sayımız daha fazla olsaydı propensity skor öncesi ve sonrası yapılan istatistik analizler farklı sonuçlar verebilirdi. Bu durum çalışmamızın en önemli kısıtıdır.

Cochran (1968), ortak değişken veya değişkenin alt sınıflamasına bağlı olarak yapılan beş tabakaya ayırmanın biası (sistemik hatayı) % 90 olarak azalttığını göstermiştir [7]. Bu nedenle çalışmamız için propensity skoru beş tabakaya ayırma yöntemi tercih edilmiştir. Tabakalara ayırma sonucunda tam eşleştirme sağlanamamış, denek sayısı 478 den 175 düşmüştür çünkü tedavi ve kontrol grupları propensity skorlarının birbirini tutmayan değişim aralıklarına sahiptir [1]. Çalışmamızda delirium tanısı alan hastaların minimum ve maksimum propensity skorları 0,00687 ve 1,00, delirium tanısı almayan hastalar için ise minimum ve maksimum propensity skorlar 0,509 ve 0,00 dır. En yüksek propensity skora sahip delirium tanısı alanlar (>0,509) ve en düşük propensity skora sahip delirium tanısı almayanlar (<0,00687) çalışmadan çıkarılmıştır.

Propensity skorun tanımında bahsettiğimiz gibi Rosenbaum ve Rubin (1983) propensity skor hesaplanması ve sonrasında yapılan uygulamaları tamamlanmış veri seti için geliştirmiştir. Ancak günümüzde propensity skorun kayıp veriler için hesaplanması henüz hazır istatistik programlarında mevcut değildir. Kayıp verilere sahip bir veri seti için çalışmalar sürdürülmektedir [3]. Özellikle kısmi kayıp veri seti için D'Agostino'nun çalışmaları önem taşımaktadır [8][9].

Özellikle, propensity skor uygulayarak rassal olmayan çalışmalarla ve geriye dönük çalışmalarla elde edilen sonuçlar, hemen hemen ileriye dönük rassal çalışmalar gibi kuvvetli sonuçlar verir. Propensity skor, istatistik analizlerde geniş bir şekilde özellikle uygulamalı tıp alanında kullanılmaktadır. Rassal klinik denemelerin maliyeti arttıkça ve çoğu araştırmacının daha az pahalı olan araştırmalara yani gözleme dayalı çalışmalara yönelmesi sonucunda propensity skorun kullanımı artış göstermektedir. Propensity skor yöntemi, çalışmanın tasarım aşamalarına (eşleştirme veya tabakalara ayırma) katıldığında en iyi faydayı ürettiği görülmüştür. Bu fayda zaman ve paradan tasarruf ederek gerçek tedavi etkisinin daha kesin tahminlerini içerir. Bu tasarruf ile çalışma için uygun olmayacak bireylerin çalışmaya alınmasından kaçınılacaktır [3]. Çalışmanın en önemli tarafı ise propensity skor uygulanarak örnekleme deki sistemik hatanın azaltılmasıyla, (retrospective) rassal olmayan geriye dönük çalışmaların, ileriye dönük (prospective) rassal çalışmalar kadar yakın sonuçlar vermesidir [10][11].

Kaynaklar

1. Paul R. Rosenbaum (1995), *Observational Studies*, New York-Berlin-Heidelberg London Paris Tokyo Hong Kong Barcelona Budapest, Springer-Verlag.
2. Rosenbaum, P. R. and Rubin, D. B. 'The central role of the propensity score in observational studies for causal effects', *Biometrika*, 70, 41–55 (1983).

3. D'Agostino RB, Jr. 'Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group.' *Stat Med* 1998; 17: 2265–2281.
4. Rosenbaum, P. R. and Rubin, D. B. 'Constructing a control group using multivariate matched sampling methods that incorporate the propensity score', *American Statistician*, 39, 33–38 (1985).
5. Rubin, D. B. 'Matching methods that are equal percent bias reducing: some examples', *Biometrics*, 32, 109–120 (1976).
6. Rosenbaum, P. R. and Rubin, D. B. 'Reducing bias in observational studies using subclassification on the propensity score', *Journal of the American Statistical Association*, 79, 516–524 (1984).
7. Cochran, W. G. 'The effectiveness of adjustment by subclassification in removing bias in observational Studies', *Biometrics*, 24, 205–213 (1968).
8. D'Agostino, R. B. Jr. 'Estimating propensity scores when covariates have either ignorable or nonignorable missing values', Ph.D. thesis, Harvard University, 1994.
9. D'Agostino, R. B. Jr. and Rubin, D. B. 'Estimating and using propensity scores with partially missing data', submitted to *Journal of the American Statistical Association*, (1977).
10. Boening A., Friedrich C., Hedderich J., Schoettler J., Fraund S., Cremer J. T. ' Early and medium-term after on-pump and off-pump coronary artery surgery: a propensity score analysis', *Ann Thorac Surg* (2003) 76: 2000-2006.
11. Weitzen S., Lapane K.L., Toledano A. Y., Hume A. L., Mor V. 'Principles for modeling propensity scores in medical research: a systematic literature review' *pharmacoepidemiology and drug safety* (2004) 13: 841–853.

KORONER ARTER HASTALIĞININ TAHMİNİNDE SINIFLANDIRMA TEKNİKLERİNİN PERFORMANSLARININ KARŞILAŞTIRILMASI

İmran Kurt¹, Mevlüt Türe², A. Turhan Kürüm³

Osmangazi Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı¹

Trakya Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı²

Trakya Üniversitesi Tıp Fakültesi Kardiyoloji Anabilim Dalı³

e-mail: ikurt@trakya.edu.tr

Özet

Bu çalışmada, koroner arter hastalığının tahmin etmede sınıflandırma yöntemlerinin performanslarının karşılaştırılması amaçlandı. Geriye dönük araştırmanın verileri 1245 kişiden (865 hasta ve 380 kontrol) elde edildi. Bu veri setinde; lojistik regresyon analizi (LR), sınıflandırma ve regresyon ağacı (classification and regression tree, CART), çok tabakalı algılayıcı (multi-layer perceptron, MLP), radyal tabanlı fonksiyon (radial basis function, RBF) ve kendi kendini organize eden model (self-organizing feature maps, SOFM) yöntemlerinin sınıflandırma performansları ROC eğrisi yöntemi kullanılarak karşılaştırıldı. Yaş, cinsiyet, aile hikayesi, sigara içme, diyabet, sistemik hipertansiyon, hiperkolesterol ve vücut kitle indeksi (VKI) değişkenleri bağımsız değişkenler olarak alındı. ROC eğrisi altında kalan alanlar MLP, LR, CART, RBF ve SOFM yöntemleri için sırasıyla 0.783, 0.753, 0.745, 0.721 ve 0.675 bulundu. MLP, bu veri setinde KAH'ı en iyi tahmin eden yöntem olarak bulundu. LR ve CART ikinci en iyi yöntemler olarak bulunurken RBF üçüncü sırada yer aldı. SOFM'in tahmin performansı diğer modellerden daha düşük bulundu.

Anahtar Kelimeler: Lojistik Regresyon, Sınıflandırma ve Regresyon Ağaçları, Sinir Ağları, Koroner Arter Hastalığı.

1. GİRİŞ

Koroner arter hastalığı (KAH), dünyadaki en önemli sağlık problemlerinden biridir [1]. KAH ve miyokard infarktüsünün tanısında rol oynayan pek çok risk faktörü vardır. Özellikle aile hikayesi, sigara kullanımı,

hipertansiyon, hiperkolesterol, diyabet ve obezite, KAH'ın major risk faktörleri olarak tanımlanmaktadır [2-6]. KAH'ın risk faktörlerinin tanımlanması, tanı ve tedavi yönünden çok önemlidir.

Sınıflandırma tekniklerinin karşılaştırılmasına yönelik farklı alanlarda pek çok çalışma yapılmıştır. Moisen ve ark. [7] çalışmalarında orman yapısının haritasını çıkarmak için doğrusal modelleri, genelleştirilmiş toplamsal modelleri, sınıflandırma ve regresyon ağacı (CART), çok değişkenli toplamsal regresyon eğrileri ve Yapay Sinir Ağları (YSA)'nı karşılaştırmışlardır. Delen ve ark. [8] göğüs kanserli hastaların yaşayıp yaşamayacağını tahmin etmede lojistik regresyon analizi (LR), karar ağacı (C5) ve YSA'ları karşılaştırmışlardır. Stark ve Pfeiffer [9] veterinerlik epidemiyolojisiyle ilgili bir veri seti üzerinde LR, karar ağaçları (ID3, C4.5, CHAID, CART) ve YSA'yı karşılaştırmışlardır. Colombet ve ark. [10] kardiyovasküler hastalığın risk faktörlerini tahmin etmek için LR, CART ve YSA'nın performanslarını karşılaştırmışlardır. Ture ve ark. [11] hipertansiyonun tahmininde çok tabakalı algılayıcı (MLP), radyal tabanlı fonksiyon (RBF), LR, esnek ayırma analizi, çok değişkenli toplamsal regresyon eğrileri, CART, CHAID ve QUEST yöntemlerini sınıflandırma performansları bakımından karşılaştırmışlardır.

Bu çalışmada, KAH olan ve olmayan grupları ayıran en iyi modeli tahmin etmede LR, CART, MLP, RBF ve kendi kendini organize eden model (SOFM) yöntemlerinin sınıflandırma performanslarının karşılaştırılması amaçlandı.

2. GEREÇ ve YÖNTEMLER

2.1. Veri

Geriye dönük ve olgu-kontrol araştırması olarak gerçekleştirilen çalışmamız, Ocak 1999-Şubat 2003 tarihleri arasında Trakya Üniversitesi Tıp Fakültesi Kardiyoloji Polikliniğine gelen 1245 birimden 891 erkek (% 71.6) ve 354 kadın (% 28.4) oluşmaktadır. Anjiyografi sonucu koroner damarlarda herhangi bir lezyonun saptanmadığı 380 birim kontrol grubunu oluştururken, en az bir koroner arterde stenoz \geq % 50 saptanan 865 birim de hasta grubunu oluşturdu. Çalışmada bütün birimler yaş, cinsiyet, aile hikayesi, sigara kullanımı, diyabet, sistemik hipertansiyon, hiperkolesterol ve vücut kitle indeksi (VKİ) risk faktörleri bakımından incelendi.

Sınıflandırma modelleri meydana getirilmeden önce veri seti rasgele üç sete bölündü. Veri setinin % 60'ı (n=747) eğitim setine, % 20'si (n=249) test setine ve % 20'si (n=249) geçerlilik setine ayrıldı.

2.2. Sınıflandırma Yöntemleri

Lojistik Regresyon Analizi: LR, bağımlı değişkenin kategorik ve bağımsız değişkenlerin karışık ölçekli olması durumunda belirli bir dağılım varsayımına bağlı kalmaksızın bağımlı değişken ile bağımsız değişkenler arasındaki neden-sonuç ilişkisinin belirlenmesinde ve aynı zamanda bağımsız değişkenlerin etkilerine dayanarak verilerin sınıflandırılmasında kullanılan bir yöntemdir [12-14].

p bağımsız değişken için LR modeli

$$P(Y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}}$$
 olup $\beta_0, \beta_1, \dots, \beta_p$, regresyon katsayılarıdır [13, 14].

Sınıflandırma ve Regresyon Ağacı: CART, ikili karar ağacı oluşturan bir yöntemdir. Ağaçtaki her bir düğümde, her bir bağımsız değişken için gelişim skoruna dayalı olarak en iyi kesim noktası (sürekli değişkenler için) ya da en iyi kategori grupları (kategorik değişkenler için) oluşturulur. En iyi kestirici değişken seçimi, kategorik bağımlı değişkenler için Gini ya da twoing, sürekli bağımlı değişkenler için en küçük kareler sapması indeks hesaplamalarına göre yapılmaktadır. Burada amaç, bağımlı değişkenle ilgili verinin mümkün olduğunca homojen alt setlerinin meydana getirilmesidir [15]. Bu çalışmada, kategorik bağımlı değişkenler için kullanılan Gini indeksi kullanılmıştır.

Gini indeksi:

t düğümünde Gini indeksi aşağıdaki gibi tanımlanır,

$$g(t) = \sum_{j \neq i} p(j|t) p(i|t)$$

Burada i ve j bağımlı değişkenin kategorileridir.

Böylece, bir düğümdeki birimler çapraz kategorilere dağıldığında bile, Gini indeksi $1-(1/k)$ 'nin maksimum değerini alır. Burada k bağımlı değişkenin kategorilerinin sayısıdır. Düğümdeki bütün birimler aynı kategoride olduğunda, Gini indeksi 0'a eşittir.

t düğümünde s dalı için Gini kritik fonksiyonu

$$\Phi(s, t) = g(t) - p_L g(t_L) - p_R g(t_R)$$

eşitliğiyle tanımlanır. Burada p_L , soldaki düğüme gönderilen birimlerin oranı ve p_R , sağdaki düğüme gönderilen birimlerin oranıdır. s dalı, $\Phi(s,t)$ fonksiyonunun değerini maksimum yapmak için seçilir. Bu değer, ağaçtaki gelişim olarak ifade edilir [15].

Çok Tabakalı Algılayıcı: MLP, geriye yayılma algoritmasıyla çalışan ileri beslemeli bir YSA'dır. MLP, sinir ağı tarafından elde edilen çıktıyı gözlenen değerlerle karşılaştırarak en iyi ağı belirlemeye çalıştığı için danışmanlı öğrenme stratejisiyle çalışır. MLP ağı, girdi ve çıktı tabakasının yanında bir ya da daha fazla gizli tabakadan oluşabilir. Gizli tabakadaki nöronlar, girdi vektörü ile çıktı vektörleri arasındaki doğrusal ve doğrusal olmayan ilişkilerin açıklanması için gerekli bilginin kayba uğramadan taşınmasına izin verir. MLP, özellikle sınıflandırma problemlerinde yaygın olarak kullanılan en popüler YSA'dır [11, 16, 17].

MLP, gözlenen ve beklenen çıktı arasındaki farktan kaynaklanan hatayı en küçük yapan en iyi ağırlıkları belirlemek için geriye yayılma algoritmasını kullanır. Bu algoritma, ağ hatasını en küçük yapmak için ağırlıklarının adım adım hesaplanması sürecidir.

Geriye yayılma algoritmasının t 'inci adımdan $(t+1)$ 'inci adıma geçişteki en genel ifadesi,

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_i(t) x_j(t)$$

eşitliğiyle gösterilir. Burada η , öğrenme oranı parametresi ve $\delta_i(t)$, t 'inci adımda i nöronunun yerel hatasıdır [16-18].

Radyal Tabanlı Fonksiyon: RBF, MLP'ye alternatif olarak geliştirilen ileri beslemeli bir YSA algoritmasıdır. RBF, ileri beslemeli YSA yapılarına benzer şekilde girdi, gizli ve çıktı tabakasından oluşur. Ancak girdi tabakasından gizli tabakaya geçişteki dönüşüm, radyal tabanlı aktivasyon fonksiyonlarını kullanan doğrusal olmayan bir dönüşümdür.

RBF, öğrenme sürecinde danışmansız ve danışmanlı öğrenme olmak üzere karma strateji kullanır ve böylece öğrenme süreci iki aşamaya ayrılır. Danışmansız öğrenme stratejisi ile gerçekleştirilen ilk aşama, girdi verilerine gizli tabakada radyal tabanlı fonksiyon uygulanması sürecidir. Danışmanlı öğrenme stratejisi ile gerçekleştirilen ikinci aşama ise, gizli tabakadan elde edilen çıktıların doğrusal dönüşüm uygulanarak ağ çıktıları ile gözlenen çıktılar arasındaki hatayı en küçük yapan ağırlıkların belirlenmesi sürecidir [11,16,17].

RBF'de gizli tabakadaki her bir nöron, ortalaması μ_i ve varyansı σ_i olan ve $G_i(x) = \exp\left[-\frac{1}{2\sigma_i^2} \sum_{k=1}^p (x_k - \mu_{ik})^2\right]$ eşitliğiyle gösterilen çok

değişkenli Gauss fonksiyonunu kullanır. Burada p bağımsız değişken sayısı ve i gizli tabakada yer alan herhangi bir nöronu gösterir. Bu fonksiyonlar, radyal tabanlı fonksiyonlar olarak isimlendirilir. Sonuç olarak, çıktı tabakasında doğrusal dönüşüm uygulanarak elde edilen ağ çıktıları

$$f(x) = \sum_{i=1}^n w_{ji} (G_i(x))$$

eşitliğiyle elde edilir [16,17].

Kendi Kendini Organize Eden Model: SOFM, özellikle sınıflandırma problemlerinde kullanılan geri beslemeli bir yapay sinir ağıdır. Bu tür ağlarda ileri beslemeli ağların aksine, tabakalar arasındaki bağlantıya ilave olarak tabakadaki her bir nöron da birbirleriyle bağlantılıdır ve sinir ağına geri besleme bağlantıları vardır.

SOFM, öğrenme sürecinde danışmansız öğrenme stratejisini kullanan bir sinir ağıdır. Kohonen tarafından geliştirilen bu ağlar, rekabete dayalı öğrenme ile girdi verisinin yapısına göre kendilerini organize ederek nöronların ağırlıklarını ayarlarlar. En uygun çıktıya sahip nöron diğerlerine göre baskın olur ve sadece bu nöronun çıktı sağlaması söz konusudur [16, 17].

SOFM ağlarının çıktısı, MLP gibi danışmanlı öğrenme stratejisi ile çalışan sinir ağlarına girdi olarak kullanılabilir. Bu durumda, SOFM kullanılarak boyutu indirgenen girdi uzayı verileriyle sınıflandırma yapılır. Böylece, girdi uzayının boyutu indirgenirken, girdi uzayından elde edilen bilgi kayba uğramamış olur.

SOFM ağları genel olarak; girdi tabakası ve 2 boyutlu bir düzlemi gösteren danışmansız tabakadan oluşur. Öğrenme süreci, rasgele olarak belirlenen küçük ağırlıklarla (w_j) başlar. Her bir girdi verisi için SOFM, t'inci adımda Öklit uzaklığı kriterini kullanarak rekabet sonucunda kazanan j^* nöronunu bulur: $j^* = \operatorname{argmin}_j \|x(t) - w_j\|$ ($j=1, \dots, N^2$). Burada $x(t) = [x_1(t), \dots, x_n(t)]$, t'inci adımda girdi vektörü ve N^2 , danışmansız tabakadaki toplam nöron sayısıdır. Öklit uzaklığı, girdi vektörüne en yakın

ağırlık vektörüne sahip nöronu bulmak için kullanılır. SOFM ağında, rekabeti kazanan ve onun etrafındaki komşu nöronların ağırlıkları t'inci adımdan (t+1)'inci adıma geçişte,

$$w_j(t+1) = w_j(t) + \eta(t)\Lambda_{j,j^*}(t)[x(t) - w_j(t)]$$

eşitliğiyle güncelleştirilir. Burada $\eta(t)$ öğrenme oranı parametresi ve $\Lambda_{j,j^*}(t)$ j ve j* nöronlarının komşuluklarını belirleyen komşuluk fonksiyonudur. SOFM, $\eta(t)$ ve $\Lambda_{j,j^*}(t)$ 'yi en iyi sonuçlar elde edilene kadar değiştirir [16,17].

3. Bulgular

Bağımsız değişkenlere ilişkin tanımlayıcı istatistikler Tablo 1'de verilmiştir. Yaş değişkeninin normal dağılıma uygunluğu Shapiro-Wilk testi ile incelendi ve normal dağılıma uygun olmadığı için gruplar arasında fark olup olmadığı Mann-Whitney U testi ile analiz edildi. İsimsel değişkenler ise Ki-kare testi ile analiz edildi.

Tablo 1. Bağımsız değişkenlerin gruplara göre tanımlayıcı istatistikleri

Bağımsız değişkenler	KAH		p
	KAH+ (n=865)	KAH- (n=380)	
Cinsiyet (erkek/kadın)	4.1	1.1	<0.001
Yaş (yıl) medyan (25%-75% persantil)	60 (52-67)	55 (45-64)	<0.001
VKİ (%)			0.024
Zayıf	0.8	2.6	
Normal	30.9	27.6	
Şişman	68.3	69.7	
Sigara kullanımı (%)	57.9	42.4	<0.001
Diyabet (%)	23.6	9.2	<0.001
Sistemik hipertansiyon (%)	47.1	41.6	0.074
Hiperkolesterol (%)	44.7	41.3	0.262
Aile hikayesi (%)	25.2	21.3	0.139

Cinsiyet, yaş, VKİ, sigara kullanımı ve diyabet bakımından KAH olan ve olmayan gruplar arasında istatistiksel olarak anlamlı fark olduğu bulundu (sırasıyla p<0.001, p<0.001, p<0.05, p<0.001 ve p<0.001).

Sınıflandırma Yöntemlerinin Karşılaştırılması:

Eğitim seti ve test seti için sınıflandırma yöntemlerinin duyarlılık, özgüllük ve doğruluk oranları Tablo 2'de verilmiştir. Bütün modellerin eğitim seti için duyarlılık oranları % 91.6-% 97.9, özgüllük oranları % 9.9-% 46.4 ve doğruluk oranları % 70.4-% 78.7 aralığında değer alırken test seti için duyarlılık oranları % 89.5-% 98.9, özgüllük oranları % 7.4-% 47.1 ve doğruluk oranları % 73.9-% 79.9 aralığında değer aldı.

Tablo 2. Eğitim ve test setleri için modellerin duyarlılık, özgüllük ve doğruluk oranları

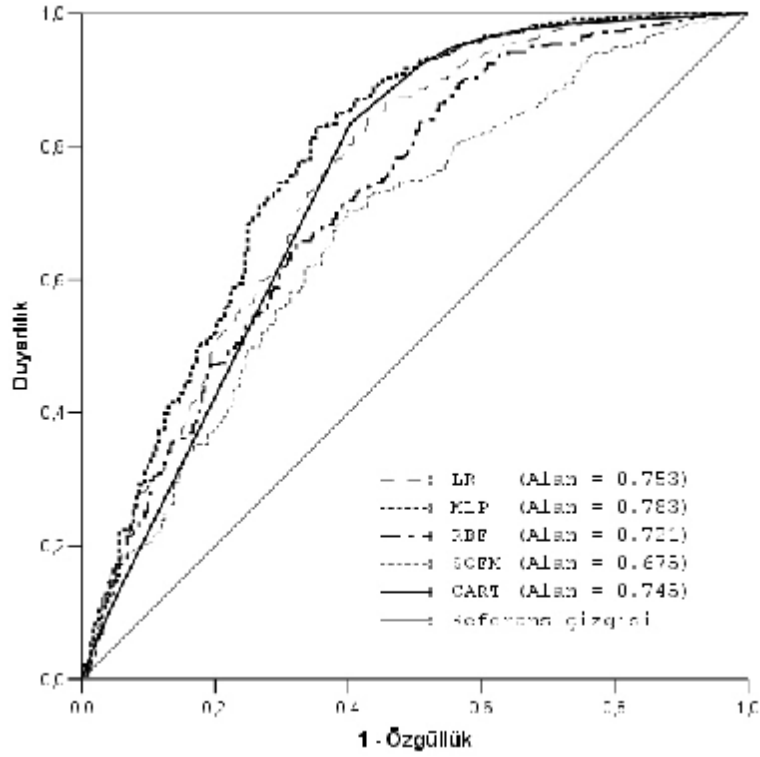
Model	Eğitim seti			Test seti		
	Duyarlılık (%)	Özgüllük (%)	Doğruluk (%)	Duyarlılık (%)	Özgüllük (%)	Doğruluk (%)
LR	91.8	43.3	76.7	92.3	45.6	79.5
CART	94.9	42.9	78.7	92.3	47.1	79.9
MLP	93.4	46.4	78.7	91.7	45.6	79.1
RBF	91.6	39.1	75.2	89.5	42.6	76.7
SOFM	97.9	9.9	70.4	98.9	7.4	73.9

LR, CART, MLP, RBF ve SOFM yöntemlerinin ROC eğrisi altında kalan alanları 0.675-0.783 aralığında değer aldı (Şekil 1). ROC eğrisi altında kalan alanlar ise Hanley ve McNeil [19] tarafından geliştirilen yöntemle karşılaştırıldı (Tablo 3). LR ile MLP, LR ile RBF, LR ile SOFM yöntemleri arasında ROC eğrisi altında kalan alanlar bakımından anlamlı fark olduğu bulundu (sırasıyla $p=0.001$, $p=0.015$ ve $p<0.001$). CART ile MLP, CART ile SOFM arasında ROC eğrisi altında kalan alanlar bakımından anlamlı fark olduğu bulundu (sırasıyla $p=0.026$ ve $p=0.003$). MLP ile RBF, MLP ile SOFM ve RBF ile SOFM arasında ROC eğrisi altında kalan alanlar bakımından anlamlı fark olduğu bulundu (sırasıyla $p<0.001$, $p<0.001$ ve $p=0.013$). Bununla birlikte LR ile CART ve CART ile RBF arasında ROC eğrisi altında kalan alanlar bakımından anlamlı fark olmadığı bulundu (sırasıyla $p=0.663$ ve $p=0.250$).

Şekil 1'de de görüleceği gibi KAH olan ve olmayan grupları ayırmada performansı en iyi olan yöntem MLP olarak bulundu. Bu yöntemi LR ve CART izlerken üçüncü iyi yöntem ise RBF olarak bulundu. SOFM ise diğer modellerden daha düşük performans göstererek grupları ayırmada en yetersiz yöntem olarak bulundu.

Tablo 3. ROC eğrisi altında kalan alanlar arasındaki fark ($Alan_i$), standart hatalar (SH) ve yöntemlerin performansları

	LR	CART	MLP	RBF
CART	$Alan_i = 0.008$ $SH(Alan_i) = 0.019$ p = 0.663			
MLP	$Alan_i = 0.030$ $SH(Alan_i) = 0.009$ p = 0.001	$Alan_i = 0.038$ $SH(Alan_i) = 0.017$ p = 0.026		
RBF	$Alan_i = 0.032$ $SH(Alan_i) = 0.013$ p = 0.015	$Alan_i = 0.024$ $SH(Alan_i) = 0.021$ p = 0.250	$Alan_i = 0.062$ $SH(Alan_i) = 0.014$ p < 0.001	
SOFM	$Alan_i = 0.078$ $SH(Alan_i) = 0.021$ p < 0.001	$Alan_i = 0.070$ $SH(Alan_i) = 0.023$ p = 0.003	$Alan_i = 0.109$ $SH(Alan_i) = 0.020$ p < 0.001	$Alan_i = 0.046$ $SH(Alan_i) = 0.019$ p = 0.013

**Şekil 1.** Sınıflandırma yöntemlerinin ROC eğrileri

4. Tartışma

Kardiyovasküler hastalık, dünyada olduğu gibi Türkiye'de de ölümlerin en önemli nedenlerinden biridir [20]. Kardiyovasküler hastalık, genel olarak birkaç risk faktörünün birleşmesiyle ortaya çıkmaktadır.

Bu çalışmada, KAH'ı tahmin etmek için LR, CART ve üç YSA algoritmasının karşılaştırılması amaçlandı. Tahmin modellerinin performansları duyarlılık, özgüllük, doğruluk oranları ve ROC eğrilerine göre değerlendirildi. ROC eğrileri altında kalan alanlar, Hanley ve McNeil [19] tarafından geliştirilen yöntemle karşılaştırıldı.

Önceki çalışmalar incelendiğinde, farklı alanlardaki veri setlerinde en iyi sınıflandırma performansına sahip olan yöntemler farklılık göstermektedir. Moisen ve Frescino [7], genelleştirilmiş toplamsal modeller ve çok değişkenli toplamsal regresyon eğrileri yöntemlerinin orman karakteristiklerini tahmin etmede YSA, CART ve doğrusal modellerden daha iyi olduğunu bildirmişlerdir. Delen ve ark. [8], karar ağacı yöntemlerinden C5'in göğüs kanseri tanısı alan birimlerin yaşayabilirliğini tahmin etmede LR ve YSA'dan daha iyi olduğunu bildirmişlerdir. Stark ve Pfeiffer [9], veterinerlik epidemiyolojisi ile ilgili bir veri setinin açıklayıcı veri analizi için ID3, C4.5, CHAID ve CART yöntemlerinin daha uygun olduğunu bildirmişlerdir. Colombet ve ark. [10], kardiyovasküler hastalığın risk faktörlerini tahmin etmede CART'ın tahmin performansının LR ve YSA'dan biraz daha düşük olduğunu bildirmişlerdir. Ture ve ark. [11] hipertansiyonun tahmininde YSA yöntemlerinden MLP ve RBF'nin sınıflandırma performanslarının istatistiksel algoritmalar ve karar ağacı yöntemlerinden daha iyi olduğunu bildirmişlerdir. Ture ve ark. çalışmalarına benzer olarak çalışmamızda, MLP'nin tahmin performansının diğer yöntemlerden daha iyi olduğu ve bunun aksine RBF'nin çok iyi bir performans sergilemediği görüldü. Bu sonuçlar yöntemlerin veriden veriye değişik performanslar gösterdiğini ortaya koymaktadır.

CART ve LR, yöntemleri sağlık alanıyla ilgili veri setlerindeki gizli, karmaşık yapıları ve ilişkileri ortaya çıkarmada yetenekli olmalarına rağmen çalışmamızda eğitim ve test setinde MLP'den daha düşük performans gösterdiler. Çalışmamızdaki veri setinde MLP'nin KAH'ı tahmin etmek için daha iyi bir sınıflandırma yöntemi olduğuna karar verildi. Bununla birlikte, yaş, cinsiyet, aile hikayesi, sigara kullanımı, sistemik hipertansiyon, diyabet, hiperkolesterol ve VKİ'nin KAH'ın tahmininde kullanılabilir değişkenler olduğu sonucuna varıldı. Bütün modellerin duyarlılık oranı yüksek iken, çalışmamıza fiziksel aktivite, lipoprotein (a), hiperurisemi ve homosistein gibi değişkenlerin dahil edilmemesinden kaynaklanan bazı kısıtlılıkları nedeniyle modellerin özgüllük oranları düşük bulundu.

Her yöntem, klinik uygulamayla ilgili bazı karakteristikleri gösterir. Birincisi, CART'ın ağaç gösterimi, medikal düşünceye yakındır ve tahmin sonuçlarının daha iyi bir şekilde anlaşılmasına yardımcı olabilir. İkincisi, mevcut veriyle iterasyonlarla öğrenme süreci devam ettirildiğinde MLP ile bulunan modelde değişebilir. Bu yöntemler, bilgisayar destekli olarak gerçekleştirilen karar destek sistemlerindeki riskin yorumu ve gösterimine yardım edebilir ve mevcut istatistiksel modellere tamamlayıcı olabilirler. Bu modeller, KAH'ın önlenmesi ve tedavisi için klinik kılavuzların ve sağlık politikalarının tasarlanması konusunda bilgi veren kapsamlı analitik bir yapı sağlar.

Çalışmamızda gerçek bir veri seti kullanarak yöntemleri karşılaştırmamızdaki amaç, sağlık alanından elde edilen veri setlerindeki veri yapısının genel eğilimi hakkında bilgi edinmek ve bundan sonraki çalışmalarda sağlık alanında karşılaşılan sınıflandırma problemleri için en iyi yöntemin seçiminde araştırmacılara ışık tutmaktır. Geçmişte tek bir yöntemle bağlı kalınarak yapılan sınıflandırma çalışmalarının sonuçlarının yeterliliği konusunda yeterince bilgi yoktur ve veriyi daha iyi açıklayabilecek performansı daha iyi olan yöntemler araştırılmamıştır. Bu bilgiler ışığında araştırmacılara, gelecekte yapılacak KAH ile ilgili olan ya da olmayan çalışmalarında tek bir yöntemle bağlı kalarak verinin değerlendirilmesinden kaçınılması ve veriyi daha iyi açıklayarak en iyi performansı veren yöntemlerle modelleme çalışmaları yapılması gerektiğini önermekteyiz.

Kaynaklar

1. Backer GD, Ambrosioni E, Borch-Johnsen K, et al. European guidelines on cardiovascular disease prevention in clinical practice. Third Joint Task Force of European and other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of eight societies and by invited experts). *European Heart Journal* 2003, 24: 1601-1610.
2. Chobanian AV, Bakris GL, Black HR, et al. National Heart, Lung and Blood Institute Joint National Committee on Prevention, detection, Evaluation, and Treatment of High Blood Pressure; National High Blood Pressure Education Program Coordinating Committee. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood pressure: the JNC 7 report. *JAMA* 2003; 289: 2560-2572.
3. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults, Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on

- Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). JAMA 2001; 285(19): 2486-2497.
4. Herrington DM, Kesler K, Reiber JC, et al. Arterial compliance adds to conventional risk factors for prediction of angiographic coronary artery disease. Am Heart J 2003; 146: 662-667.
 5. Rahman MA, Chaudhury HS, Malek MA, Khaled MA. Risk factors of angiographically defined coronary artery disease in Bangladesh. Cardiovascular Pathology 2004; 13: S80-S138.
 6. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus, Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 2003; 26: S5-20.
 7. Moisen GG, Frescino TS. Comparing five modelling techniques for predicting forests characteristics. Ecological Modelling 2002; 157: 209-225.
 8. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. Artificial Intelligence in Medicine (in press).
 9. Stark KDC, Pfeiffer DU. The application of non-parametric techniques to solve classification problems in complex data sets in veterinary epidemiology-an example. Intelligent Data Analysis 1999; 3: 23-35.
 10. Colombet I, Ruelland A, Chatellier G, et al. Models to predict cardiovascular risk: comparison of CART, multilayer perceptron and logistic regression. Proc AMIA Symp 2000; 156-160.
 11. Ture M, Kurt I, Kurum AT, Ozdamar K. Comparing classification techniques for predicting essential hypertension. Expert Systems with Applications (in press).
 12. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: John Wiley & Sons; 2000.
 13. Ozdamar K. Paket Programlarla İstatistiksel Veri Analizi-1. Eskişehir: Kaan Kitabevi; 2004.
 14. Sharma S. Applied Multivariate Techniques. New York: John Wiley & Sons; 1996.
 15. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Monterey: Wadsworth and Brooks/Cole; 1984.
 16. Haykin S. Neural Network: A Comprehensive Foundation. Upper Saddle River, NJ: Prentice Hall; 1999.
 17. Principe J, Euliano NR, Lefebvre WC. Neural and Adaptive Systems: Fundamentals Through Simulations. New York: John Wiley & Sons Inc; 1999.

18. Kurt İ. Yapay Sinir Ağları ile Lojistik Regresyon Analizi'nin Karşılaştırılması (Yüksek Lisans Tezi). Edirne: Trakya Üniversitesi; 2004.
19. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; 148(3): 839-843.
20. Turkish Heart Report. Turkish Society of Cardiology. Yenilik Publication 2000; 11-25.

NEURAL NETWORKS İLE LOJİSTİK REGRESYON YÖNTEMLERİNİN EKSİK ÖLÇÜMLÜ VEYA UÇ DEĞERLER İÇEREN VERİ SETLERİ ÜZERİNDE KARŞILAŞTIRILMASI

İlker Ünal, Refik Burgut, Yaşar Sertdemir

Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana,
TÜRKİYE

Giriş

Sağlık alanında sınıflama amaçlı yapılan çalışmalarda başvuru yöntemleri arasında Lojistik Regresyon modelleri ve Neural Networks yaklaşımı bulunmaktadır. Bu yöntemler birçok çalışmada performansları yönünden karşılaştırılmış ve üstünlükleri sergilenmiştir. Bu çalışmada ise eksik ölçümlü ve uçdeğerli veri setlerinin yöntemlerin performansına olan etkisi incelenmiştir.

Anahtar Sözcükler: Eksik ölçümlü değerler, Uçdeğerler, Lojistik Regresyon, Neural Networks, ROC Eğrileri.

Gereç ve Yöntem

Çalışmada Lojistik Regresyon (LR) ve Neural Networks (NNs) yöntemleri üretilen eksik ölçümlü ve uçdeğerli veri setleri kullanılarak karşılaştırılmıştır. Veri üretiminde koşullu olasılıklar kullanılmıştır. Eksik ölçümlü veri setleri yüzde 10, 20 ve 30 oranlarında eksik ölçüm içerecek şekilde seçilmiştir. Uç değer içeren veri setlerinde ise yüzde 5 ve 10 oranında uç değerler bulunmaktadır. Örneklem büyüklüğü yöntemlerin performansını etkilediğinden veri setleri küçük (200 vakalık) ve büyük (1000 vakalık) olmak üzere iki kere oluşturulmuştur. Oluşturulan veri setleri yöntem performanslarını doğru ölçebilmek için iki alt örnekleme (eğitim ve test) bölünmüşlerdir. Yöntemlerin karşılaştırılmasında, elde edilen Receiver Operating Characteristic (ROC) eğrileri altında kalan alanları kullanılmıştır. Veri oluşturma işlemi ve analiz kısmı alan değerleri için güven aralığı elde edilebilmesi için 100 kere tekrarlanmıştır.

Bulgular

Eksik ölçümlü büyük veri setlerinin tümünde Neural Networks yaklaşımı ile daha iyi sonuçlar elde edilmiştir. Buna göre % 10 eksik ölçümlü

olan test veri setinde elde edilen ortalama alan deęerleri ve standart sapmaları NNs için 0.792 ve 0.014, LR için 0.759 ve 0.015'dir. Ancak 200 vakalık küçük veri setinde LR ile daha iyi alan deęerleri elde edilmiştir, LR için 0.755 ve 0.034, NNs için 0.720 ve 0.036'dır. % 20 ve 30 oranları için de bu deęerler benzerdir. % 5 oranında uç deęerler içeren büyük veri setinde ise bu deęerler NNs için 0.786 ve 0.014, LR için 0.755 ve 0.015 olarak bulunmuştur. Bu deęerler küçük veri setinde eksik ölçümlü veri setinde olduğu gibi LR yöntemi lehine daha üstün bulunmuştur, LR için 0.749 ve 0.034, NNs için 0.716 ve 0.036. % 10 oranında uç deęerli veri setinde de benzer sonuçlar elde edilmiştir.

Sonuç

Bu çalışma, eksik ölçümlü vakalar veya uç deęerler içeren büyük veri setlerinin analizinde Neural Networks yaklaşımının, veri setinin küçük olması durumunda ise Lojistik Regresyon yaklaşımının daha uygun olacağını göstermektedir.

YAPAY SİNİR AĞI İLE GRUPLANMIŞ ZAMANLI VERİLERİN ANALİZİ

Bahar Taşdelen¹, Sema Erden¹, Hakan Kaleağası², Aynur Özge²

Mersin Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, MERSİN¹

Mersin Üniversitesi Tıp Fakültesi Nöroloji Anabilim Dalı, MERSİN²

e-mail: bahartasdelen@mersin.edu.tr

Özet

Yapay sinir ağları, verilerin dağılımı konusunda kısıtlayıcı olmaması, karmaşık yapıdaki prognostik faktörün bir arada incelenebilmesi ve bu faktörlerin doğrusal olmayan etkilerinin modellenmesi nedeniyle sensörlü (censored) zaman verileri için yaygın olarak kullanılmaktadır. Yapay sinir ağları lojistik aktivasyon fonksiyonu kullanılarak modellenmiş ise, lojistik regresyon modelinin doğrusal olmayan genelleştirilmiş hali olarak düşünülebilir. Bu nedenle son yıllarda, gruplandırılmış zaman verilerinin analizi için kullanılmaya başlanmıştır. Her ne kadar bazı literatürler, yapay sinir ağlarının değişkenlerin önem kontrolünü yapmadığını ileri sürse de, duyarlılık analizi ile bu kontrol farklı bir şekilde yapılmaktadır.

Bu çalışmada, baş ağrısı şikayeti ile Nöroloji polikliniğine başvuran 65 yaş üstü hastalara ait yaş, cinsiyet, meslek, baş ağrısı tipi (birinci derece, ikinci derece), sigara ve alkol kullanımı, baş ağrısı süresi, (1, 2, 3, 4) ve baş ağrısı şiddeti (1, 2, 3, 4) gibi değişkenler kullanılarak hastaların iyileşme durumları (tam iyileşme, kısmi iyileşme, aynı kalma, kötüleşme) tahmin edilmeye çalışılmıştır. Tam iyileşme sensorlenmiş ve iyileşme durumu baş ağrısı süresine göre gruplandırılmıştır. Veri setine çok tabakalı perseptron (MLP) ağı uygulandıktan sonra, modele etkilerinin önem derecesine göre değişkenler; 1. baş ağrısı şiddeti, 2. meslek, 3. alkol kullanımı, 4. sigara içimi şeklinde sıralanmıştır. Bu değişkenler kullanılarak her grupta doğru sınıflandırma olasılıkları % 90-% 98,4 arasında bulunmuştur.

Anahtar Kelimeler: Yapay Sinir Ağları, survival analizi, Cox regresyon.

RANDOMİZE KONTROLLÜ ARAŞTIRMALARDA BİLİMSEL KALİTENİN ÖLÇÜMÜ İÇİN YAKLAŞIMLAR: 4 FARKLI YÖNTEMİN İRDELENMESİ

Mustafa Şenocak, Necdet Süt, Ömer Uysal, Penbe Çağatay

İstanbul Üniversitesi, Cerrahpaşa Tıp Fakültesi, Biyoistatistik BD.

e-mail: msukrus@yahoo.com,

Özet

Randomize kontrollü araştırma (RKA) – Rassal denetimli çalışma - yaklaşımı, girişimli klinik çalışmalar içinde, genellikle en güvenilir sonuç üreten araştırma yöntemi olarak kabul edilir. RKA'lardan, özellikle tedaviye yönelik “Karar verme” durumlarında, üst düzey güvenilirlikte kanıt olarak yararlanıldığı için bu tip çalışmalar en çok başvurulan kaynakların da başında gelmektedir. Bu kararların doğru olabilmesinde ise araştırmanın bilimsel “Kalite” kavramı öne çıkmaktadır. Bu kavram için, araştırmanın hem yöntem bilimi, hem süreci hem de sonuçlarının güvenilirliği açısından; İçsel geçerlilik, Dışsal geçerlilik ve İstatistiksel yargılama boyutlarının üçünü de içeren yaklaşımların gerekliliği öngörülmüştür. Böylece kalite kavramının içeriği ve ölçülmesi ile ilgili pek çok öneri yöntem de ortaya çıkmış ve kısa sürede sayıları ve amaçları da artmıştır. Yaklaşımlar, öngörülen çeşitli başlık, madde, alt-madde özelliklerine uygunluğa göre belli puanlar alınarak ortaya çıkan ölçeksel değerlendirmeler (skala) veya yine öngörülmüş olan özellikler listesinden hangilerinin var olduğu denetlenerek yapılan gözden geçirmelerdir (onay listesi). Ölçeksel yaklaşımların puan değişim aralığı birbirlerinden çok farklı olabilmektedir ve çeşitli nedenlerle yayın kalitesi değerlendirmesinde kullanılabilirler. Onay listeleri ise hem yayın editörlerinin, yollanan çalışmaları denetiminde yardımcı olmaktadır hem de RKA tasarlayan, sürdüren, yayına hazırlayan araştırmacılar için bir çeşit kılavuzluk oluşturabilmektedirler. Bu yaklaşımların içeriklerinden yararlanarak sadece RKA lar için değil, genel anlamda kıyasal yargılamalara dayalı tüm araştırmaların bilimsel düzeyini ölçümleyebilecek yaklaşımlara da gidilebilmektedir.

Bu derlememizde, günümüzde en sık ve güvenilerek kullanılan 2 onay listesini (*Delphi* ve *CONSORT listeleri*) ve 2 ölçeği (*Jadad* ve *PEDro* “skalaları”) gözden geçirdik. Amacımız, uygun araştırma tiplerinde, uygun yapılandırma ve değerlendirme koşullarının kullanılması, gereğinde ölçek ve

listelerin içerik ve yapılarında değişiklik yapılabilmesi için gerekli alt yapısal bilgiyi gözden geçirmektir.

Sonuçta, ele aldığımız yöntemlerin değerlendirmesinden anlaşılacağı üzere, bir kalite değerlendirme yönetiminde hem büyük okuyucu-uygulayıcı kitlelerine ulaşabilecek yayınlardan, belli bilimsel düzeyin üzerindeki seçebilmek hem de araştırmacılara kılavuzluk etmek bekleniyorsa yöntem daha kapsamlı ve bazen karmaşık duruma gelebilmektedir. İrdelediğimiz yöntemlerde bu açılardan çeşitli çelişkili, karmaşık, eksik ve/veya fazla yönler gözlenebilmektedir. RKA gerçekleştirmek amacı ile yola çıkmış bir çalışmanın *sine qua non* yönlerinin belirlenerek, bunların yeniden ölçüt olarak alınmaması madde sayısını azaltmak yönünde bir adım olabilir. Buna karşılık değerlendirme listesinin aynı zamanda kapsamlı bir kılavuz rolü üstlenmesi de isteniyorsa, daha temel öğeleri de içeren, tanımları açık maddeler listede bulunmalıdır.

Kalite değerlendirmesi yönünde yeni içerik maddelerinin, puansal ağırlıkların, sadece belli alan veya çalışma tiplerine yönelik değerlendirmelerin, yargı sınırlarının belirlenmesi gereği saptanmaktadır.

Anahtar Kelimeler: Randomize kontrollü araştırma (RKA), Yöntembilim, skala, onaylistesi, *Delphi* listesi, *CONSORT listesi*, *Jadad* ölçeği, *PEDro* ölçeği.

Giriş

Randomize kontrollü araştırma (RKA) – Rassal denetimli çalışma - yaklaşımı, girişimli klinik çalışmalar içinde, genellikle en güvenilir sonuç üreten araştırma yöntemi olarak kabul edilir. Yaklaşımın temeli, seçim biaslarını önlemek amacı ile, araştırmaya alınan olguların, değerlendirme kümelerine tamamen rastgele dağılımını sağlamaktır. Bununla, sadece kümelerin sayısal dengesi değil aynı zamanda konu-sonucu etkileyebilecek temel özelliklerin de başlangıçta benzer dağılması amaçlanır.

Klinik uygulayıcılar, RKA'lardan, özellikle tedaviye yönelik “Karar verme” durumlarında, üst düzey güvenilirlikte kanıt olarak yararlandıkları için bu tip çalışmalar en çok başvurulan kaynakların da başında gelmektedir. (1, 2). Bu kararların doğru olabilmesinde ise araştırmanın “Kalite” kavramı öne çıkmaktadır. Bu kapsamdaki bir kalite kavramı olarak öncelikle, sistematik hata içermeyen – unbiased – sonuçlar üretmek açısından, olabilirlik yeteneği tanımı önerilmiştir (3), ancak bu yaklaşımın sadece içsel geçerliliği kapsayabileceği, buna karşılık sonuçların güvenilirliği açısından; İçsel geçerlilik, Dışsal geçerlilik ve İstatistiksel yargılama boyutlarının üçünü de içeren yaklaşımların gerekliliği öngörülmüştür (1). Bu bağlamda farklı araştırmacıların kalite açısından , “körleme”, “tedavisi kararlaştırılmış olgular –

intention to treat- değerlendirmesi” gibi özgün noktaların önemini kanıtlamaya uğraşmalarının yanı sıra (4, 5), genel olarak kavramın içerik ve sınırları üzerinde de tartışılmıştır (6). Böylece kalite kavramının içeriği ve ölçülmesi ile ilgili pek çok öneri yöntem de ortaya çıkmış, Chalmers ve arkadaşlarının konu ile ilgili belirgin ilk yaklaşımından (7) sonra kısa sürede sayıları da artmıştır. Yaklaşımlar, öngörülen çeşitli başlık, madde, alt-madde özelliklerine uygunluğa göre belli puanlar alınarak ortaya çıkan ölçeksel değerlendirmelerdir (skala) veya yine öngörülmuş olan özellikler listesinden hangilerinin var olduğu denetlenerek yapılan gözden geçirmelerdir (onay listesi). Daha 1995 de 25 skala ve 9 onay listesi (checklist) in varlığı bildirilmiştir (8). Günümüzde ise sayının çok daha artmış olduğu açıktır. Ölçeksel yaklaşımların puan değişim aralığı birbirlerinden çok farklı olabilmektedir ve çeşitli nedenlerle yayın kalitesi değerlendirmesinde kullanılabilirler. Onay listeleri ise hem yayın editörlerinin, yollanan çalışmaları denetiminde yardımcı olmaktadır hem de RKA tasarlayan, sürdüren, yayına hazırlayan araştırmacılar için bir çeşit kılavuzluk oluşturabilmektedirler. Bu yaklaşımların içeriklerinden yararlanarak sadece RKA lar için değil, genel anlamda kıyassal yargılamalara dayalı tüm araştırmaların bilimsel düzeyini ölçümlenebilecek yaklaşımlara da gidilebilmiştir (9, 10).

Yöntemler'in Değerlendirmesi

Bu çalışmada, günümüzde en sık ve güvenilir olarak kullanılan 2 onay-listesini (*Delphi ve CONSORT listeleri*) ve 2 ölçeği (*Jadad ve PEDro* “skalaları”) gözden geçirdik.

A- “Delphi” listesi

Maastricht üniversitesinden bir grup araştırmacının 1998'de biçimlendirdiği listedir (1). Çalışmaya, önce kaynak taraması ile başlanarak RKA kalite değerlendirmelerinde kullanılması öngörülen tüm özellikler derlenmiş, 213 maddelik bir öncül liste düzenlenmiştir. Daha sonra, kaynaklardaki yayınların yazarları davet edilerek, yapısal açıdan ön-elemesi yapılan listede, katılımcıların (17 kişi) her bir madde için nicel ve nitel değerlendirmeleri alınarak ve devamında açık tartışmalar yapılarak 4 tur toplantı yapılmıştır. Her toplantı, listeden eksiltilebilecek veya içeriği-yapısı değiştirilebilecek madde sorgulamasını temel almıştır. Sonuçta, “büyük çoğunluğun genel kabulü” ile biçimlenen liste aşağıdaki 8 maddeden oluşmuştur:

1- Tedavi tipine dağıtım

- a) Belli bir randomizasyon yöntemi uygulanmış mıdır?
- b) Atama-dağıtım gizliliğine uyulmuş mudur?

- 2- Kümeler, başlangıçta, en önemli prognostik özellikler açısından eşdüzeyle midir?
- 3- Çalışmaya alınma-katılma ölçütleri belirtilmiş midir?
- 4- Sonuç ölçümleyici-saptayıcı "kör" müdür?
- 5- Tedavi uygulayıcı "kör" müdür?
- 6- Olgular (Hastalar) "kör" müdür?
- 7- Temel sonuç ölçütler için nokta kestirim ve değişkenlik ölçümleri belirtilmiş midir?
- 8- Değerlendirme, "tedavisi kararlaştırılmış olgular" (intention-to-treat) çerçevesinde de yapılmış mıdır ?

Sorulara; Evet / Hayır / Bilmiyorum yanıtları verilebilecektir.

Çalışmanın başındaki listede bulunan ve böylesi araştırmalarda temel özellik başlıkları olarak varsayılmış olan ; " Örneklem sayısı ve güç" , " Araştırma tasarımı", "Etik", "Tedavi girişimleri", "Sonuç veri ölçümleme", "İzleme süreç özellikleri", "Eşlik eden girişimler", "Yan etkiler", "Tedavi uyumu", "Sonuç yorum", "Sunum" kavramları ile ilgili maddelerden hiçbiri sonuç listeye erişmemiştir.

Listenin, 3 körlülüğe önem verirken örneklem sayısının önemine aynı değeri göstermemek, özet istatistiksel bilgilendirmeyi koşul olarak içerirken, başlangıçta, ikincil prognostik öğelerin rastlantısal olarak eşdüzenli geleme-yebileceğini bunun ise uygun istatistiksel yaklaşımlarla giderilebileceğini yoksaymak gibi çelişkili tarafları gözlenmektedir. Ayrıca çok sayıda prognostik öğeden sadece bazılarının eşdüzenli bulunabilecek olması bu maddeye uyumun genel yargılanmasında çelişki oluşturabilmektedir. "Delphi" listesi özel bir skortlama öngörmemektedir ancak çalışmanın sonunda "kalite" kavramı açısından bir tanım önerilmiştir: "Kalite; çalışmanın tasarım ve sürdürülmesinde yer alan böylece içsel ve dışsal geçerliliğin sağlanması ve uygun istatistiksel yöntemin kullanılması ile sonuçların geçerliliğini sağlayan ölçütler bütünüdür."

B- "PEDro" Ölçeği

Fizyoterapi konulu randomize klinik araştırmalar – ve sistematik derlemeler - için hazırlanan bir web-veri tabanına (11) (Physiotherapy Evidence Database) alınan çalışmaların bilimsel kalite düzeyinin belirtilmesi amacı ile oluşturulmuştur. Temel olarak Delphi listesinin tüm maddeleri kullanılmıştır ancak daha nesnel özellikleri olan 3 soru daha eklenmiştir. İçeriği ve kullanım özelliklerine elektronik ortamda da erişilebilen (12), ölçek maddeleri aşağıda listelenmektedir:

- 1- Çalışmaya alınma-katılma ölçütleri belirtilmiştir.
- 2- Olgular kıyas kümelerine rastlantısal olarak atanmışlardır.
- 3- Atama-dağıtım gizliliğine uyulmuştur.
- 4- Kümeler, en önemli prognostik temel özellikler açısından, başlangıçta eşit düzeylidirler.
- 5- Tüm Olgular (Hastalar) “kör” dür.
- 6- Tedavi uygulayıcılar “kör” dür.
- 7- En az bir temel sonuç-çıktıyı ölçümleyici-saptayıcıların tümü “kör” dür.
- 8- En az bir temel sonuç-çıktının ölçümü, her kümeye başlangıçta atanmış tüm olguların en az % 85’inde saptanabilmiştir.
- 9- Her olgu, sonuçları açısından, başlangıçta atandığı küme içinde değerlendirilmiştir. Bu sağlanamadı ise en az bir temel sonuç-çıktı; “tedavisi kararlaştırılmış olgular” (intention-to-treat) çerçevesinde değerlendirilmiştir.
- 10- En az bir temel sonuç-çıktı açısından kümeler arası istatistiksel kıyaslama sonuçları bulunmaktadır.
11. Çalışmada, en az bir temel sonuç-çıktı açısından hem özekselle “nokta” değerler hem de değişim değerleri bulunmaktadır.

Öngörülen özelliğin bildirildiği maddelere, Evet/Hayır yanıtları verilmektedir. İlk madde dışındaki konular puanlanmakta ve “Evet” yanıtları “1” değeri almaktadır böylece ölçekten alınabilecek en yüksek değer 10 olabilmektedir.

Ölçek için bir yargılama sınırı öngörülmemiştir ve PEDro veritabanında, çalışmalar çok düşük de olsa aldıkları puanlarla sunulmaktadır. Daha gelişmiş içerikli olduğu için bu ölçeğin kullanımına Delphi listesinden daha fazla rastlanmaktadır ve sonuçlarının geçerliliği, tutarlılığı da kanıtlanmış durumdadır (13). Burada, Delphi listesinden gelen ve kıyas kümelerinin başlangıç eşdüzenliliğini öngören 4. maddenin en düşük ICC (intra-class correlation coefficient) ye sahip olması (0.36: 0.27-0.44) dikkat çekicidir (13).

Randomize bir klinik araştırmanın kıyassal bir değerlendirmeyi temel alacağı, bunun bilimsel geçerliliği olabilmesi için de biyoistatistiksel bir yargılamanın zorunluluğu ve bu çerçevede istatistiksel özet değerlerin de sunulmasının bir gereklilik olduğu düşünülürse, 10. ve 11. maddeler, “olmaz ise olmaz” özelliklerin boşuna denetimi olarak düşünülebilir. PEDro ölçeğinin sadece Fizyoterapi konulu randomize klinik araştırmalarla sınırlanmayıp daha geniş boyutta kullanılmasına ve bu yönde de değerlendirilmesine rastlanmaktadır bu bağlamda ölçek maddelerine uygunluk oranlarını da elde

etme olanağı ortaya çıkmıştır. “Stroke” ile ilgili 272 yayının değerlendirildiği bir çalışmada 6. maddenin uygunluğu (Tedavi uygulayıcıların körlüğü) ancak % 8.5 oranında – rehabilitasyon yöntemi, çalışma özelliğinin koşullarına bağlı olarak – sağlanabilmişken, Olguların “körlüğü” benzer nedenlerle % 21.7 düzeyinde kalmıştır. Buna karşılık rasgele dağıtım ve ikili kıyaslama bildiri oranları % 97’nin üzerindedir ve başlangıç koşullarının benzeşliği de ilginç şekilde % 90’ın üzerinde bildirilmiştir (14).

C- “Jadad” Ölçeği

Jadad ve arkadaşlarının, ağrı ile ilgili araştırmalarda , “körlemenin” kaliteye katkısını irdelemek amacı ile tasarladıkları bir ölçektir (3). 7 madde içermektedir; bunların ilk 5 tanesi kalite açısından olumluluk göstergesidir ve varlıkları puan getirir, son ikisi ise, irdelenerek puan verilmiş maddelerin uygun yöntem içermemeleri durumunda puan indirimi öngörmektedirler. Bu yaklaşımla Jadad ölçeği 0-5 arası puan almaktadır:

	<u>Ev</u>	<u>Hayır</u>
1- Çalışma “randomize” olarak tanımlanmış mıdır? (Randomize, rastgele-rassal dağıtım, rassallama terimlerinden birinin yöntem açıklamasında kullanılmış olması gerekir)	1	0
2- Randomizasyonu sağlamak için uygulanan yöntem tanımlanmıştır ve “uygun” bir yöntemdir. (Rastgele sayılar yöntemi, bilgisayar üretimli , vs)	1	0
3- Çalışma “çift kör” olarak tanımlanmış mıdır?	1	0
4- Çift körleme için uygulanan yöntem tanımlanmıştır ve “uygun” bir yöntemdir. (Benzeş plasebo, etkin plasebo, vs)	1	0
5- Kayıplar ve çalışmadan çıkartılan olgular tanımlanmış, açıklanmış mıdır ?	1	0
6- Randomizasyonu sağlamak için uygulanan yöntem tanımlanmış ama yetersiz, uygunsuz bir yöntem ise “1” puan indirilir.(Dönüşümlü atama, doğum tarihine göre atama, vs) .		
7- Çalışma “çift kör” olarak tanımlanmış ama yetersiz, uygunsuz bir yöntem ise “1” puan indirilir.(Benzeş kullanılmaksızın tablek, enjeksiyon kullanımı gibi).		

Jadad ölçeği oldukça yaygın bir kullanım bulmuştur ve başlangıç için öngörüldüğü kullanım sınırlarının da dışına çıkarak farklı tıp alanlarında uygulanmıştır. Kullanıcılar kendi gereksinimleri veya düşünceleri çerçevesinde değişiklikler gerçekleştirerek uygulama yapmışlardır; bunlardan dikkat çeken birinde randomizasyon kullanıldığından söz edilmesine karşın (madde 1) yöntemle ilişkin tanımlama yapılmadı ise bir puan indirilmiştir. Benzer şekilde körlemenin bildirilmesi ile bir puan kazanılmış ama zorunlu olarak çift bildirilmesi değil, çalışmanın zorunlu özelliklerine bağlı olarak gereğinde “tek” kör olarak bildirilmesi de kabul edilmiştir. Körlük bildiri yapılmamış olması – çalışma kör olmayabilir – puana etkili olmamış ama bildirimle rağmen yöntemde yanlışlık bulunması bir puan azaltmıştır (15).

D- CONSORT Listesi

Gözden geçirilen 3 değerlendirme yaklaşımı kısıtlı sayıda madde ile en kısa yoldan bilimsel kalite değerlendirmesi yapabilmeyi amaçlamaktadırlar. Bu çerçevede bir RKA'nın taşınması zaten zorunlu gözükken özellikleri –çelişkili durumlar gözlenirse de - tekrar gözden geçirmemektedirler veya tersine tüm RKA lar için genellenemeyecek özgün ayrıntıları önemseyebilmektedirler. CONSORT listeleri ise en geniş anlamda alınabilecek tüm özellikleri ile RKA ları değerlendirebilme olanağı sağlamak için oluşturulmuşlardır. CONSORT (Consolidated Standarts for Reporting Trials: Deneysel araştırma yayınlanmasında bütünleştirilmiş standartlar) un oluşturulması ile ilgili ilk çabalar 1990'ların ortasında başladı ve biyoistatistikçiler, epidemiyologlar, tıp araştırmacıları, tıpsal dergilerin editörlerinden oluşan uluslar arası bir çalışma grubu tarafından hazırlanan 30 maddelik ilk liste yayınlandı (16). Yaklaşım tıp dergilerinin editörlerince büyük destek gördü ve ICMJE (International Committee of Medical Journal Editors), CSE (Council of Science Editors), WAME (World Association of Medical Editors) tarafından önerilerek pek çok önemli derginin yayın kabul etme ölçütleri arasına girdi. Günümüzde, içlerinde Lancet, NEJM gibi çok önemli dergilerin de bulunduğu 200 civarında tıp dergisi de bu liste maddelerine uymak koşulunu uygulamaktadırlar, listede şimdilik bir Türk dergisi bulunmaktadır – Journal of Neurological Science- (17). CONSORT, kendine ait web sitesinde ilgili tüm bilgileri ve madde ayrıntılarını açıklamaktadır (17). Yapılan araştırmalar ile CONSORT kullanımının RKA'ların bilimsel kalitesini arttırarak, hatalı sonuç değerlendirmelerini azalttığı ve tedavi etkinliği kararlarının alınmasında yararlı olduğu kanıtlandı (18, 19). Listenin daha güçlendirilmesi amacı ile süreç içinde çeşitli düzenlemelere de gidilerek 2001 de, 22 maddelik son liste kullanıma sunuldu (20). Liste denetiminin yanı sıra, olguların araştırma süreci boyunca sayısal, oransal dağılımlarını yansıtan bir de akış çizelgesi hazırlanması önerilmektedir. CONSORT listesi, bir randomize klinik araştırmanın sadece bilimsel geçerliliğini değerlendirme yararının ötesinde, çalışmanın gerek süreci gerekse de yazım düzeni

çerçevesinde denetim maddeleri de içermektedir. Bu nedenle yayın hazırlayanlar için de bir kılavuz niteliği taşımaktadır. Aşağıda listelendiği üzere, ideal bir RKA yayınının 5 temel bölümü ile (Başlık-Özet / Giriş / Yöntem /Sonuçlar / Tartışma) ilgili maddeler, gereğinde alt bölümler de oluşturarak –randomizasyon ile ilgili 3 madde bulunmaktadır– değerlendirilmektedir.

BAŞLIK ve ÖZET

1- Olguların girişim kümelerine dağıtımına ilişkin bilgilendirmenin (“randomizasyon”, “randomize”, “rastgele (rassal) atama” gibi terimler kullanılarak Başlık ve/veya, Özet’de bulunması.

GİRİŞ

2- Çalışmanın bilimsel tabanının ve gerekçesinin açıklanması.

YÖNTEM

3- Olgular: Katılanların için seçilme ölçütleri ve verilerin toplanma koşul ve yerlerinin açıklanması

4- Girişimler: Kümelerde girişimlerin ayrıntısının verilmesi, ne zaman ve nasıl uygulandıklarının bildirilmesi.

5- Amaç: Özgün amaç(lar) ve varsayımların bildirilmesi.

6- Sonuç çıktılar (Outcomes): Birincil ve ikincil sonuç çıktı tiplerinin açık olarak tanımlanması, gereğinde, ölçüm koşullarının iyileştirilmesi için alınmış yöntemlerin de açıklanması.

7- Örneklem Büyüklüğü: Örneklem Büyüklüğünün nasıl saptandığı ve gereğinde interim çözümleme ve “çalışma durdurma” koşullarının açıklanması.

8- Randomizasyon (sistemin açıklanması)

Randomizasyon (rastgele stama) düzeninin nasıl oluşturulduğunun açıklanması. Varsa kısıtlamalar, düzenlemelerle ilgili (Bloklama, katmanlama gibi) ayrıntıların bildirilmesi.

9- Randomizasyon (*allocation concealment*: atanma gizliliği)

Randomizasyon düzeni sıralamasının nasıl uygulandığı (ör.: numaralı kutucuklar, telefon sistemi) ve atama sonuçlarının girişimler başlayana dek gizli tutulması ile ilgili ayrıntıların açıklanması.

10- Randomizasyon (*implementation*: düzenleme ekibi-işlevi)

Randomizasyon düzenini üretenlerin, kayıtları yapanların, olguları kümelerine aktaranların bildirilmesi.

11- Körlük (Blinding): Olguların, girişimleri gerçekleştirenlerin, sonuçları ölçümleyen-kayıtlayanların “körlük” durumlarının açıklanması. Yapıldı ise körlük başarısının değerlendirilmesi.

12- İstatistiksel yöntemler: Sonuçların (ana ölçütler) değerlendirilmesi için kullanılan yöntemlerin açıklanması, alt küme çözümlenmeleri, düzeltme çözümlenmeleri gibi ek düzeltmelerin ayrıntılarının verilmesi.

SONUÇLAR

13- Olgu akışı: Olguların, çalışma süreci içindeki sayılarının değişimi (akış diagramı yararlı olur.) Çalışmaya uygun olanların, atananların, çalışmayı tamamlayanların, sonuçları değerlendirilebilenlerin sayıları. Varsa çalışma protokolündeki süreç içi değişikliklerinin nedenleri ile açıklanması.

14- Veri toplama zamanlaması: İzleme süresinin ve veri toplama zamanlamalarının tarihlerinin bildirilmesi.

15- Temel özellikler (baseline data): Tüm kümelerin, demografik, klinik, temel-öncül özelliklerinin bildirilmesi

16- Değerlendirilen olgu sayıları: Her değerlendirme için, her kümede değerlendirmeye katılan (payda) olgu sayısının bildirilmesi, değerlendirmenin "intention-to treat" tedavisi kararlaştırılmış olgular değerlendirmesi olup olmadığının bildirilmesi. (Uygun durumlar için sonuçların % olarak değil de pay/payda şeklinde verilmesi yeğlenmelidir.)

17- Sonuçlar–Kestirimler: Her birincil ve ikincil sonuç için, her kümenin özetsel bilgisi, Etki büyüklüğü (*effect size*) kestirimlerinin ve kesinlik sınırlarının (ör.: % 95 CI) bildirilmesi.

18- Ek değerlendirmeler: Ana değerlendirmenin yanı sıra gerçekleştirilen alt-küme çözümlenmelerinin ve/veya düzeltmeli değerlendirmelerin açıklanması. Bunların hangilerinin önceden kararlaştırılmış hangilerinin ayrıntıya inmek adına ortaya çıktığının bildirilmesi.

19- Beklenmedik olaylar- Yan etkiler: Tüm girişim kümelerindeki "Beklenmedik olayların" ve yan etki bildirimlerinin dökümü.

TARTIŞMA

20- Yorumlama: Sonuçların, başlangıç varsayım(lar)ı çerçevesinde, çalışmada yapılabilmemiş olan bias ve belirsizliklerin, çoklu değerlendirmelerin, ek çözümlenmelerin de göz önünde bulundurularak değerlendirilmesi.

21- Genelleme: Çalışma sonuçlarının, "genellenebilirliği" (dışsal geçerlilik).

22-Genel sonuç: Geçerli kanıtlar çerçevesinde sonuçların genel yorumlaması

Değerlendirme, her maddeye eğer, yeterince sağlanmış ise , "1" puan vererek gerçekleştirilmektedir. Bazı araştırmalarda, çalışmanın özellikleri çerçevesinde, tüm maddelerin değerlendirilebilir durumda olmamaları ile de karşılaşılabilir (21). Toplam değerlendirme için, genellikle kabul gören bir yargı sınırı önerisi bulunmamaktadır. Bazı araştırmalarda "düşük",

“orta” , “yüksek” şeklinde, puanlama aralığını 3 eşit parçada ele alan ayırım yaklaşımları önerildiği görülmüştür (9, 22). Burada aynı zamanda “yöntem-bilimsel-bilimsel kalite” ile “yayınlama” kalitesinin kesin bağıntısı olamaya-bileceği sorunsalı da ortaya konmuştur (22). CONSORT listesi sıklıkla yayın kümelerinin genel yorumlamasını yapmak için de kullanılmaktadır (9, 21, 22, 23); bunlarda, ele alınan yayınlarda, hangi oranda (veya doğrudan sayıları ile), hangi maddenin sağlanmış olduğu yönünde değerlendirme yapılmaktadır.

CONSORT uyma koşulu ile yayın kabul eden dergiler kendi kabulleri çerçevesinde belli eksikliklere göz yumabilir gözükmedirler (24).

Tartışma ve Sonuç

Özellikle “kanıta dayalı tıp” uygulamaları açısından RKA’lar büyük önem taşımaktadırlar. Klinik karar süreçlerini etkileme özellikleri nedeni ile sonuçları açısından yüksek güvenilirlik taşımaları, bunun sağlanabilmesi için ise bilimsel “kalite” yönünden yüksek olmaları gerekir. Kalite değerlendirme yöntemi arayışları, büyük okuyucu-uygulayıcı kitlelerine ulaşabilecek yayınlardan, belli bilimsel düzeyin üzerindeki seçebilmeyi ayrıca araştırmacılara kılavuzluk etmeyi amaçlar.

Ele aldığımız yöntemlerin değerlendirmesinden anlaşılacağı üzere, bu iki amacın birlikte taşınması değerlendirme yönteminin daha kapsamlı ve bazen karmaşık olmasını getirebilmektedir. RKA gerçekleştirmek amacı ile yola çıkmış bir çalışmanın *sine qua non* yönlerinin belirlenerek, listeye konmaması madde sayısını azaltmak yönünde bir adım olabilir. Böylesi “olmaz ise olmaz” özellikleri bile taşımayan çalışmalar herhangi bir ölçek değerlendirmesi aşamasına bile erişemeyeceklerdir. Buna karşılık değerlendirme listesinin aynı zamanda kapsamlı bir kılavuz rolü üstlenmesi de isteniyorsa, daha temel öğeleri de içeren, tanımları açık maddeler listede bulunmalıdır.

Farklı tıp bilimlerinin sürdürdükleri RKA’ların, alanın koşulları çerçevesinde özgünlükler göstereceği açıktır. Bu nedenle geniş kapsamlı listelerin tüm maddelerinin her alanda değerlendirilemediği, bazı alanların ise kendi değerlendirmeleri için listelere ek maddeler oluşturdukları gözlenmektedir (25). Genel kullanımlı bir değerlendirme yaklaşımı herhalde ortak-ana özelliklerle ilgili maddeleri içermeli ayrıca uygun alanlar için modüler ekler kullanarak sonuçta belli değerlendirme aralığını korumalıdır. Bu görüşün CONSORT oluşturucuları tarafından da düşünüldüğü saptanmaktadır (26). Ayrıca farklı araştırma tipleri için içeriği doğal olarak değişik olacak listelerin oluşturulmasına da gidilmektedir; Randomize araştırmaların sistematik derlemeleri için QUORUM, gözlemsel araştırmaların sistematik

derlemeleri için MOOSE, tanı testleri arařtırmaları için STARD, gözlemsel arařtırmalar için STROBE bunlara uygun örnekler olarak bildirilmektedir (26).

Kılavuz ölçek düzeyinde kapsamlı listelerde, Jadad ölçeğinde gözlemlendiđi gibi bazı temel özelliklerin deđerlendirmesi sadece puan katkısı ile oluşmamalı ve geređinde puan azaltımına gidilmelidir.

Listenin özelliđi ne olursa olsun, tüm maddelerin aynı puan ađırlılıđında olması da tartıřılan bir yöndür ve herhalde ađırlıklı katkı düzeylerinin biçimlendirilmesine gidilmelidir.

Ölçekler boyutunda yargı deđerlendirme sınırları netleřtirilmiř deđildir, sadece puanların verilmesi okuyucuyu-uygulayıcıyı da kararları açısındandır belirsizlik içinde bırakabilmektedir.

Sonuç olarak, kalite deđerlendirmesi yönünde yeni maddelerin, puansal ađırlıkların, ölçek deđerlendirme "alanlarının", yargı sınırlarının belirlenmesi, arařtırma tasarımı boyutunda, gerekmektedir.

Kaynakça

1. Verhagen AP, de Vet HCW, de Bie RA, Kessels AGH, Boers M, Bouter LM, Knipschild PG. The Delphi List: A Criteria List for Quality Assessment of Randomized Clinical Trials for Conducting Systematic Reviews Developed by Delphi Consensus, *Jornal of Clinical Epidemiology*, 51(12):1235–1241, 1998.
2. Sauerland S, Davis TRC. The Consolidated Standards O Reporting Trials Consort):Better Presentation of Surgical Trials in The Journal of Hand Surgery, *Journal of Hand Surgery*, 29B(6): 621–624, 2004.
3. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials*, 17: 1–12, 1996.
4. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association*, 273:408-412, 1995.
5. Ruiz-Canela M, Martinez-Gonzalez MA, de Irala-Estevez J. Intention to treat analysis is related to methodological quality. *British Medical Journal*, 320:1007-1008, 2000.
6. Verhagen AP, de Vet HCW, de Bie RA, Boers M, van den Brandt PA. The art of quality assessment of RCTs included in systematic reviews, *Journal of Clinical Epidemiology*, 54: 651–654, 2001.

7. Chalmers TC, Smith H, Blackburn B, Silverman B, Schoeder B, Reitman D. A method for assessing the quality of a randomized control trial. *Controlled Clinical Trials*, 2:31–49, 1980.
8. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists, *Controlled Clinical Trials*, 16:62-73, 1995.
9. Boulware LE, Daumit GL, Frick KD, Minkovitz CS, Lawrence RS, Powe NR. Quality of Clinical Reports on Behavioral Interventions for Hypertension, *Preventive Medicine*, 34: 463–475, 2002.
10. Çağatay P., Şenocak M, Baykal E. Randomize Klinik Deneyleerin Değerlendirilmesinde Bir Ölçek Önerisi. *Cerrahpaşa J Med* 31(1):49-55, 2000.
11. Sherrington C, Herbert RD, Maher CG, Moseley AM. A database of randomized trials and systematic reviews in physiotherapy, *Manual Therapy*, 5(4): 223-226, 2000.
12. PEDro. The Physiotherapy Evidence Database (PEDro) frequently asked questions: how are trials rated? http://www.pedro.fhs.usyd.edu.au/scale_item.html . Erişim : 12 Mayıs 2005 .
13. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro Scale for Rating Quality of Randomized Controlled Trials, *Physical Therapy*, 83:713-721, 2003.
14. Bhogal SK, Teasell RW, Foley NC, Speechley MR. The PEDro scale provides a more comprehensive measure of methodological quality than the Jadad Scale in stroke rehabilitation literature, *Journal of Clinical Epidemiology*, 2005, Article in press.
15. Gummesson C, Atroshi I, Ekdahl C. The Quality of Reporting and Outcome Measures in Randomized Clinical Trials Related to Upper-Extremity Disorders, *The Journal of Hand Surgery*, 29A(4):727-734, 2004.
16. Begg CB, Cho MK, Eastwood S. Improving the quality of reporting of randomized controlled trials: the CONSORT statement, *JAMA-Journal of The American Medical Association*, 276:637-639, 1996.
17. CONSORT: [on-line]: <http://www.consort-statement.org>. Erişim 10 Mayıs 2005.
18. Moher D, Jones A, Lepage L, for the CONSORT Group. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA-Journal of The American Medical Association*, 285:1992-1995, 2001.

19. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Guyatt GH. The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the Consolidated Standards of Reporting Trials (CONSORT) checklist, *Controlled Clinical Trials*, 23:380–388, 2002.
20. Moher D, Schulz KF, Altman D, for the CONSORT Group. The Consort Statement: Revised Recommendations For Improving The Quality Of Reports Of Parallel-Group Randomized Trials 2001, *EXPLORE*, 1(1):40-45, 2005.
21. Halpern SH, Darani R, Douglas MJ, Wight W, Yee J. Compliance with the CONSORT checklist in obstetric anaesthesia randomised controlled trials, *International Journal of Obstetric Anesthesia*, 13:207–214, 2004.
22. Huwiler-Müntener K, Jüni P, Junker C, Egger M. Quality of Reporting of Randomized Trials as a Measure of Methodologic Quality, *JAMA- Journal of The American Medical Association*, 287:2801-2804, 2002.
23. Mills EJ, Ping Wu T, Gagnier J, Devereaux PJ. The quality of randomized trial reporting in leading medical journals since the revised CONSORT statement, *Controlled Clinical Trials* (2005), Baskıda.
24. Mills E., Wu P., Gagnier J., Heels-Ansdell D., Montori VM. An analysis of general medical and specialist journals that endorse CONSORT found that reporting was not enforced consistently. *Journal of Clinical Epidemiology*, 58 : 662–667, 2005.
25. Loannidis JP., Evans SJ., Gøtzsche PC., O'Neill RT., et al. Better Reporting of Harms in randomized Trials: An Extension of the CONSORT Statement . *Ann Int Medicine*, 141: 781-789, 2004.
26. Moher D., Altman DG., Schulz KF., Elbourne DR. Opportunities and challenges for improving the quality of reporting clinical research: CONSORT and beyond. *Canadian Medical Association Journal*, 171: 349-351, 2004.

HASTA MEMNUNİYETİNE İLİŞKİN VERİ SETİNDE BOYUT İNDİRGEME YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Mevlüt Türe¹, İmran Kurt², Zekeriya Aktürk³

Trakya Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı¹

Osmangazi Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı²

Trakya Üniversitesi Tıp Fakültesi Aile Hekimliği Anabilim Dalı³

e-mail: [ture@trakya.edu.tr](mailto:tire@trakya.edu.tr)

Özet

Bu çalışmada, 2005 yılında Trakya Üniversitesi Tıp Fakültesi hastanesine gelen hastaların hastane personelinden memnuniyetini ölçen maddeler üzerinde klasik temel bileşenler analizi (TBA), genelleştirilmiş temel bileşenler analizi (GTBA), yapay sinir ağları (YSA) ile doğrusal temel bileşenler analizi (YSA-TBA) ve YSA ile doğrusal olmayan temel bileşenler analizi (YSA-DOTBA) yöntemlerinin karşılaştırılması amaçlandı. Hasta memnuniyeti ile ilgili veri setinde doktor, hemşire, radyoloji teknisyeni, laboratuvar teknisyeni ve diğer personel için elde edilen bileşenlerin en büyük varyans açıklayıcılık oranları YSA-TBA ve YSA-DOTBA'dan elde edildi. Bu da araştırmacılara klasik yöntemlere göre daha büyük varyans açıklayıcılık oranı veren YSA'nın boyut indirgemedeki kullanılması gerektiğini göstermektedir.

Anahtar Kelimeler: Temel Bileşenler Analizi, Yapay Sinir Ağları, Doğrusal Olmayan Temel Bileşenler Analizi

1. Giriş

İncelenen olaylarda değişkenlerin birbirleriyle ilişkili ve değişken sayısının fazla olması sonuçların yorumlanması, özetlenmesi ve çok değişkenli istatistiksel analizlerin uygulanmasını güçleştirmektedir. Temel bileşenler analizi (TBA), değişkenler arasındaki bu bağımlılık yapısının yok edilmesi ve p sayıda değişkenin doğrusal bileşeni olan m sayıda değişkene ($p > m$) indirgemesi amacıyla geliştirilmiş bir yöntemdir [1, 2]. TBA oransal, aralıklı ya da sıralı ölçekli değişkenler arasında sadece doğrusal ilişkileri incelemektedir. TBA'ya alternatif olarak geliştirilen genelleştirilmiş temel bileşenler analizi (GTBA) ise isimsel, sıralı, aralıklı ve oransal ölçekli değişkenlerden oluşan veri setlerinde kullanılabilen optimal ölçekleme yöntemidir.

Günümüzde TBA ve GTBA dışında boyut indirgeme yaklaşımı olarak kullanılan bir diğer yöntem ise yapay sinir ağları (YSA)'dır. Biyolojik sinir ağlarından esinlenilerek geliştirilen YSA, insan beyninin bilgiyi taşıma ve depolama gibi hareketlerini taklit eden ve değişken yapısı konusunda herhangi bir kısıtlama getirmeksizin değişkenler arasındaki ilişkileri hem doğrusal hem de doğrusal olmayan biçimde ele alarak temel bileşenleri belirleyebilme yeteneğine sahip bir yöntemdir.

Bazı araştırmacılar bu yöntemlerin performanslarını farklı alanlardaki veri setlerinde incelemişlerdir. Dong ve McAvoy [3], görüntü sıkıştırma TBA ve yapay sinir ağları ile doğrusal olmayan temel bileşenler analizi (YSA-DOTBA)'ni karşılaştırmışlardır. Monahan [4], iklim ile ilgili bir veri setini kullanarak TBA ile YSA-DOTBA yöntemlerinin boyut indirgeme performanslarını karşılaştırmıştır. Albanis ve Batchelor [5], uzun dönemli kredi sürekliliğinin değerlendirilmesine ilişkin veri setinde finansal oranların boyut indirgemesinde TBA, YSA-TBA ve YSA-DOTBA'yı karşılaştırmışlardır. Hsieh [6], Pasifik Denizi'nin yüzey sıcaklığı verilerini kullanarak TBA, döndürülmüş TBA ve YSA-DOTBA yöntemlerini karşılaştırmışlardır.

Bu çalışmada, 2005 yılında Trakya Üniversitesi Tıp Fakültesi Hastanesi'ne gelen 294 hastanın hastane personelinden memnuniyetini ölçen soruların boyutunun indirgenmesinde TBA, GTBA, yapay sinir ağları ile doğrusal temel bileşenler analizi (YSA-TBA) ve doğrusal olmayan temel bileşenler analizi (YSA-DOTBA) yöntemlerinin karşılaştırılması amaçlandı.

2. GEREÇ ve YÖNTEMLER

2.1. Veri

Çalışmamız, 2005 yılında Trakya Üniversitesi Tıp Fakültesi Eğitim Uygulama ve Araştırma Hastanesi polikliniklerine herhangi bir rahatsızlık nedeni ile başvuran 294 hastadan (% 46.3 erkek ve % 53.7 kadın) oluşmaktadır. Hastalara, hastane personelinden memnuniyetlerini değerlendirmek için alt kategorileri (Çok kötü-1, Kötü-2, Orta-3, İyi-4, Çok iyi-5) 5'li Likert ölçekli 31 maddeden oluşan Hasta Memnuniyeti Anket Formu uygulandı.

Anket verilerinin güvenilirlik katsayısı (Cronbach α) 0.96 bulundu. Ayrıca 31 maddelik anket soruları doktor, hemşire, radyoloji teknisyeni, laboratuvar teknisyeni ve diğer personel memnuniyeti olmak üzere 5 gruba ayrıldı. Her bir grubun güvenilirlik katsayısı; doktor için 0.96, hemşire için 0.93, radyoloji teknisyeni için 0.91, laboratuvar teknisyeni için 0.93 ve diğer personel için 0.84 bulundu (Tablo 1).

Tablo 1. Gruplara ilişkin güvenilirlik katsayıları

Soru		Cronbach α
Doktor memnuniyetine ilişkin maddeler		
1	Doktorun görüşmeler sırasında yeterli zaman ayırması	0.96
2	Doktorun sorunlarınızı ona söylemenizi kolaylaştırması	
3	Doktorun tıbbi bakımınızla ilgili kararlara sizi de katması	
4	Doktorun sizi dinlemesi	
5	Doktorun şikayetlerinizi çabuk geçirmesi	
6	Doktorun sizi muayene etmesi	
7	Doktorun test ve tedavilerin amaçlarını açıklaması	
8	Doktorun istediğiniz bilgileri vermesi	
9	Doktorun kaygı/endişelerinizle başa çıkmanıza yardımcı olması	
10	Doktorun tavsiyelerine uymanızın önemini kavramanıza yardımcı olması	
11	Doktorun önceki görüşmelerde yaptıklarını ve söylediklerini bilmesi	
12	Doktorun size karşı saygısı	
13	Doktorun ilgisi / dinlemesi	
14	Doktorun güler yüzlü olması	
15	Doktorun sizi bilgilendirmesi	
Hemşire memnuniyetine ilişkin maddeler		
1	Hemşirenin size karşı saygısı	0.93
2	Hemşirenin ilgisi / dinlemesi	
3	Hemşirenin güler yüzlü olması	
4	Hemşirenin sizi bilgilendirmesi	
Radyoloji teknisyeni memnuniyetine ilişkin maddeler		
1	Radyoloji teknisyeninin saygısı	0.91
2	Radyoloji teknisyeninin ilgisi / dinlemesi	
3	Radyoloji teknisyeninin güler yüzlü olması	
4	Radyoloji teknisyeninin sizi bilgilendirmesi	
Laboratuvar teknisyeni memnuniyetine ilişkin maddeler		
1	Laboratuvar teknik personelinin saygısı	0.93
2	Laboratuvar teknik personelinin ilgisi / dinlemesi	
3	Laboratuvar teknik personelinin güler yüzlü olması	
4	Laboratuvar teknik personelinin sizi bilgilendirmesi	
Diğer personel memnuniyetine ilişkin maddeler		
1	Kayıt/evrak personelinin size karşı ilgisi ve sizi dinlemesi	0.84
2	Kayıt/evrak personelinin güler yüzlü olması	
3	Kayıt/evrak personelinin size karşı saygısı	
4	Güvenlik görevlilerinin tutumu	

2.2. Temel Bileşenler Analizi

Temel bileşenler analizi, aralarında ilişki bulunan p değişkenin açıkladığı yapıyı, aralarında ilişki bulunmayan daha az sayıda ($m < p$) orijinal değişkenlerin doğrusal bileşenleri olan değişkenlerle ifade etme yöntemidir.

Temel bileşenler analizinde p boyutlu uzay yerine m boyutlu bir uzayda çalışarak boyut indirgemesi amaçlanır. Temel bileşenlerin elde edilmesinde;

$$X' = [X_1, X_2, \mathbf{K}, X_p]$$

p değişkene ilişkin girdi vektörünün kovaryans matrisinin özdeğerleri

$$\lambda_1 \geq \lambda_2 \geq \mathbf{K} \geq \lambda_p \geq 0$$

biçiminde elde edilir. Temel bileşenler varyanslarının büyüklüklerine göre sıralanır. Orijinal değişkenlik ile p temel bileşenin ($Y_1, Y_2, \mathbf{K}, Y_p$) varyansları toplamı birbirine eşit olacaktır. Ayrıca toplam varyans özdeğerler toplamına eşittir:

$$\text{Toplam Varyans} = \lambda_1 + \lambda_2 + \mathbf{K} + \lambda_p = \text{Var}(Y_1) + \text{Var}(Y_2) + \dots + \text{Var}(Y_p)$$

Temel bileşenler sırasıyla özdeğerlerin büyüklük sıralamasına göre elde edilir. En büyük özdeğerli temel bileşen varyans açıklayıcılığı en büyük olan birinci bileşen olacaktır. Bu özdeğerlere karşılık gelen özvektör matrisi,

$$e' = [e_1, e_2, \mathbf{K}, e_p]$$

biçiminde olacaktır. Buna göre X veri matrisinin doğrusal bileşenleri

$$Y = e'X'$$

eşitliğiyle elde edilir [1, 2, 7].

2.3. Genelleştirilmiş Temel Bileşenler Analizi

GTBA, karışık yapıdaki optimal ölçek düzeyindeki değişkenlerin tek sette değişimini hesaplamayı amaçlar. GTBA'da orijinal veri seti yerine, çok az bilgi kaybı ile yeni ve daha küçük veri seti kullanılır. Bu analiz yöntemi; isimsel, sıralı, aralıklı ve oransal ölçekli değişkenler arasındaki ilişkileri doğrusal olmayan yaklaşım ile analiz ettiği gibi birimler arasındaki ilişkileri de

ortaya koyar. GTBA optimal ölçeklemeli TBA olarak da bilinmektedir. Seçilen boyut sayısında değişkenler ile obje skorları arasındaki korelasyonu maksimize ederek çözüme ulaşan bir yöntemdir [8, 9].

GTBA'da değişken setindeki değişkenlerin bazıları sıralı (Likert Ölçeği), bazıları isimsel ölçekli olabilmektedir [9, 10].

GTBA'da kayıp fonksiyon minimize edilerek çözümlenir. Gifi sisteminde kayıp fonksiyon

$$\sigma(X; Y_1, \dots, Y_j) = J^{-1} \sum_{j=1}^J \text{SSQ}(X - G_j Y_j) = J^{-1} \text{tr}(X - G_j Y_j)'(X - G_j Y_j)'$$

minimize edilir. Burada;

Y_j ($k_j \times p$) boyutlu çoklu kategori açıklayıcılıkları matrisi ($Y_j = q_j \beta_j'$)

q_j h_j değişkeninin dönüştürülmüş şekli ($q_j = \frac{Y_j \beta_j}{\beta_j' \beta_j}$)

β_j bileşen yükleri matrisi ($\beta_j = \frac{(Y_j' D_j q_j)}{q_j' D_j q_j}$)

G_j ($n \times k_j$) boyutlu j değişkeninin gösterge matrisi

X ($n \times p$) boyutlu obje skorları matrisi ($X = J^{-1} \sum_{j=1}^J G_j Y_j$)

k_j j değişkeninin kategori sayısı

j değişken sayısı

D_j G_j 'nin diagonal matrisidir [9, 10].

2.4. Yapay Sinir Ağlarıyla Doğrusal Temel Bileşenler Analizi

Biyolojik sinir ağlarından esinlenilerek geliştirilen yapay sinir ağları boyut indirgeme için yaygın olarak kullanılan klasik temel bileşenler analizine alternatif bir yöntemdir.

p girdi verisini temsil edecek m ($m < p$) doğrusal çıktı bileşeni,

$$y_j(n) = \sum_{i=1}^p w_{ij}(n) x_i(n) \quad j=1,2,\dots,m$$

formülüyle elde edilir. Burada $w_{ij}(n)$, n'inci adımda i girdi değişkeni ile nöronunu bağlayan ağırlıktır. Sanger'ın kuralına göre ağırlıklar adım adım değiştirilir,

$$\Delta w_{ji}(n) = \eta \left[y_j(n) x_i(n) - y_j(n) \sum_{k=1}^j w_{ki}(n) y_k(n) \right]$$

$i=1,2,\dots,p$

Burada η , öğrenme oranı parametresidir. Yapay sinir ağlarıyla temel bileşenler analizinde, korelasyon matrisi hesaplanmaksızın girdinin korelasyon fonksiyonunun özvektörlerinin hesaplanmasıdır. Bu öğrenme yöntemi, TBA ağının ağırlıklarının girdi veri matrisinin ilk m temel bileşenine yakınsamasını sağlar. Yani öğrenme süreci sonunda ağırlıklar, özvektörlere yakınsayacaktır. Öğrenme yöntemine göre temel bileşenler sırayla elde edilir: birinci nöronun ağırlıkları yakınsadıktan sonra ikinci en büyük özdeğerle ilgili özvektöre ikinci nöronun ağırlıkları yakınsayacaktır [5,11,12].

2.5. Yapay Sinir Ağlarıyla Doğrusal Olmayan Temel Bileşenler Analizi

YSA-DOTBA'nın genel amacı, orijinal veri setinden gelen bilginin maksimum miktarını kaybetmeden yeni ve daha küçük veri seti elde etmektir. DOTBA, doğrusal olmayan bir özellik çıkarma yaklaşımıdır.

Yapay sinir ağları, boyut indirgemek için girdi, çıktı ve üç gizli tabaka olmak üzere 5 tabaka kullanır. Girdiler aynı zamanda ağın gözlenen çıktıları oluşturur. Doğrusal olmayan temel bileşenler analizinin uygulanabilmesi için ağın ikinci ve dördüncü tabakaları sigmoid aktivasyon fonksiyonlu olmalıdır. Aksi durumda ağ doğrusal temel bileşenler analizi uygulayacaktır. Bunun yanında boyut indirgeme ve çıktı tabakaları doğrusal aktivasyonlu olmalıdır. Boyut indirgeme tabakası olan üçüncü tabaka, p girdi nöronuna sahip birinci ve beşinci tabakadan daha az nörona ($m < p$) sahip olmalıdır.

YSA'nın amacı,

$$E = \sum_{i=1}^p (X_i - X'_i)^2$$

gözlenen çıktılar (X_i) ile beklenen çıktılar (X'_i) arasındaki hatayı minimize etmektir. Burada p , girdi yada değişken sayısıdır ($i=1,2,\dots,p$). Boyut indirgeme tabakasının boyutu girdi ve çıktı tabakasından daha küçük olduğu için ağ girdiyi temsil edebilecek kadar iyi bir temsilciyi bulmak için zorlanır. Öğrenme süreci uygun bir çözümlerle sonuçlanırsa, girdinin iyi bir temsilcisi boyut indirgeme tabakasında elde edilecektir. Bu durum boyut indirgeme tabakasında meydana gelen veri sıkıştırmasına rağmen girdilerin çok az bir kayba uğrayarak tahmin edilmesi anlamına gelmektedir. Bu yüzden boyut indirgeme tabakası çıktıları, doğrusal olmayan temel bileşenleri vermektedir:

$$h_2 = \sum_{i=1}^m w_{jk} f \left(\sum_{k=1}^{N_i} w_{ki} X_i \right)$$

YSA'nı minimum hataya ulaştıran en iyi ağırlıkları elde etme süreci yaygın olarak kullanılan geriye yayılma algoritmasıyla gerçekleştirilir [5, 6, 10, 12-14].

3. Bulgular

TBA ve GTBA yöntemlerinde veri setinin tamamı kullanıldı. YSA için veri seti % 80'i ($n= 235$) eğitim ve % 20'si ($n= 59$) geçerlilik seti olmak üzere rasgele 2 sete ayrıldı.

Boyut indirgeme yöntemlerinden elde edilen ve her bir gruptaki maddeleri temsil eden 1 bileşenin varyans açıklayıcılık oranları doktor için % 60.2-% 84.9, hemşire için % 80.5-% 88.7, radyoloji teknisyeni için % 79.2-% 96.1, laboratuvar teknisyeni için % 82.1-% 90.9 ve diğer personel için % 66.5-% 86.1 aralığında değişti (Tablo 2).

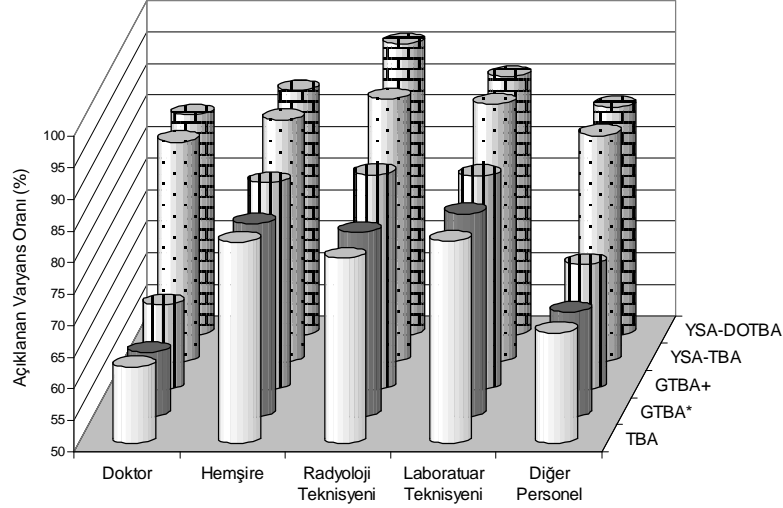
Tablo 2. Yöntemlerin bileşene göre açıklanan varyans oranları

Bileşen Adı	Değişken Sayısı	Bileşen Sayısı	Açıklanan Varyans Oranı (%)				
			TBA	GTBA*	GTBA*	YSA-TBA	YSA-DOTBA
Doktor	15	1	62.2	60.2	63.4	84.9	84.9
Hemşire	4	1	82.0	80.5	82.7	88.4	88.7
Radyoloji Teknisyeni	4	1	79.4	79.2	84.0	91.7	96.1
Laboratuvar Teknis.	4	1	82.1	82.1	83.8	90.9	90.9
Diğer personel	4	1	67.6	66.5	69.8	85.9	86.1

*: Değişkenler oransal ölçekli alınmıştır.

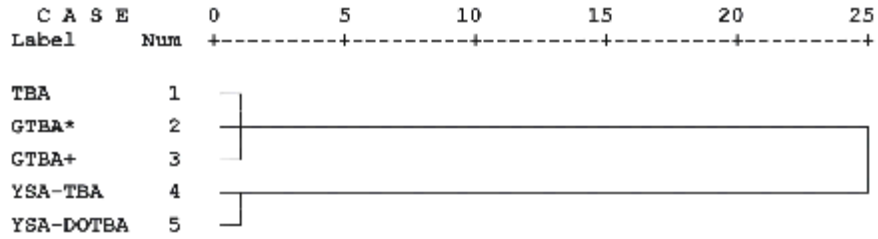
†: Değişkenler sıralı ölçekli alınmıştır.

Tablo 2 ve Şekil 1’de görüleceği gibi doktor, hemşire, radyoloji teknisyeni, laboratuvar teknisyeni ve diğer personel için en büyük varyans açıklayıcılık oranları YSA-TBA ve YSA-DOTBA yöntemlerinden elde edilmiştir. Sıralı ölçekli GTBA yöntemi, oransal ölçekli GTBA ve TBA’ya göre daha iyi performans göstermiştir.



Şekil 1. Yöntemlerin bileşene göre açıklanan varyans oranları (%)

Varyans açıklayıcılık oranları açısından yöntemler aşamalı kümeleme analizi ile gruplandırıldı. Dendrogram grafiğinden de görüleceği gibi yöntemler iki kümeye ayrıldı. Birinci kümede TBA, oransal ölçekli değişkenler için GTBA ve sıralı ölçekli değişkenler için GTBA yöntemleri yer alırken, ikinci kümede YSA-TBA ve YSA-DOTBA yöntemleri yer aldı (Şekil 2). Böylece açıklanan varyans oranı kriterine göre ikinci kümede yer alan yöntemlerin (YSA-TBA ve YSA-DOTBA) değişkenleri 1 bileşene indirgemede en iyi performansı gösterdiği bulundu.



Şekil 2. Dendrogram grafiği

4. Tartışma

Genellikle incelenen olaylar birden fazla değişkenin etkisi altında kalmakta ve yapılan çalışmaların geçerli ve güvenilir olması için incelenen olayı etkileyen birden çok değişkenin birlikte ele alınması gerekmektedir. Fakat değişkenlerin çok sayıda olması ve birbirleriyle ilişkili olması verilerin analizini güçleştirmektedir. Boyut indirgeme yöntemlerinden, p boyutlu uzayda çalışmak yerine veriyi m boyutlu ($p > m$) uzaya indirgeyerek ilişkisiz bileşenler belirlemede ve verilerin başka yöntemlerde kullanılmak üzere hazırlanmasından sıkça yararlanılmaktadır. TBA, doğrusal bileşenleri belirlemek için geliştirilmiş bir yöntem olduğu için doğrusal olmayan ilişkileri göz önünde bulundurmaz. Bu nedenle veri setindeki değişkenler arasında eğrisel ilişkiler varsa bu değişkenleri temsil edecek en iyi bileşenleri belirlemede yetersiz kalmaktadır.

Albanis ve Batchelor [5], uzun dönemli kredi sürekliliğinin değerlendirilmesine ilişkin veri setinde YSA-TBA ile YSA-DOTBA'nın boyut indirgeme performansının TBA'dan daha iyi olduğunu göstermişlerdir. Dong ve McAvoy [3], görüntülerin sıkıştırılmasında YSA-DOTBA'nın TBA'ya göre daha iyi olduğunu göstermişlerdir. Monahan [4], iklim ile ilgili bir veri setini kullanarak TBA ile YSA-DOTBA yöntemlerinin boyut indirgeme performanslarını karşılaştırmış ve YSA-DOTBA'nın daha güçlü bir yöntem olduğunu belirlemiştir. Hsieh [6], Pasifik Denizi yüzey sıcaklığıyla ilgili veri setini kullanarak TBA, döndürülmüş TBA ve YSA-DOTBA yöntemlerini karşılaştırmış ve YSA-DOTBA'nın daha iyi olduğunu belirlemiştir. Çalışmamızda ise hasta memnuniyeti ile ilgili veri setinde doktor, hemşire, radyoloji teknisyeni, laboratuvar teknisyeni ve diğer personel için elde edilen bileşenlerin en büyük varyans açıklayıcılık oranları YSA-TBA ve YSA-DOTBA yöntemlerinden elde edilmiştir. Sıralı ölçekli GTBA yöntemi, oransal ölçekli GTBA ve TBA'ya göre daha iyi performans göstermiştir.

Sonuç olarak, varyans açıklayıcılık oranı daha yüksek olabilecek iken daha küçük açıklayıcılık oranına sahip bileşenlerle çalışmaların yürütülmesi yerine pek çok alanda ve pek çok klasik yöntem alternatif yöntemler içeren YSA'nın boyut indirgemede de kullanılması gerektiği görülmektedir. YSA'nın, doğrusal ilişkilerin yanında doğrusal olmayan ilişkileri de göz önünde bulundurarak en iyi bileşenleri belirleme yeteneği göz ardı edilmemelidir.

Kaynaklar

1. Özdamar K. Paket Programlar ile İstatistiksel Veri Analizi 2. Eskişehir: Kaan Kitabevi; 1999.

2. Tatlıdil H. Uygulamalı Çok Değişkenli İstatistiksel Analiz. Ankara: Akademi Matbaası; 1996.
3. Dong D, MacAvoy TJ. Batch tracking via nonlinear principal component analysis. *AIChE Journal* 1996; 42(8): 2199-2208.
4. Monahan AH. Nonlinear principal component analysis by neural networks: Theory and applications to the Lorenz system. *Journal of Climate* 2000; 13: 821–835.
5. Albanis GT, Batchelor RA. Assessing the Long-Term Credit Standing Using Dimensionality Reduction Techniques Based on Neural Networks-An Alternative to Overfitting. In the Proceedings of the SCI 99/ISAS 99 Conference, Orlando, US, 1999b.
6. Hsieh WW. Nonlinear principal component analysis by neural networks. *Tellus* 2001a; 53A: 599-615.
7. Sharma S. *Applied Multivariate Techniques*. New York: John Wiley & Sons; 1996.
8. SPSS INC. *SPSS Categories 10.0*. Chicago: SPSS inc.; 1999.
9. Gifi A. *Nonlinear Multivariate Analysis*. Chichester: Wiley; 1990.
10. Michailidis G, De Leeuw J. Multilevel Homogeneity analysis with differential weighting. *Computational Statistics & Data Analysis* 2000; 32(3-4): 411-442.
11. Principe JC, Euliano NR, Lefebvre WC. *Neural and Adaptive Systems: Fundamentals Through Systems*. New York: John Wiley & Sons; 2000.
12. Hassoun MH. *Fundamentals of Artificial Neural Networks*. Cambridge: Mass:MIT Pres; c1995.
13. Daszykowski M, Walczak B, Massart DL. A journey into low-dimensional spaces with autoassociative neural networks. *Talanta* 2003; 59: 1095-1105.
14. Fotheringham D, Baddeley R. Nonlinear principal components analysis of neuronal spike train data. *Biol Cybern.* 1997; 77(4): 283-288.

SEÇKİN TÜRK TIP DERGİLERİNDEKİ RANDOMİZE KONTROLLÜ ARAŞTIRMALARIN YÖNTEMBİLİMSEL KALİTESİ

Mustafa Şenocak, Necdet Süt, Ömer Uysal

İstanbul Üniversitesi, Cerrahpaşa Tıp Fakültesi, Biyoistatistik BD.

e-mail: msukrus@yahoo.com,

Özet

Araştırma tasarımlarının bilimsel kanıt özelliği sıralamasında Randomize kontrollü Araştırmalar (RKA) en yüksek kaliteye sahip kabul edilirler. Randomize kontrollü araştırma yayınları klinik kararları doğrudan etkileyen güçlü bilimsel kanıtlar sağlar ve sağlıksal girişimlerini kıyaslamada en doğru yöntemi oluştururlar. Bu düzeye ancak, uygun araştırma tasarımı, iyi seçilmiş yöntembilimsel özellikler ve uygulama ile erişilebilir ve sonuçlar güvenilir olarak ortaya çıkar. Bu çerçevede, çalışmanın özelliklerinin okuyucuya da uygun biçimde açıklanması, araştırmanın kurgusunu değerlendirmesi ve geçerliliğini yargılaması açısından önem taşır. CONSORT listesi de, RKA yayınlarının bilimsel kalite düzeyini yükseltmek, değerlendirmek ve denetlemek amacı ile oluşturulmuştur ve bu düzeydeki araştırmaların üretilmesinde kılavuzluk edebilmesinin yanı sıra, RKA'ların yöntembilimsel kalitesini yargılamak amacı ile de kullanılmaktadır.

Amacımız, EMBASE veya Index Medicus'da yer alan, CONSORT listesine uygunluk bildirmemiş, rastgele seçilmiş Türk Tıp dergilerindeki RKA'ların, liste maddelerine uyumlarını, gözlemsel bir araştırma ile saptamaktır. *Ulusal Travma Dergisi* (Index Medicus), *Anadolu Kardiyoloji Dergisi* (EMBASE, Index Medicus), ve *Turkish Journal of Pediatrics* (EMBASE, Index Medicus) dergilerindeki, 2002-2004 yıllarında yayınlanmış 14 randomize klinik araştırma, CONSORT listesinin 19 maddesine uygunlukları çerçevesinde değerlendirilmiştir.

Ortalama uygunluk puanı 8.9 / 19 olarak saptanırken, ancak bazı çok temel yöntembilimsel özelliklerle ilgili maddelerin yeterince yerine getirilmiş olduğu gözlenmiştir. Randomizasyon yönteminin açıklanması, Örneklem büyüklüğünün hesaplanarak yola çıkılması, sonuçlarda güven aralıklarının sunulması, çok değişkenli çözümlere yer verilmiş olması gibi son derece önemli yöntembilimsel özelliklerin ise çok zayıf olduğu saptanmıştır.

Sonuç olarak, en güvenilir olması gereken bilimsel yayınlarımızın bile ne yazık ki çok temel yöntembilimsel özellikleri sağlamakta sorunlu olduğu ortaya çıkmaktadır. Bu bağlamda, en azından temel yöntembilimsel özellikleri sağlama yaklaşımını sağlamlaştırana dek, dergilerimizin RKA'lar için CONSORT'a olabildiğince uyum zorunluluğunu öngörmeleri ve Biyoistatistik eğitimi içinde deney tasarımı ve yöntembilimine verilen ağırlığın artırılması önerilebilir.

Anahtar Kelimeler: Randomize Klinik Araştırma, CONSORT, Türk Tıp Dergileri, Yöntembilim, Bilimsel kalite.

Giriş

Randomize kontrollü araştırmalar (RKA), girişimli klinik çalışmalar içinde, en seçkin tasarıma sahiptirler (1) Yeni tedavi yöntemlerinin, hastalıkların doğal süreci, nedenleri ve risk faktörleri çerçevesinde değerlendirilmesinde klinisyenin tedavi kararlarında yüksek güvenilirlikte kanıt olarak işe yararlar. Tasarımları ve uygulamaları özen isteyen, kendilerine özgü yöntembilimsel özellikleri ve süreçleri olan çalışmalar oldukları için de tüm orijinal çalışmalar içinde ancak düşük bir oran oluşturabilirler. İngiliz çocuk cerrahları birliği 1996-2000 arasındaki çalışmalarında RKA oranını % 1 (2) bildirirken, yakın yıllarda, Çin İç hastalıkları dergilerinde bu oran % 10.1(3), Rus Üroloji dergilerinde % 0.08 (4), British journal of Surgery'de % 5 olarak bildirilmiştir (5).

Tüm araştırmalarda olduğu gibi RKA'lar da sistematik hatalara (bias) açıktırlar (6) ve tasarımlarının zayıflığı, yeterince açık olmayan veya uygun olarak sunulmamış yöntembilimsel özellikleri, kanıtların çelişkili olmasını ve kararlarda belirsizliği beraberinde getirir (7). Çalışmada "bilimsel kalite" sağlanarak sonuçların ve buna bağlanacak kararların güvenilir olması klinisyen okuru ilgilendirdiği kadar, benzer çalışmalar yürütecek olan araştırmacılara yol gösterme açısından da önemlidir. Bu çerçevede, yöntembilimsel güvenilirliği sağlamak böylece de sonuçların doğruluğunu oluşturabilmek yolundaki ilkelerin sağlanması ve ölçülmesi ile ilgili çok sayıda yöntem ortaya çıkmıştır. Yaklaşımlar, çalışmanın çeşitli özelliklerini puanlayan ölçeksel değerlendirmeler (skala) veya yine öngörülmuş olan özellikler listesinden hangilerinin var olduğu denetlenerek yapılan gözden geçirmeler şeklinde ortaya çıkar (onaylistesi). Daha 1995 de 25 ölçek (skala) ve 9 onaylistesi (checklist) bulunurken (8) günümüzde bu sayı çok daha artmıştır. Onaylisteleri hem yayın editörlerinin, yollanan çalışmaları denetiminde yardımcı olmaktadır hem de RKA tasarlayan, sürdüren, yayına hazırlayan araştırmacılar için bir çeşit kılavuzluk oluşturabilmektedirler.

Bu tür listelerin en geniş kapsamlısı ve en geniş kullanım alanına sahip olanı CONSORT (Consolidated Standarts for Reporting Trials:

Deneysel araştırma yayınlanmasında bütünleştirilmiş standartlar) listesidir (9) Oluşturulan ilk liste 30 maddelik olarak 1996 da yayınlandı (10). En geniş anlamda alınabilecek tüm özellikleri ile RKA ları değerlendirebilme olanağı sağlamak için oluşturulan bu yaklaşım tıp dergilerinin editörlerince de destek gördü ve önerilerek pek çok önemli derginin yayın kabul etme ölçütleri arasına girdi. Günümüzde içlerinde çok önemli dergilerin de bulunduğu 200 civarında tıp dergisi – tek Türk dergisi: Journal of Neurologic Sciences - bu liste maddelerine uymak koşulunu öngörmektedirler (9). Listenin daha yetkinleştirilmesi için süreç içinde çeşitli düzenlemelere de gidilerek 2001 de, 22 maddelik son liste kullanıma sunuldu (11). Yapılan araştırmalar ile CONSORT kullanımının RKA'ların bilimsel kalitesini arttırarak, hatalı sonuç değerlendirmelerini azalttığı ve tedavi etkinliği kararlarının alınmasında yararlı olduğu kanıtlanmıştır (12, 13, 14). Bununla birlikte CONSORT maddelerinin içerdiği ilkeler zaten bir bilimsel çalışmanın taşıması gereken temel özellikler olduğundan, listeye uyum öngörmeyen ama bilimsel değeri olan araştırmalar yayınlayan dergilerde de yüksek düzeyde sonuçlar gözlenmektedir (13) dolayısı ile liste geniş anlamda bir kalite değerlendiricisi olduğunu bu yönde de kanıtlamaktadır.

Ülkemizin bilim dünyası içindeki yerinin pek yükseklerde olmadığı bilinmektedir, atıf almış çalışmalar göz önünde bulundurularak yapılan değerlendirmede, 2003 de 22. sıradadır ancak nüfus büyüklüğü düzeltmesi ile 34. sıraya inmektedir (15). Bu yöndeki belirlemeler, SCI yayın sayısı, araştırmacı başına yayın sayısı, atıf sayısı, vs gibi özel ölçütlerle ortaya konabilir. RKA'lar araştırma yöntembiliminin en tipik özelliklerini taşıdıklarından bilimsel düzey açısından da oldukça iyi fikir verebilirler. Tüm araştırmalar içinde RKA'ların oranı gerçek bilimsel uğraşı eğiliminin düzeyini yansıtırken CONSORT değerlendirmesinden alınacak puan da bu uğraşın yöntembilimsel nitelik düzeyini gösterebilir.

Bu anlayış çerçevesinde, tıpsal araştırma düzeyimizin iyi bir sahnesi olarak düşünülebilecek olan üst düzey tıp dergilerimizde, RKA 'ların sıklığını ve yöntembilimsel kalite ölçütü olarak ortalama CONSORT puanlarını değerlendirdik. Kısa bir süreç olması nedeni ile (2001-2004), anlamlı bir değişim beklentisinde olmamakla birlikte, bu ölçütlerin son yıllardaki farklılaşımını da ek olarak gözlemledik.

Yöntem

EMBASE ve/veya INDEX MEDICUS'da bulunan 3 Türk tıp dergisi rastgele belirlenerek, içerdikleri RKA'ların, orijinal çalışmalar içindeki sıklığı, Bu RKA yayınlarının CONSORT listesi açısından kalite düzeyi, 2001-2002 yıllarındaki durumun, 2003-2004 dönemine değişimi açılarından irdelendi. Değerlendirme hem dergiler düzeyinde hem de toplamda gerçekleştirildi.

CONSORT Listesi

Çalışmamızda, 22 maddeli, gözden geçirilmiş 2001 CONSORT listesinin (11), 1-19. maddeleri kalite değerlendirmesi için kullanıldı, çalışmaların konusu olan alanla doğrudan ilgili yargılamalar içeren 20-22 maddeler değerlendirmemize alınmadı. Yargılamada maddelerin kavramsal yöntem bilim içeriği göz önünde bulunduruldu. Seçilen dergiler CONSORT'a uygunluk bildirmiş dergiler değildir. Listenin tüm maddeleri Tablo-4 de sunulmaktadır.

Dergiler ve RKA seçimi

Belirlenmiş olan 3 dergi'nin, Cerrahpaşa Tıp Fakültesi Kütüphanesinden ve/veya internet üzerinden, 2002-2004 yılları tüm sayılarına erişilerek değerlendirmeye alındı:

Ulusal Travma Dergisi, (Index Medicus)

Anadolu Kardiyoloji Dergisi, (EMBASE, Index Medicus)

Turkish Journal of Pediatrics, (EMBASE, Index Medicus, SCI), IF=0.18

Bu dergilerin 2002-2004 yılları sayılarındaki tüm "orijinal çalışmalar", başlık ve/veya özetlerinde "randomize", "rastgele", "randomizasyon", "rassal" terimleri bulunan tüm çalışmalar doğrudan erişilerek gözden geçirildi. İleri yönelik olmayan ve koşut kollu deney özelliği taşımayan çalışmalar değerlendirmeye alınmadı. Çalışmaların 4'ü ratlar (Ulusal Travma dergisi) üzerinde gerçekleştirilen deneysel araştırmalardır diğerleri ise farklı tedavi yöntemlerinin çeşitli sonuç özellikler açısından kıyaslandığı hasta çalışmalarıdır.

Dergilerin dönemlere göre araştırma sayıları ve RKA'ların dökümü aşağıda sunulmaktadır.

Tablo 1. Dergilerin, dönemlere göre ve toplamda orijinal çalışma, RKA sayıları ve oranları

Dergi	2001-2002			2003-2004			Toplam		
	Orj. Çalışma	RKA	%	Orj. Çalışma	RKA	%	Orj. Çalışma	RKA	%
Ulusal Travma	83	4	4.8	71	5	7.04	154	9	5.84
Anadolu Kardiyoloji	42	0	0	55	4	7.27	97	4	4.12
Turkish J Pediatrics	77	0	0	78	1	1.28	155	1	0.64
TOPLAM	202	4	1.98	204	10	4.90	606	14	2.31

Değerlendirme

Değerlendiriciler (MŞ, NS, ÖU) her yayını birbirlerinden “kör” olarak, her madde açısından değerlendirmişlerdir. Sonuç yargının Evet (1 puan) ve Hayır (0 puan) olarak belirlenmesi ilke olarak kabul edilmiştir. Ayrıca birden fazla kavramsal özellik içeren (ör.: Madde 4, 9, 10, 13, 17 gibi) maddelerde beklenenlerin belli eksikliklerle yerine getirilmiş olması durumu ile karşılaşılmış ve böylesi durumlarda “Eksik” (0.5 puan) yargısı kullanılmasına karar verilmiştir. Tüm yayınlar için bireysel değerlendirmeler bittikten sonra araştırmacılar hep birlikte yayınlara ilişkin yargıları sıra ile gözden geçirmişlerdir. Farklı yargıların gözlemlendiği çalışma/madde durumları için açık olarak tartışılmış ve çoğunluğun anlaştığı puan uygulanmıştır. Böylece her yayının her yargı maddesi açısından tek bir puanı bulunmaktadır.

Sonuç puanlar Ort. \pm SD olarak sunulmuştur, Toplam için değişim aralığı (R) verilmektedir. Tüm değerlendirmeler “19” üzerinden olduğu için, kolay anlaşılması çerçevesinde (%) başarı düzeyine de çevrilerek sunulmuştur.

Sonuçlar

Dönemsel değişimle ilgili değerlendirme için, 2 dergide 2001-2002 yıllarına ait, RKA olmaması nedeni ile toplam üzerinden irdeleme yapıldı. Buna göre 2001-02 aralığındaki sıklık ile 2003-04 aralığındaki sıklık arasında istatistiksel anlamlı fark bulunamadı (Fisher kesin olasılık testi: $p=0.17$) ancak bu sonucun genelde sıklıkların düşük olması ve örneklem hacminin yetersiz olması nedeni ile ortaya çıktığı düşünülmelidir çünkü ilk dönemde RKA içermeyen her iki dergide de sonrasında RKA'lar yayınlanmış, daha önce bulunduranda ise sayısal artış gözlemlenmiştir. Bu çerçevede yönelimin arttığına ilişkin ipucu akla getirilmelidir.

İrdelenen 19 madde toplamı açısından elde edilen ortalama puan düzeyleri, hem dergiler hem de toplam çerçevesinde aşağıda tablolanmaktadır, 2001-2002 yıllarına ait, 2 dergide RKA olmaması nedeni ile sağlıklı irdeleme yapılamayacağından toplam üzerinden değerlendirme yapılmaktadır.

Tablo 2. Dergilerde ve toplamda, yayınların CONSORT puan düzeyleri

	N	Ort.	SD	%	R
Ulusal Travma	9	8.6	0.79	45.3	7 – 9
Anadolu Kardiyoloji	4	8.5	0.41	44.7	8 – 9
Turkish J Pediatrics	1	12.5	-	65.8	12.5
TOPLAM	14	8.9	1.22	46.8	7 – 12.5

Yanıt tiplerinin puanları çerçevesinde her maddenin dergilerde ve toplamda almış olduğu puan düzeyleri aşağıda tablanmaktadır. Dergiler ve maddeler arası bir kıyaslama örnek sayılarının yetersizliği nedeni ile amaçlanmamıştır.

Tablo 3. CONSORT listesi maddelerinin puan düzeyleri

Madde (kısa açıklama)	U. Trv.	A. Krd.	TJP	Toplamda
1 Randomizasyon bildirilmesi	0.5	0.5	1	0,53
2 Gerekçe	1	1	1	1
3 Seçilme, koşullar, yer	0.67	1	1	0.78
4 Girişim ayrıntısı	1	1	1	1
5 Varsayım	0.5	0.63	1	0.57
6 Çıktı tipleri	0.89	1	1	0.93
7 Örneklem büyüklüğü hesabı	0	0	0	0
8 Randomizasyon yöntemi ayrıntısı	0	0	0	0
9 Randomizasyon Sıralama-gizleme	0	0	0	0
10 Randomizasyonu gerçekleştirenler	0	0	0	0
11 Körlük	0	0.13	0	0,03
12 İstatistiksel yöntemler	1	1	1	1
13 Sürece bağlı örneklem sayısı değişimi	0.39	0.38	1	0,43
14 Zamanlar	0.39	0.13	1	0,36
15 Öncül durum değerleri	0.78	0.5	0.5	0,68
16 Payda bildirim	1	0.5	1	0,86
17 Sonuçlar özet ve dağılım düzeyleri	0.28	0.5	0.5	0,36
18 Ek çözümlenmeler varlığı	0.11	0.13	0.5	0,14
19 Yan etkiler	0.17	0.13	1	0,21

Tartışma

En kaliteli tıp dergilerindeki RKA'larda bile CONSORT öncesi dönemde, önemli yöntemsel özellik olarak düşünülen koşullara bile uyumun çok düşük olabildiği saptanmıştır (16). Günümüzde, bilimsel kalite çok yükselmesine karşın, RKA'ların tümünde, CONSORT'a uymayı öngörsün veya öngörmesin, zaten tüm maddelere uygunluk gözlenmemektedir (13). Anlaşıldığı üzere hem maddeler boyutunda farklılıklar söz konusudur hem de kesin bir uygunluğa her durumda rastlanmamaktadır.

Bunun nedenlerinden biri çalışmanın özelliklerinin tüm maddelerin gerçekleşmesine doğası gereği uygun olmamasıdır ancak ne yazık ki bir diğeri bilimsel kalite yetersizliğidir çünkü CONSORT'a uyum öngörmeyen ancak yayınları açısından bilimsel kalite düzeyi ile öne çıkan dergilerde yöntem-bilimsel özelliklere özen gösterildiği saptanmaktadır (13).

İrdelediğimiz dergilerden 2 sindeki ortalama düzey böylesi çalışmalar için beklenenin oldukça altındadır (8.5, 8.6) ve böylece başarı oranı olarak da % 50'nin altında düzeyler gözlenmektedir (% 44.7, 45.3) Turkish Journal of Pediatrics'de tek yayın saptanabildiğinden kuramsal ortalamanın üzerinde gözükene düzey bile tartışmalıdır. Toplamda ise ancak 8.9 (19) luk bir ortalama düzeyle, orta düzey bir kalitenin bile altında kaldığımız söyle-nabilir.

Maddeler irdelendiğinde ise 19 maddeden sadece 9 unda ≥ 0.5 ortalama puana erişilebildiği gözlenmektedir (madde: 1-6, 12,15,16) . 3 derginin birden ≥ 0.5 değeri taşıdığı maddeler de aynı maddelerdir. Ancak bu maddeler yöntem-bilimsel açıdan varlığı tartışılmaz özelliktedir ve sağlanmamış olsalardı büyük olumsuzluk göstergesi olabileceklerken sağlanmış olmalarının aslında özel bir önem taşıdığı düşünülmemelidir. Diğer tüm maddeler ise ne yazık ki büyük yetersizlikler yansıtmaktadır: Hiçbir yayında, güç ve güvenilirliğe, beklentinin niceliğine dayalı bir örneklem hesabına rastlanmamaktadır (madde 7). Yine hiçbir yayında randomizasyon'un nasıl gerçekleştirildiğine ilişkin herhangi bir açıklama bulunmamaktadır (madde 8). Madde 9 ve 10'daki randomizasyon ayrıntısı ile ilgili özellikler zaten CONSORT uyumlu en önemli dergilerde bile ortalamanın altında kaldığı için (14) belki büyük eksiklikler olarak gözlenmemelidir ancak hem başlık hem de özetinde randomizasyonun vurgulanmış olduğu yayınların (madde 1) çok ender olması (1/ 14) dikkat çekicidir.

Körlük'le ilgili puan da yetersiz görülmektedir ancak hem çalışmaların bir kısmının deneysel olması hem de bu konunun da tam uyumlu özellik gösteremediğinin saptanmış olması (14, 17) zayıf bir açıklama olabilir. Sürecin zamansal ayrıntılarının verilmemesi ve süreç içindeki örneklem değişimlerinin dizgeli şekilde aktarılmaması günümüzdeki yöntem-bilim anlayışına göre oldukça büyük eksikliklerdir (madde 13, 14). Sonuçların nicel sunumlarında ise güven aralığı kullanımının yaygınlaşmamış olması dikkat çekmektedir (madde 17), istatistiksel açıdan gözlenen bir diğer eksiklik de çoklu değerlendirmelerin veya en azından bunların gerekliliğine ilişkin irdemelerin yapılmamış olmasıdır.

Sonuç olarak, bilimsel sorumluluk taşıyan yayınların, CONSORT listesine uyum bildirim olmaksızın, temel yöntem-bilimsel özellikleri zaten taşıdıkları bilinmesine karşın bizim seçkin yayınlarımızda, kısıtlı sayıdaki özellikler dışında, bilimsel kaliteyi yansıtacak düzeyin tutturulabildiği söylenememektedir. CONSORT listesindeki maddeler, bir yayının sonuçlarının

geçerli ve güvenilir olması açısından özellik taşıyan maddelerdir, eklere gereksinim olduğu ve/veya bazı maddelerin tam uyumunun, o çalışmanın özellikleri çerçevesinde, beklenmemesi gerektiği zaten bilinmektedir (17). Ancak bizim irdelememiz, yayınlarımızın ne yazık ki çok temel özellikleri bile sağlamakta sorunlu olduğunu ortaya koymaktadır. Bu bağlamda, en azından temel yöntembilimsel özellikleri sağlama yaklaşımını sağlamlaştıran dek, dergilerimizin RKA'lar için CONSORT'a (olabildiğince) uyum zorunluluğunu öngörmeleri önerilebilir ve Biyoistatistik eğitimi içinde deney tasarımı ve yöntembilimine verilen ağırlığın artırılması önerilebilir.

Tablo 4. CONSORT Listesi

- 1- Olguların girişim kümelerine dağıtımına ilişkin bilgilendirmenin ("randomizasyon", "randomize", "rastgele (rassal) atama" gibi terimler kullanılarak Başlık ve/veya, Özet'de bulunması.
- 2- Çalışmanın bilimsel tabanının ve gerekçesinin açıklanması.
- 3- Katılanların için seçilme ölçütleri ve veriler*in toplanma koşul ve yerlerinin açıklanması
- 4- Kümelerde girişimlerin ayrıntısının verilmesi, ne zaman ve nasıl uygulandıklarının bildirilmesi.
- 5- Özgün amaç(lar) ve varsayımların bildirilmesi.
- 6- Birincil ve ikincil sonuç çıktı tiplerinin açık olarak tanımlanması, gereğinde, ölçüm koşullarının iyileştirilmesi için alınmış yöntemlerin de açıklanması.
- 7- Örneklem Büyüklüğünün nasıl saptandığı ve gereğinde interim çözümlenme ve "çalışma durdurma" koşullarının açıklanması.
- 8- Randomizasyon düzeninin nasıl oluşturulduğunun açıklanması. Varsa kısıtlamalar, düzenlemelerle ilgili (Blokla, katmanlama gibi) ayrıntıların bildirilmesi.
- 9- Randomizasyon düzeni sıralamasının nasıl uygulandığı (ör.: numaralı kutucuklar, telefon sistemi) ve atama sonuçlarının girişimler başlayana dek gizli tutulması ile ilgili ayrıntıların açıklanması.
- 10- Randomizasyon düzenini üretenlerin, kayıtları yapanların, olguları kümelerine aktaranların bildirilmesi.
- 11- Olguların, girişimleri gerçekleştirenlerin, sonuçları ölçümleyen-kayıtlayanların "körlük" durumlarının açıklanması. Yapıldı ise körlenme başarısının değerlendirilmesi.
- 12- Sonuçların (ana ölçütler) değerlendirilmesi için kullanılan yöntemlerin açıklanması, alt küme çözümlenmeleri, düzeltme çözümlenmeleri gibi ek düzeltmelerin ayrıntılarının verilmesi.
- 13- Olguların, çalışma süreci içindeki sayılarının değişimi (akış diagramı yararlı olur.) Çalışmaya uygun olanların, atanmaların, çalışmayı tamamlayanların, sonuçları değerlendirilebilenlerin sayıları. Varsa çalışma protokolündeki süreç içi değişikliklerinin nedenleri ile açıklanması.
- 14- İzleme süresinin ve veri toplama zamanlamalarının tarihlerinin bildirilmesi.
- 15- Tüm kümelerin, demografik, klinik, temel-öncül özelliklerinin bildirilmesi.

16- Her değerlendirme için, her kümede değerlendirmeye katılan (payda) olgu sayısının bildirilmesi, değerlendirmenin "intention-to treat" tedavisi kararlaştırılmış olgular değerlendirmesi olup olmadığının bildirilmesi. (Uygun durumlar için sonuçların % olarak değil de pay/payda şeklinde verilmesi yeğlenmelidir.)

17- Her birincil ve ikincil sonuç için, her kümenin özetsel bilgisi, Etki büyüklüğü (*effect size*) kestirimlerinin ve kesinlik sınırlarının (ör.: % 95 CI) bildirilmesi.

18- Ana değerlendirmenin yanı sıra gerçekleştirilen alt-küme çözüm-lemelerinin ve/veya düzeltmeli değerlendirmelerin açıklanması. Bunların hangilerinin önceden kararlaştırılmış hangilerinin ayrıntıya inmek adına ortaya çıktığının bildirilmesi.

19- Tüm girişim kümelerindeki "Beklenmedik olayların" ve yan etki bildirimlerinin dökümü.

20- Sonuçların, başlangıç varsayım(lar)ı çerçevesinde, çalışmada yapılabilmemiş olan bias ve belirsizliklerin, çoklu değerlendirmelerin, ek çözümlerinin de göz önünde bulundurularak değerlendirilmesi

21- Çalışma sonuçlarının, "genellenebilirliği" (dışsal geçerlilik).

22- Geçerli kanıtlar çerçevesinde sonuçların genel yorumlaması.

Kaynakça

1. Hwiler-Muntener K., Juni P., Junker C., Egger M. Quality of reporting of randomized trials as a measure of methodologic quality, JAMA, 287: 2801-4, 2002.
2. Curry JI., Reeves B., Stringer MD. Randomized controlled Trials in pediatric surgery: could we do better?, J Pediatr Surg, 38 (4): 556-9, 2003.
3. Dong BR. Poster. Cochrane Colloquium. Randomized controlled Trials in the leading Chinese internal medical journals. Baltimore, Maryland, 1998.
4. Rafalsky V., Andreeva I., Riabkova E. Handsearching of Randomized Controlled Trials in Russian Urological Journals, <http://www.antibiotic.ru/en/pdfs/p079.pdf> :Erişim 14 Haziran 2005.
5. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomized Trials in surgery: problems and possible solutions. BMJ, 324: 1448-1451, 2002.
6. Schulz KF., Chalmers I., Hayes RJ. Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials, JAMA, 273(5):408-12, 1995.
7. Thakur A., Wang EC., Chiu TT., Chen W., Ko CY., Chang JT., Atkinson JB., Fonkalsrud EW. Methodology Standards Associated With Quality Reporting in Clinical Studies in Pediatric Surgery Journals.. Journal of Pediatric Surgery, 36 (8) :1160-1164, 2001.

8. Moher D., Jadad AR., Nichol G., Penman M., Tugwell P., Walsh S. Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists, *Controlled Clinical Trials*, 16:62-73, 1995.
9. CONSORT: <http://www.consort-statement.org>. Erişim 10 Mayıs 2005.
10. Begg C., Cho M., Eastwood S., et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement, *JAMA*, 276:637-9, 1996.
11. Moher D., Schulz KF., Altman D. for the CONSORT Group. The Consort Statement: Revised Recommendations For Improving The Quality Of Reports Of Parallel-Group Randomized Trials 2001, *EXPLORE*, 1(1):40-45, 2005.
12. Moher D., Jones A., Lepage L., for the CONSORT Group. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation, *JAMA*, 285:1992-1995, 2001.
13. Devereaux PJ., Manns BJ., Ghali WA., Quan H., Guyatt GH. The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the Consolidated Standards of Reporting Trials (CONSORT) checklist, *Controlled Clinical Trials*, 23:380-388, 2002.
14. Mills EJ., Wu TP., Gagnier J., Devereaux PJ. The quality of randomized trial reporting in leading medical journals since the revised CONSORT statement. Baskıda. *Controlled Clinical Trials*, 2005.
15. Ak MZ, Gülmez A. Atıf indekslerine göre Türkiye'nin bilimsel yayın performansının analizi. 3. Ulusal Bilgi, Ekonomi ve Yönetim kongresinde bildiri. Kasım 2004, Eskişehir.
16. DerSimonian R., Charetta LJ., McPeck B., Mosteller F. Reporting on methods in clinical trials, *J Pediatr Surg*, 38(4): 556-559, 2003.
17. Moher D, Altman DG, Schulz KF, Elbourne DR. Opportunities and challenges for improving the quality of reporting clinical research: CONSORT and beyond. *Canadian Medical Association. Journal.*: 2004; 171: 349-351.

SEÇKİN KANSER DERGİLERİNDEKİ RANDOMİZE KONTROLLÜ ARAŞTIRMALARIN “CONSORT” ÖLÇÜTLERİ İLE DEĞERLENDİRİLMESİ

Mustafa Şenocak, Ömer Uysal, Necdet Süt, Hilal Köksalan

İstanbul Üniversitesi, Cerrahpaşa Tıp Fakültesi, Biyoistatistik BD.

e-mail: msukrus@yahoo.com,

Özet

Araştırma tasarımlarının sıralamasında Randomize kontrollü Araştırmalar (RKA) en yüksek bilimsel kaliteye sahip kabul edilirler. Randomize kontrollü araştırma yayınları klinik kararları doğrudan etkileyen güçlü bilimsel kanıtlar sağlar ve sağlıksal girişimleri kıyaslamada en doğru yöntemi oluştururlar ve uygun tasarım, yöntembilimsel özellikler ve uygulama ile güvenilir sonuçlara erişilebilir. Bu çerçevede, çalışmanın özelliklerinin okuyucuya da uygun biçimde açıklanması, araştırmanın kurgusunu değerlendirmesi ve geçerliliğini yargılaması açısından çok önem taşır. CONSORT listesi de, RKA yayınlarının bilimsel kalite düzeyini yükseltmek, değerlendirmek ve denetlemek amacı ile oluşturulmuştur. İçerdiği maddelerin bazı tıp alanları için uygunsuz olmasının yanı sıra bazı alanlar için de özgün yeni noktalara gereksinim olduğu, bunu için de CONSORT'un sürekli güncellenmesi gerektiği saptanmıştır.

Amacımız, CONSORT listesine uygunluk bildirmemiş Kanser dergilerindeki RKA'ların, liste maddelerine uyumlarını, gözlemsel bir araştırma ile saptamaktır. CCR (Clinical Cancer Research) ve JNCI (Journal of National cancer Institute) dergilerinde, 2002-2004 yıllarında yayınlanmış 33 randomize klinik araştırma, CONSORT listesinin 19 maddesine uygunlukları çerçevesinde değerlendirilmiştir.

Ortalama uygunluk puanı 15.03 / 19 olarak saptanırken, bazı maddelerin öngördüğü özelliklerin tam olarak yerine getirilmiş olduğu da gözlenmiştir. “Randomizasyon yöntembilimi” ve “Körlük özellikleri” ile ilgili bazı anahtar özelliklerin ise çok yetersiz olarak yerine getirildiği dikkat çekmektedir.

Sonuç olarak, bilimsel sorumluluk taşıyan yayınların, CONSORT listesine uyum bildirimi olmaksızın, temel yöntembilimsel özellikleri zaten taşıdıkları gözlenmektedir. Farklı tıp alanlardaki RKA'lar için yeni özgün

soruların gereksiniminin yanı sıra, bazı maddelerin de tüm çalışmalar için gerekli gözükmemesi CONSORT benzeri listelerde, bir ana yapıya ek olarak, ilgili tıp alanına özgü modüler bölümlerin bulunduğu kurgular modellemenin gereğini düşündürmektedir.

Anahtar Kelimeler: Randomize Klinik Araştırma, CONSORT, Kanser, Yöntembilim.

Giriş

Yeni ilaçların ve diğer tedavi yöntemlerinin, hastalıkların doğal süreci, nedenleri ve risk faktörleri çerçevesinde değerlendirilmesi klinisyenin tedavi kararlarını doğru verebilmesi açısından son derece önemlidir. Bu yöndeki bilgiler ise güvenilir kanıtlar sunabilen çalışmalardan elde edilebilirler. Klinik çalışmaların tasarımlarının zayıflığı, yeterince açık olmayan veya uygun olarak sunulmamış yöntembilimsel özellikleri, kanıtların çelişkili olması ve kararlarda belirsizliği beraberinde getirir (1). Randomize kontrollü araştırmalar (RKA), girişimli klinik çalışmalar içinde, en seçkin tasarıma sahiptirler (2) ve özellikle tedaviye yönelik “Karar verme” durumlarında, üst düzey güvenilirlikte kanıt olarak işe yararlar. Ancak tüm araştırmalarda olduğu gibi RKA’lar da sistematik hatalara (bias) açıktırlar (3) ve bunların oluşmaması için yöntembilimsel ilkeler belirlenebilir; Randomizasyonun bozukluğu ve bu atama eşitsizliğinin oluşturduğu biaslar, kıyasal girişimler çerçevesinde oluşan kurgu eşitsizlikleri, sonuçların değerlendirilmesinde yapılan hatalar, izlemede yiten veya çıkartılan olguların irdelenmesinde uygunsuzluklar bunların en tipikleridir (2) ve sonuçların bilimsel “kalitesini” bozmaktadırlar.

Bu çerçevede yöntembilimsel güvenilirliği sağlamak böylece de sonuçların doğruluğunu oluşturabilmek yolundaki ilkelerin sağlanması ve ölçülmesi ile ilgili çok sayıda yöntem ortaya çıkmıştır. Yaklaşımlar, öngörülen çeşitli başlık, madde, alt-madde özelliklerine uygunluğa göre belli puanlar alınarak ortaya çıkan ölçüsel değerlendirmeler veya yine öngörülmuş olan özellikler listesinden hangilerinin var olduğu denetlenerek yapılan gözden geçirmelerdir (onaylistesi). Daha 1995 de 25 ölçek (skala) ve 9 onaylistesi (checklist) bulunurken (4) günümüzde bu sayının çok daha artmış olduğu açıktır. Onay listeleri hem yayın editörlere, yollanan çalışmaları denetiminde yardımcı olmaktadır hem de RKA tasarlayan, sürdüren, yayına hazırlayan araştırmacılar için bir çeşit kılavuzluk oluşturabilmektedirler.

Bu tür listelerin en geniş kapsamlısı ve en geniş kullanım alanına sahip olanı CONSORT (Consolidated Standarts for Reporting Trials: Deneysel araştırma yayınlanmasında bütünleştirilmiş standartlar) listesidir (5).

Listenin oluşturulması ile ilgili ilk çabalar 1990 ların ortasında başladı ve biyoistatistikçiler, epidemiologlar, tıp araştırmacıları, tıpsal dergilerin editörlerinden oluşan uluslar arası bir çalışma grubu tarafından hazırlanan 30 maddelik ilk liste yayınlandı (6). En geniş anlamda alınabilecek tüm özellikleri ile RKA ları değerlendirebilme olanağı sağlamak için oluşturulan bu yaklaşım tıp dergilerinin editörlerince de büyük destek gördü ve önerilerek pek çok önemli derginin yayın kabul etme ölçütleri arasına girdi. Günümüzde, içlerinde Lancet, NEJM gibi çok önemli dergilerin de bulunduğu 200 civarı tıp dergisi de bu liste maddelerine uymak koşulunu uygulamaktadırlar. Listenin daha güçlendirilmesi amacı ile süreç içinde çeşitli düzenlemelere de gidilerek 2001 de, 22 maddelik son liste kullanıma sunuldu (7). Yapılan araştırmalar ile CONSORT kullanımının RKA'ların bilimsel kalitesini arttırarak, hatalı sonuç değerlendirmelerini azalttığı ve tedavi etkinliği kararlarının alınmasında yararlı olduğu kanıtlanmıştır (8, 9, 10). Böylece liste, cerrahiden, anesteziye, psikiyatriye dek tıbbın çeşitli alanlarında geniş kullanım bulmuştur (11, 12, 13, 14). Bu duruma rağmen, listeye uymayı öngören dergilerde bile çeşitli yönlerden eksiklikleri olan yayınlar gözlenmiştir (15). Aslında bu durum çalışmaların yöntembilimsel açıdan yetersiz olmalarından ortaya çıkmak durumunda değildir, gerçekten de çeşitli tıp alanlarındaki randomize klinik çalışmalarda alanın özellikleri nedeni ile CONSORT koşullarına uyumsuz gözükecek veya o alan için çok önemli olmasına karşın CONSORT'un içermediği özellikleri gereksinen çalışma koşulları bulunabilir (16). Bu durum listenin hazırlayıcıları tarafından da belirlendiği için, sürekli bir yenileme, gereksinim saptama, gereğinde eskiden yapıldığı gibi madde eksiltme süreci ile listenin güncel tutulmaya çalışılması önlemleri gündemdedir (17).

Biz de bu çalışmamızda, onkoloji gibi, sonuçları açısından çok özel önem ve riskler taşıyan bir alanda yapılmış randomize klinik araştırmaları CONSORT ölçütleri açısından ele alarak, eksiklik gibi gözükebilecek olan halbuki belki de alanın koşullarına özgü durumları ortaya koymaya çalıştık. Amaç, bu yönde, çelişkili sonuçlar üreten, yöntembilimsel maddeleri ortaya koyarak, farklı alanların bilimsel eksiklik olarak değil kendilerine özgü nedenlerle değişik koşullar taşıyabileceklerini ortaya koymaktır. Yaklaşım, farklı alanlar için, CONSORT içerisinde özgün modüler yapılar kurgulanması gerektiğini ortaya koyacağı gibi, alanlara özgü ek maddelerin tasarımı içinde gelecekteki araştırmacılara yol gösterebilecektir.

YÖNTEM

Çok önemli iki kanser dergisindeki RKA'lar CONSORT listesi maddeleri açısından değerlendirildi, önem taşıyan ve daha az uyulmuş görünen maddelerin saptanmasına çalışıldı.

CONSORT Listesi

Değerlendirmemizde, 22 maddeli, gözden geçirilmiş 2001 CONSORT listesi kullanıldı (7) Listedeki 20, 21 ve 22 maddeler, yargılama açısından çalışmaların konusu olan alanla doğrudan ilgiyi zorunlu kılmaktadır – burada onkoloji- dolayısı ile bizim değerlendirmemize alınmamışlardır. Diğer bazı maddeler de alana özgü görünebilirler ancak yargılamada bu maddelerin kavramsal yöntem bilim içeriği göz önünde bulundurulmuştur. Biz de doğrudan araştırma yöntem bilimi ile ilgili bulunan ilk 19 maddeyi yargılamada kullandık. Seçilen dergiler özellikle CONSORT'a uygunluk bildirmemiş dergiler olarak seçildi. Böylece zorunluluk olmadan, kanser alanındaki randomize klinik araştırmaların özelliklerini saptama olanağı ortaya çıkacaktır. Listenin tüm maddeleri Tablo-1 de sunulmaktadır.

Dergiler ve RKA seçimi

İstanbul Üniversitesi, Onkoloji Enstitüsü kütüphanesinin abone bulunduğu ve 2002-2004 yılları tüm sayılarına erişilebilen, en yüksek IF (Impact Factor) düzeyine sahip ve salt kanserle ilgili çalışmaları yayınlayan 2 dergi seçildi. Bunlardan *Journal of National Cancer Institut*'ün IF değeri 13.84, *Clinical Cancer Research*'ün ki ise 6.51 dir. Dergilerin ikisi de CONSORT'a uygun RKA yayınlama yükümlülüğü bildirmiş değillerdir.

Bu dergilerin 2002-2004 yıllarına ait sayılarına ilişkin MEDLINE taraması, "randomized", "clinical trial" anahtar kelimeleri ile gerçekleştirildi ve ortaya çıkan listedeki tüm çalışmalar doğrudan dergiden erişilerek gözden geçirildi. Çalışmalardan, Faz II ve ötesi olan, tamamlanmış, ileri yönelik, kıyasal tedavi içerikli (tanımsal veya önlemsel özellikte olanlar dışındakiler) ve orijinal çalışma özelliğinde (kısa bildirim olanlar dışındakiler) olanlar değerlendirilmiştir. Bu çerçevede *Journal of National Cancer Institut* 'den 17, *Clinical Cancer Research* 'den 16 çalışma değerlendirmeye alınmıştır.

Tablo 1. CONSORT Listesi

- 1- Olguların girişim kümelerine dağıtımına ilişkin bilgilendirmenin ("randomizasyon", "randomize", "rastgele (rassal) atama" gibi terimler kullanılarak Başlık ve/veya, Özet' de bulunması.
- 2- Çalışmanın bilimsel tabanının ve gerekçesinin açıklanması.
- 3- Katılanların için seçilme ölçütleri ve verilerin toplanma koşul ve yerlerinin açıklanması
- 4- Kümelerde girişimlerin ayrıntısının verilmesi, ne zaman ve nasıl uygulandıklarının bildirilmesi.
- 5- Özgün amaç(lar) ve varsayımların bildirilmesi.
- 6- Birincil ve ikincil sonuç çıktı tiplerinin açık olarak tanımlanması, gereğinde, ölçüm koşullarının iyileştirilmesi için alınmış yöntemlerin de açıklanması.

7- Örneklem Büyüklüğünün nasıl saptandığı ve gereğinde interim çözümlenme ve “çalışma durdurma” koşullarının açıklanması.

8- Randomizasyon düzeninin nasıl oluşturulduğunun açıklanması. Varsa kısıtlamalar, düzenlemelerle ilgili (Blokla, katmanlama gibi) ayrıntıların bildirilmesi.

9- Randomizasyon düzeni sıralamasının nasıl uygulandığı (ör.: numaralı kutucuklar, telefon sistemi) ve atama sonuçlarının girişimler başlayana dek gizli tutulması ile ilgili ayrıntıların açıklanması.

10- Randomizasyon düzenini üretenlerin, kayıtları yapanların, olguları kümelerine aktaranların bildirilmesi.

11- Olguların, girişimleri gerçekleştirenlerin, sonuçları ölçümleyen-kayıtlayanların “körlük” durumlarının açıklanması. Yapıldı ise körlük başarısının değerlendirilmesi.

12- Sonuçların (ana ölçütler) değerlendirilmesi için kullanılan yöntemlerin açıklanması, alt küme çözümlenmeleri, düzeltme çözümlenmeleri gibi ek düzeltmelerin ayrıntılarının verilmesi.

13- Olguların, çalışma süreci içindeki sayılarının değişimi (akış diagramı yararlı olur.) Çalışmaya uygun olanların, atanmaların, çalışmayı tamamlayanların, sonuçları değerlendirilebilenlerin sayıları. Varsa çalışma protokolündeki süreç içi değişikliklerinin nedenleri ile açıklanması.

14- İzleme süresinin ve veri toplama zamanlamalarının tarihlerinin bildirilmesi.

15- Tüm kümelerin, demografik, klinik, temel-öncül özelliklerinin bildirilmesi.

16- Her değerlendirme için, her kümede değerlendirmeye katılan (payda) olgu sayısının bildirilmesi, değerlendirmenin “intention-to treat” tedavisi kararlaştırılmış olgular değerlendirilmesi olup olmadığının bildirilmesi. (Uygun durumlar için sonuçların % olarak değil de pay/payda şeklinde verilmesi yeğlenmelidir.)

17- Her birincil ve ikincil sonuç için, her kümenin özet bilgisi, Etki büyüklüğü (*effect size*) kestirimlerinin ve kesinlik sınırlarının (ör.: % 95 CI) bildirilmesi.

18- Ana değerlendirmenin yanı sıra gerçekleştirilen alt-küme çözümlenmelerinin ve/veya düzeltmeli değerlendirmelerin açıklanması. Bunların hangilerinin önceden kararlaştırılmış hangilerinin ayrıntıya inmek adına ortaya çıktığının bildirilmesi.

19- Tüm girişim kümelerindeki “Beklenmedik olayların” ve yan etki bildirimlerinin dökümü.

20- Sonuçların, başlangıç varsayım(lar)ı çerçevesinde, çalışmada yapılabilmemiş olan bias ve belirsizliklerin, çoklu değerlendirmelerin, ek çözümlenmelerin de göz önünde bulundurulması değerlendirilmesi

21- Çalışma sonuçlarının, “genellenebilirliği” (dışsal geçerlilik).

22- Geçerli kanıtlar çerçevesinde sonuçların genel yorumlanması.

Değerlendirme

Değerlendiriciler (MŞ, ÖU, NS) her yayını birbirlerinden “kör” olarak, her madde açısından değerlendirmişlerdir. Sonuç yargının Evet (1 puan) ve Hayır (0 puan) olarak belirlenmesi ilke olarak kabul edilmiştir ancak benzer

çalışmalarda rastlandığı üzere (7) bazı maddelerin uygulanamaz (çalışmanın koşullarına uygun değil) olduğu durumlarla karşılaşılmıştır. Ayrıca birden fazla kavramsal özellik içeren (ör.: Madde 4, 9, 10, 13, 17 gibi) maddelerde beklenenlerin belli eksikliklerle yerine getirilmiş olması ile de karşılaşılabilmektedir. Böylesi durumlarda “Eksik” (0.5 puan) yargısı kullanılmasına karar verilmiştir. Tüm yayınlar için bireysel değerlendirmeler bittikten sonra araştırmacılar hep birlikte yayınlara ilişkin yargıları sıra ile gözden geçirmişlerdir. Farklı yargıların gözlemlendiği çalışma/madde durumları için açık olarak tartışılmış ve çoğunluğun anlaştığı puan uygulanmıştır. Böylece her yayının her yargı maddesi açısından tek bir puanı bulunmaktadır. 34 yayın toplamında, her madde için hangi yargı sonuçlarının bulunduğu 19 X 4'lük bir tablo olarak belirlenmiştir.

Sonuçlar

Değerlendirilen 19 madde toplamı açısından elde edilen ortalama düzeyler aşağıda tablanmaktadır; JNCI'de iki yayında yan etki değerlendirmesi koşullara bağlı olarak yapılamadığından, bunlarda puanlama “18” üzerinden gerçekleştirilmektedir.. Ortak değerlendirmeye yatkınlık açısından alınan puanlar, alınabilecek puanların – genelde 19- yüzdesi üzerinden de sunulmuştur.

Tablo 2. Dergilerde ve toplamda, yayınların CONSORT puan düzeyleri

DERGİ	Ort.	SD	Değişim Aralığı	% uygunluk ort.
CCR	13.75	1.82	11.5-16.5	72.37
JNCI *	16.24	1.85	12.5-19	85.89
Toplamda	15.03	2.20	11.5-19	79.33

Çalışmanın amacı dergilerin düzeyini kıyaslamak olmamakla birlikte JNCI'deki yayınların düzeyi CCR'ye göre istatistiksel açıdan anlamlı olarak yüksek bulunmaktadır ($p < 0.001$).

Farklılaşma özellikle bazı maddelerin puanlaması çerçevesinde ortaya çıkmaktadır, bu yöndeki irdelenmeyi gerçekleştirmek için öncelikle 33 yayın toplamında, her maddenin hangi tip puanları aldığı irdelenmiştir.

Tablo 3 de bu dağılımın genel durumu ve tip yanıtların yayınlar geneli çerçevesindeki dağılımları sunulmaktadır. Maddenin özellik duru-

munun uygun olmadığı koşul sadece yan etkileri ilgilendiren 19. başlıkta görülmüştür.

Tablo 3. CONSORT listesi maddelerinde yanıt tiplerinin dağılımı

Madde	Evet		Hayır		Eksik		Uygun değil	
	N	%	n	%	n	%	n	%
1	27	81.82	0	0.00	6	18.18	0	0.00
2	33	100.00	0	0.00	0	0.00	0	0.00
3	33	100.00	0	0.00	0	0.00	0	0.00
4	33	100.00	0	0.00	0	0.00	0	0.00
5	30	90.91	0	0.00	3	9.09	0	0.00
6	32	96.97	0	0.00	1	3.03	0	0.00
7	23	69.70	7	21.21	3	9.09	0	0.00
8	14	42.42	10	30.30	9	27.27	0	0.00
9	6	18.18	21	63.64	6	18.18	0	0.00
10	1	3.03	29	87.88	3	9.09	0	0.00
11	9	27.27	22	66.67	2	6.06	0	0.00
12	32	96.97	0	0.00	1	3.03	0	0.00
13	24	72.73	0	0.00	9	27.27	0	0.00
14	24	72.73	2	6.06	7	21.21	0	0.00
15	33	100.00	0	0.00	0	0.00	0	0.00
16	33	100.00	0	0.00	0	0.00	0	0.00
17	26	78.79	0	0.00	7	21.21	0	0.00
18	25	75.76	2	6.06	6	18.18	0	0.00
19	25	75.76	3	9.09	3	9.09	2	6.06

Yanıt tiplerinin puanları çerçevesinde her maddenin dergilerde ve toplamda almış olduğu puan düzeyleri aşağıda tablanmaktadır. Bu yönden, dergiler veya maddeler arası bir kıyaslama amaçlanmamakla birlikte dikkat çekici farklılıklar vurgulanacaktır.

Genelde 7., 8., 9., 10. ve 11. başlıklar dışında başarılı sonuçlar gözlenmektedir. Dergiler arasındaki farklar 8.,9. ve 11. maddelerde özellikle ortaya çıkmaktadır. 10. madde en az yeterli yanıtı sahip gözükmemektedir.

Tablo 4. CONSORT listesi maddelerinin puan düzeyleri

Madde (kısa açıklama)	CCR	JNCI	Toplamda
1 Randomizasyon bildirilmesi	0,88	0,94	0,91
2 Gerekçe	1	1	1
3 Seçilme, koşullar, yer	1	1	1
4 Girişim ayrıntısı	1	1	1
5 Varsayım	0,91	1	0,96
6 Çıktı tipleri	1	0,97	0,99
7 Örneklem büyüklüğü hesabı	0,62	0,85	0,74
8 Randomizasyon yöntemi ayrıntısı	0,34	0,76	0,56
9 Randomizasyon Sıralama-gizleme	0,09	0,44	0,27
10 Randomizasyonu gerçekleştirenler	0,03	0,12	0,08
11 Körlük	0,19	0,41	0,30
12 İstatistiksel yöntemler	0,97	1	0,99
13 Sürece bağlı örneklem sayısı değişimi	0,72	1	0,86
14 Zamanlar	0,66	1	0,83
15 Öncül durum değerleri	1	1	1
16 Payda belirtilmesi	1	1	1
17 Sonuçlar özet ve dağılım düzeyleri	0,78	1	0,89
18 Ek çözümlenmeler varlığı	0,72	0,97	0,85
19 Yan etkiler	0,84	0,87	0,86

Tartışma

En kaliteli tıp dergilerindeki RKA'larda bile (NEJM, Lancet, BJM), CONSORT öncesi (1980) dönemde, önemli yöntembilimsel özellik olarak düşünülen koşullara bile uyumun çok düşük olabildiği saptanmıştır; örneğin randomizasyon yönteminin açıklanmasına % 19, örneklem hesabı ile ilgili olarak "güç" irdelenmesine % 12, çalışmaya alınma koşullarının açıklanmasına % 37 yayında rastlandığı bildirilmektedir (18). Günümüzde, bilimsel kalite çok yükselmesine karşın, RKA'ların tümünde, CONSORT'a uymayı öngörsün veya öngörmesin, zaten tüm maddelere uygunluk gözlenmekte değildir (9). Ancak bazı maddeler açısından dergi tipleri arasında açık farklar gözlenebilirken bazı diğerlerinde çalışma tipinin temel yöntembilimsel özelliği olması nedeni ile farklılık bulunmamaktadır. Örneğin randomizasyon gizliliği (madde 9) CONSORT'a uyan yayınlarda % 57, diğerlerinde % 35 olarak bildirilmişken, kıyas kümelerinin öncül durumlarının irdelenmesi (madde 15)

her iki dergi tipinde % 90'ın üzerindedir. Randomizasyonun oluşturulma yönteminin açıklanması ise CONSORT dergilerinde % 76 iken diğerlerinde anlamlı olarak düşük ($p < 0.001$) olarak % 45 bulunmuştur. Veri değerlendiricinin körlüğü gibi bir özgün ayrıntıya ise uyum bildiren dergilerde % 8 bildirmeyenlerde % 0 oranlarında rastlanmıştır (9). Anlaşıldığı üzere hem maddeler boyutunda farklılıklar söz konusudur hem de kesin bir uygunluğa her durumda rastlanmamaktadır. Bunun nedenlerinden biri bilimsel kalite yetersizliği olsa da bir diğeri çalışmanın özelliklerinin tüm maddelerin gerçekleşmesine doğası gereği uygun olmamasıdır. Bizim çalışmamızda da bu yönden bazı maddeler dikkat çekmektedir. Randomizasyonu gerçekleştirenlerin (madde 10) açıklanması bizde sadece 1 yayında (% 3) tam olarak bulunmaktadır ve bu madde en düşük puana sahiptir. Obstetrik alanında gerçekleştirilmiş RKA'ların irdelendiği bir araştırmada aynı maddenin uygunluğu % 10 ile yine çok düşük olarak saptanmıştır. Aynı çalışma örneklem sayısı hesaplanması için % 61 oranını verirken bizde tam yanıt oranı % 70 dir. (19). Çalışmamızda tam bildirim % 42 olarak saptanan randomizasyon yönteminin açıklanması özelliği, içinde NEJM'in de bulunduğu – çalışma yayınlarının da % 50 sini içererek – CONSORT uyumu bildirmeyen dergilerde % 45 olarak bildirilmiştir (9).

Bu maddeler (7., 8.) yöntembilimsel açıdan tartışılmaz özelliktedir ve daha özen gösterilmesinin gereği anlaşılmaktadır. Aynı şekilde randomizasyon sonuçlarının süreç içinde gizli tutulması yaklaşımının düşük uyum taşınması (tam yanıt %18.2), varlığı doğal gözükken özelliğın durumunun bildiriminde eksiklik olabileceğini düşündürmektedir. “Körlük” gibi RKA için çok yaşamsal olarak düşünölebilecek bir madde, tam yanıt oranı açısından çalışmamızda % 27 düzeyindedir. İlk bakışta RKA özellikleri ile çelişkili gözükken bu durumun, maddenin içeriği çerçevesinde tam olarak sağlanması aslında zor olabilmekte hatta konu çerçevesinde olanaksız olabilmektedir. Gerçekten de oftalmolojik RKA'ları konu eden bir çalışmada sonuçları kaydedenlerin ve değerlendirenlerin körlük oranı ancak % 4.5 oranındadır (19). Obstetrik çalışmasında olguların körlüğü bile % 66'dır (19). Buradan anlaşılacağı gibi körlüğü, olgu, tedavi uygulayıcı, sonuç saptayıcı üçlü boyutunda her seferinde sağlayabilmek söz konusu değildir; bu durumun gerçekliği, ilk CONSORT listesinde yer alan (6) ve körlüğün üçlü denetimini 3 ayrı madde ile yapan yaklaşımın yeni listede tek soruda toplanması ile de gözlenmektedir.

Yan etkiler ise RKA olmasına karşın, bazı çalışmalarda, tedavi özellikleri açısından (genellikle cerrahi tabanlı) veya çalışmanın önlemsel, tanımsal konuları nedeni ile doğal olarak bulunamayabilmektedir.

Diğer tüm maddeler RKA'ların zaten temel özelliklerini yansıtmaktadırlar ve nerede ise tam olarak yerine getirilme çabası çalışmamızdaki yayınlarda da yansımaktadır.

Sonuç olarak, bilimsel sorumluluk taşıyan yayınların, CONSORT listesine uyum bildiri olmaksızın, temel yöntem bilimsel özellikleri zaten taşıdıkları gözlenmektedir. Liste, bu durumu gerek yazarların gerek editörlerin çalışmalarını denetlemesi açısından yarar sağlamaktadır ancak bir zorunluluk listesi olarak uygulanması söz konusu değildir. Farklı alanlardaki RKA'lar için yeni özgün soruların gereğinin yanı sıra, bazı maddelerin de tüm çalışmalar için gerekli gözükmemesi CONSORT benzeri listelerde bir ana yapıya ek olarak, ilgili tıp alanına özgü modüler bölümlerin bulunduğu kurguların daha verimli olabileceğini düşündürmektedir.

Kaynakça

1. Thakur A., Wang EC., Chiu TT., Chen W., Ko CY., Chang JT., Atkinson JB., Fonkalsrud EW. Methodology Standards Associated With Quality Reporting in Clinical Studies in Pediatric Surgery Journals, *Journal of Pediatric Surgery*, 36(8):1160-1164, 2001.
2. Huwiler-Muntener K., Juni P, Junker C., Egger M. Quality of reporting of randomized trials as a measure of methodologic quality, *JAMA*, 287:2801–2804, 2002.
3. Schulz KF., Chalmers I., Hayes RJ. Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials, *JAMA*, 273(5):408-412, 1995.
4. Moher D., Jadad AR., Nichol G., Penman M., Tugwell P., Walsh S. Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists, *Controlled Clinical Trials*, 16:62-73, 1995.
5. CONSORT: <http://www.consort-statement.org>. Erişim 10 Mayıs 2005.
6. Begg C., Cho M., Eastwood S. et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement, *JAMA*, 276:637–639, 1996.
7. Moher D., Schulz KF., Altman D. for the CONSORT Group. The Consort Statement: Revised Recommendations For Improving The Quality Of Reports Of Parallel-Group Randomized Trials 2001, *EXPLORE*, 1(1):40-45, 2005.
8. Moher D., Jones A., Lepage L., for the CONSORT Group. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation, *JAMA*, 285:1992-1995, 2001.
9. Devereaux PJ., Manns BJ., Ghali WA., Quan H., Guyatt GH. The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the

Consolidated Standards of Reporting Trials (CONSORT) checklist, *Controlled Clinical Trials*, 23:380–388, 2002.

10. Mills EJ., Wu TP., Gagnier J., Devereaux PJ. The quality of randomized trial reporting in leading medical journals since the revised CONSORT statement. *Baskıda. Controlled Clinical Trials*, 2005.
11. Sauerland S., Davis TRC. The Consolidated Standards O Reporting Trials (Consort): Better Presentation Of Surgical Trials In The Journal Of Hand Surgery, *Journal of Hand Surgery (British and European Volume)* 29B(6): 621–624, 2004.
12. Halpern SH., Darani R., Douglas MJ., Wight W., Yee J. Compliance with the CONSORT checklist in obstetric anaesthesia randomised controlled trials. *International Journal of Obstetric Anesthesia*, 13:207 – 214, 2004.
13. Grimes DA. The “CONSORT” Guidelines for Randomized Controlled Trials in Obstetrics & Gynecology, *Obstetrics & Gynecology*, 100(4): 631-632, 2002.
14. Khan A., Khan SR., Leventhal RM., Krishnan KRR., Gorman JM. An Application of the Revised CONSORT Standards to FDA Summary Reports of Recently Approved Antidepressants and Antipsychotics, *BIOL PSYCHIATRY*, 52:62–67, 2002.
15. Mills E., Wu P., Gagnier J., Heels-Ansdell D., Montori VM. An analysis of general medical and specialist journals that endorse CONSORT found that reporting was not enforced consistently. *Journal of Clinical Epidemiology*, 58:662–667, 2005.
16. Ioannidis JPA., Evans SJW., Gøtzsche PC., O'Neill RT. et al Better Reporting of Harms in Randomized Trials: An Extension of the CONSORT Statement, *Annals of Internal Medicine*, 141(10);781-789, 2004.
17. Moher D., Altman DG., Schulz KF., Elbourne DR. Opportunities and challenges for improving the quality of reporting clinical research: CONSORT and beyond. *Canadian Medical Association Journal*, 171: 349-351, 2004.
18. DerSimonian R., Charetta LJ., McPeck B., Mosteller F. Reporting on methods in clinical trials, *J Pediatr Surg*, 38(4):556-559, 2003.
19. Sanchez-Thorin JC., Cortes MJ., Montenegro M., Villate N. The Quality of Reporting of Randomized Clinical Trials Published in Ophthalmology, *Ophthalmology*, 108:410-415, 2001.

KLİNİK DENEMELERDE HATA KAYNAKLARI VE YANLILIK TIPLERİ

Mehmet Ali Sungur, Emine Arzu Kanık

Mersin Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Mersin

E-mail: malisungur@mersin.edu.tr

Özet

Bu çalışmada klinik denemelerde ortaya çıkan, denemenin geçerliliğini etkileyen ve çalışmanın yanlış bulgularla sonuçlanmasına neden olabilecek hata kaynaklarından söz edilmektedir. Bir araştırma planlanırken belki de en önemli sorun bu araştırmanın geçerli bulgular üretebilecek olup olmadığıdır. Denemeleri tek kör, çift kör olarak planlamak ve uygulamaları deneklere rasgele dağıtmak gibi teknikler geçerliliği artırabilirler ancak garanti edemezler. Ayrıca planlanan çalışma için bu gibi teknikler uygun olmayabilir ya da her zaman uygulama imkanı olmayabilir. Yani geçerliliğin nasıl sağlanabileceği ve denemenin nasıl düzenleneceği konusunda daima her koşula uygun olan bir yöntem bulunmamaktadır. Dolayısıyla tüm araştırmalarda olduğu gibi klinik denemelerde de çalışmanın geçerliliğini tehdit eden ve yanlış bulgularla sonuçlanmasına neden olabilecek hata kaynakları söz konusudur. Bu hata kaynakları rasgele hata (random error) ve sistematik hata (systematic error) olmak üzere 2 gruba ayrılır. Rasgele hata şansa karakterize edilir ve örneklem genişliğine bağlı olarak değişebilir. Rasgele hata genellikle doğrudan sonucu etkilememekte, parametre tahmininde güven aralığının genişlemesi gibi olumsuzluklara neden olmaktadır. Sistematik hata şansa açıklanamaz ve örneklem genişliğine bağlı değildir. Sistematik hata genellikle yanlılıkla (bias) açıklanır. Yanlılık, denemenin herhangi bir aşamasında ortaya çıkabilen ve verileri gerçekten saptırarak doğrudan çalışma sonucunun yanlış bulunmasına neden olan sistematik hatalardan kaynaklanır. Literatürde rasgele hata ve sistematik hataya ek olarak etki karışımı (confounding) da bazen üçüncü bir hata kaynağı olarak yer almaktadır. Etki karışımı yanlılığa benzerdir ve genellikle de karıştırılır. Bazen yanlılık tipleri arasında da sayılmaktadır. Ancak yanlılık bir denemenin planlanmasındaki veya bir değişkenin ölçümündeki hatayı içerirken etki karışımı doğru planlanmış bir denemenin veya doğru olabilecek bir ölçümün istatistik modellemede yanlış kullanılmasından dolayı sonuçların hatalı olmasına neden olur.

Çok çeşitli yanlılık tipleri vardır ve bunlar genel olarak 2 grupta sınıflandırılmaktadır: Selection Bias ve Information/Observation Bias. Sackett (1979), klinik bir denemeyi yanlılığın meydana gelebileceği 7 aşamaya bölmüş ve bu aşamaların her biri için yanlılıkların detaylı bir listesini oluşturarak 55 tip yanlılık tanımlamıştır. Buna ek olarak Spilker (1991) 6 tip yanlılık tanımlamıştır. Daha sonra Dorak (2005), Sackett'in sınıflandırmasını esas alarak yaptığı sınıflandırmada, klinik bir denemeyi yanlılığın meydana gelebileceği 6 aşamaya bölerek 103 tanesi alt tip olmak üzere toplam 121 yanlılık tipi tanımlamıştır.

Bu çalışmada literatürde karşılaşılan tüm yanlılık tipleri listelenmiş ve bazıları ilk kez Türkçe olarak isimlendirilmiştir. Bu isimlendirmeler tartışmaya açılmıştır. Ayrıca yanlılık tipleri sıklıkla karşılaşıldığı istatistik analizlere ve araştırma konularına göre gruplandırılmıştır. Özellikle meta analizi çalışmalarında karşılaşılan publication bias ve tanı testlerinin performanslarının değerlendirilmesinde önemli rol oynayan verification bias ve work-up bias detaylı olarak araştırılmıştır.

Anahtar Kelimeler: Rasgele Hata, Sistemik Hata, Yanlılık, Yanlılık Tipleri, Etki Karışımı.

KLİNİK DENEYLERDE GRUPLARIN BELİRLENMESİNDE “MİNİMİZASYON“ METODU

Özden Calay, Sevim Purisa¹, Mustafa Şenocak¹

İ.Ü. Cerrahpaşa Tıp Fakültesi Temel Bilimler Bölümü Biyoistatistik Bilim
Dalı¹

e-mail: purisa@istanbul.edu.tr

Özet

Randomize kontrollü deneyler, sağlık ile ilgili araştırmalarda altın standart araştırma metodu olarak kabul edilir. Temel olarak katılımcılar tedavi ve kontrol gruplarına rassal olarak dağıtılır ki bu da tarafsızlıkların olmasını önler. Çoğu klinik deneylerde her bir tedavi grubunda sadece eşit sayıda hasta olmasının yanı sıra benzer prognostik faktörlerin de (yaş, hastalığın evresi gibi) olması arzu edilir. Basit randomizasyonda özellikle daha büyük denemelerde iyi dengelenmiş gruplar oluşturulabilir. Fakat her zaman için bir dengesizlik riski de söz konusudur. Buna önlem olarak başka randomizasyon yöntemleri oluşturulmuştur. Katmanlı randomizasyon, prognostik faktörlerin sayısında artış olduğunda işe yarayabilir. Nedeni ise katman sayısının artması deneydeki hasta sayısının da buna bağlı olarak artmasını gerektirmesindedir. Minimizasyon ise; katmanlamanın karmaşık ileri bir biçimidir. Katmanlama da sonuçlar üzerinde anlamlı bir etki de bulunduğu inanan sadece bir veya birkaç faktör bulunurken, minimizasyon da ise tedavinin sonucuna önemli bir etkide bulunma olasılığı olan tüm önemli faktörler bulunur. Bu tekniğin katmanlamada olduğu gibi amacı klinik denemenin sonucu bakımından iki veya daha fazla tedavi grubu oluşturulurken aralarındaki farkları en aza indirmektir.

Bu derlememizde “Minimizasyon” metodu ile bilgisayar simülasyonu kullanılarak klinik deneylerde sayıca yeterli ve karakteristik özelliklerine göre daha dengeli tedavi grupları elde edilebildiğini, ayrıca basit ve katmanlı randomizasyona göre avantajlarını ve dezavantajlarını ortaya koymaya çalıştık.

Sonuçta minimizasyon metodunun klinik deneylerde tedavi gruplarının belirlenmesinde özellikle de çok fazla prognostik faktör olduğunda kullanılmasının daha uygun olacağını düşündük. Bu metod ilk olarak Taves, Pocock ve Simon tarafından ortaya çıkarılmıştır. Minimizasyon her gruptaki hastaların sayısını önceden belirlenen prognostik faktörler açısından dengeli

bir şekilde tedavi gruplarına dağıtmayı amaçlamaktadır. Çeşitli yayınlardan elde edilen sonuçlara göre simülasyon çalışmaları 'minimizasyon' metodunun sınırlı veya sınırsız (unrestricted) randomizasyonla mukayese edildiğinde daha dengeli tedavi grupları oluşmasını sağladığını göstermiştir. Minimizasyon kullanımı ile ilgili endişeler daha çok tedavi gruplarına alınacak hastaların önceden tahmin edilebilmesi ve kullanılan analiz metotları üzerinde toplanmaktadır. Klinik deneylerin analizinde minimizasyon metodu kullanıldığında deney düzeneğinin minimizasyon faktörlerine uygun olarak düzenlenmesi önerilmektedir. Minimizasyon kullanımı diğer metotlarla karşılaştırıldığında bazen organizasyonel güçlükler sebep olabilir. Buna rağmen minimizasyonun birçok yorumcu tarafından klinik deneylerde kullanılması tavsiye edilmekte, fakat seyrek olarak kullanılmaktadır. Bu derlemede sunulan ipuçlarına dayanarak minimizasyon metodunun oldukça etkili bir belirleme metodu olduğu görülmekte ve randomize kontrollü deneylerde kullanılması önerilmektedir.

Anahtar Kelimeler: Minimizasyon, randomizasyon, klinik deneyler.

RANDOMİZASYONDA KATMANLAMANNIN YERİ: BİR BENZETİM ÇALIŞMASI DOLAYISI İLE DEĞERLENDİRME

Mustafa Şenocak, Ömer Uysal, Necdet Süt

İstanbul Üniversitesi, Cerrahpaşa Tıp Fakültesi, Biyoistatistik BD.

e-mail: msukrus@yahoo.com,

Özet

Rastgele dağıtım (randomizasyon) yaklaşımı özellikle ileri yönelik klinik çalışmalar için, farklı tıpsal girişimlerin etkilerini hatasız ortaya koyabilmek açısından en önemli temel özelliklerden biri olarak kabul edilir. Yaklaşımın temeli, değerlendirilen tedavi tipi dışındaki, konu edinilen tıpsal sonuç açısından etkili olabilecek başlangıç özelliklerinin, yargılama kümelerine eşdüzeyle dağılımını sağlamaktır. Bu amaçla da temeli rassallığa dayanan ama yöntem olarak özgün tekniklerle geliştirilmiş yaklaşımlar uzun yıllardan beri üretilmiştir. Bunlar arasında en güvenilir sonuçları verebilecek yaklaşım olarak "Katmanlı örnekleme" tasarlanmıştır. Yöntem çalışma olgularının, sonucu etkilediği düşünülen demografik, klinik özellikler açısından "katmanlar" boyutunda ele alınmasını ve her katman içinde çalışmanın küme kollarına rastgele atamalar gerçekleştirilmesini esas alır.

Yöntemin bilimsel yararları ile ilgili görüşlerin yanı sıra, özellikle yüksek sayılı örneklemlerde gereksizliğini savunan yayınlar da tasarımın ortaya çıkışından beri gözlenmiştir.

Bu tartışmalara katkıda bulunmak amacı ile benzetim (simülasyon) tabanlı bir araştırma gerçekleştirdik ve sanal bir çalışma için, bunun sonuçlarını etkileyebileceği düşünülen belli katmanlar içeren bir toplumdan, katmanlama yapılmaksızın çekilen rastgele örneklemlerin sonuç dengelerini, benzetimler üreterek gözlemlemeyi amaçladık.

Bir nicel, bir binomial ve bir 3 altkümeli özellikli taşıyan, 2000 olguluk sanal bir toplum üretilmiştir. Bu önkabulsel yapı ve rastgele atama özellikleri çerçevesinde hazırlanmış olan varsayımsal toplumdan, 20, 50 ve 100 er olguluk rastgele örneklem kümeleri çekilmiş ve işlem 100 benzetim dizisi ile yinelenmiştir. Benzetimler sonunda, üretilmiş olan 20, 50 ve 100 er olgu içerikli örneklem kümelerinin, katman özellik değişkenleri açısından ikili kıyaslamalarında, kaç değerlendirmede ve hangi kritik düzeylerde istatistiksel anlamlılıkla karşılaşıldığı saptanmış, değişken kombinasyonlarındaki anlam-

lılık dağılımları irdelenmiştir. Binomial özellikteki sıklık özelliğinin sonuçlar üzerine etkisi de sorgulanmıştır.

Sonuçlarımız göre, değişkenler tekil olarak ele alındığında, kıyas kümelerinde tatminkar dengelerle karşılaşılmaktadır ve bu düzen gerek küme örneklem hacmine gerekse de benzetim sayısına bağlı gözükmemektedir. Buna karşılık konu edilen katmanların çeşitli kombinasyonlarının aynı anda dengeli geldiği örneklem durumlarına daha kısıtlı sayıda rastlanmıştır. Dolayısı ile rastgele örnekleme, katman sayısı arttıkça, kıyasal kümelerin en azından bir katmanda dengesiz ortaya çıktığı gözlenmiştir. Bir diğer olumsuzluk gözlemi de, örneklem kümeleri arasında fark bulunmakla birlikte, her kümedeki durumun aslında toplumsal düzeyden farklı olabileceğidir.

Sonuç olarak basit randomize örneklemenin, kıyas kollarındaki dengeyi sağlamakta kısıtlamalar taşıdığı anlaşılmaktadır. İlk çözümsel seçenek olarak ortaya çıkan katmanlı randomize yaklaşım, uygun koşullarda gerçekleştirildiğinde güvenilir sonuçların alındığı, Tip I ve II hatayı azaltmakta yetkin bir örnekleme yöntemi olarak klinik araştırma tasarımında yerini korumaktadır.

Anahtar Kelimeler: Randomizasyon, Katman, Örnekleme, Benzetim – simülasyon, Katmanlı örnekleme.

Giriş

Randomize – rassallanmış, rastgele – dağıtım yaklaşımı özellikle ileri yönelik klinik çalışmalar için, farklı tıpsal girişimlerin etkilerini hatasız ortaya koyabilmek açısından standart olarak kabul edilir. Yaklaşımın temeli, değerlendirilen tedavi tipi dışındaki, konu edinilen tıpsal sonuç açısından etkili olabilecek temel özelliklerin, yargılama kümelerine eşdüzeyleli dağılımını sağlamaktır (1), böylece randomizasyon “bilinen ve bilinmeyen risk öğeleri açısından çalışma kümelerini kıyaslanır duruma getirmek, olguların kümelere dağıtımında araştırmacı bias’ını gidermek ve istatistiksel yargılamaların geçerli anlamlılık düzeyleri taşımalarını sağlamak yönünde gerçekleştirilir “(2).

Randomizasyonun “basit” anlamda gerçekleştirilmesi veya yanlış, eksik yaklaşımlarla gerçekleştirilmesi - psödorandomizasyon - önemli sorunların ortaya çıkmasını engelleyemez: Kıyas kümelerine farklı sayıda olgu atanabilir, tedavi sürecini etkileyebilecek kritik özellikler açısından denge sağlanamayabilir, sonuç istatistiksel yargılamalarda düzey kaymaları oluşabilir (1, 2, 3, 4). Bu sakıncaların olabildiğince giderilmesi amacı ile de temeli rassallığa dayanan ama yöntem olarak özgün tekniklerle geliştirilmiş yaklaşımlar uzun yıllardan beri üretilmiştir. Genel yaklaşım olarak; *Sabit atamalı* ve *Uyarlamalı (adaptiv)* atamalı randomizasyon süreçleri ortaya

çıkılmıştır (2). Blok, küme, oranlı kollu, yanıt uyumsal randomizasyon, minimizasyon, bunlara örnektir (2, 3, 4, 5, 6, 7).

Bu çerçevede en güvenilir sonuçları verebilecek yaklaşım olarak “Katmanlı örnekleme” tasarlanmıştır. Yöntem iki aşamalı bir süreçtir ve ilkinde, çalışma olguları, sonucu etkilediği düşünülen demografik, klinik özellikleri açısından “katmanlar” olarak ele alınır. Her katman 2 veya daha fazla alt-özellik taşır. İkinci aşamada ise katman olguları içinden çalışmanın küme kollarına rastgele atamalar gerçekleştirilir (1, 2). Bu aşamada özellikle blok randomizasyon gibi diğer yaklaşımların da katkısından yararlanılır (1, 2).

Yöntemin bilimsel yararları ile ilgili görüşlerin yanı sıra (8), özellikle yüksek sayılı örneklemlerde gereksizliğini savunan yayınlar tasarımın ortaya çıkışından beri gözlenmiştir (9). Katman içi dengelerin sağlanmasına karşın, çalışma bütününde bunun sağlanamayabileceği (10), katmanlamanın istatistiksel testlerin gücüne etkili olduğu (11), özellikle de katman sayısının çok kısıtlı tutulması yönündeki yayınlar (12, 13), yaklaşımı her zaman güncel ve tartışılır tutmaktadır.

Bu tartışmalar doğrultusunda biz de, sanal bir çalışma için, bu araştırmanın sonuçlarını etkileyebileceği düşünülen belli katmanlar içeren bir toplumdan, katmanlama yapılmaksızın çekilen rastgele örneklemlerin sonuç dengelerini, benzetimler üreterek gözlemlemeyi amaçladık. Özellikle kısıtlı sayıda örneklem içeren ($n < 200$) kümelerle ilgili durumsal dağılımlarda dengesizlik sıklıkları ortaya konmaya çalışıldı.

YÖNTEM

Sanal toplum olgularının üretimi

Yaklaşım, 45 yaş ve üzerini tutan ve 65 yaş üzerinde yoğunlaşan, cinsiyet açısından farklı dağılım özellikleri içerebileceği düşünülen, 65 yaş üzeri nüfus oranının doğal olarak kadınlarda daha yoğun gözleneceği ve üç altkümeye dağılan bağımsız bir nitel özelliğin de sonucu etkileyebileceği bir sanal hastalık toplumu öngörmektedir. Toplam olarak hasta toplumunda 2000 olgu varsayılmış ve öncelikle eşit oranlarda kadın/erkek içermek üzere cinsiyet açısından kodlanmışlardır.

Olguların çalışmaya kabul ölçütü olarak 45-80 yaşları arasında olacağı ve Kadınlarda; 45-64 yaş aralığında: % 20, 65-80 yaş aralığında: % 80 olgu, Erkeklerde ise 45-64 yaş aralığında: % 25, 65-80 yaş aralığında : % 75 olgu dağılımını göstereceği varsayılmıştır böylece toplamda 65 yaş altı olgu oranı % 22.5 olmaktadır. Bu kabullere göre kadınlar ve erkekler için, uygun aralıklara uyan olgu sayıları kadar rastgele “yaş” tamsayı olarak

üretilek atanmıştır. Atama sonucunda Erkekler (n=1000): 67.96 ± 9.08 , Kadınlar (n=1000): 69.13 ± 8.45 yaşındadırlar ve beklendiği üzere cinsiyetler arasında anlamlı fark bulunmaktadır ($p=0.003$).

Varsayımsal niteliğin sıralı özellikte 3 alt kümesi olacağı ve toplamda % 10, 30, 60 oranlarında dağılacığı öngörülmüştür. Uygun sırayla 1, 2, 3 olarak kodlanan ve gerekli sayı kadar sıralanan olgu dizimine 1-2000 arası sıra sayıları rastgele atanarak yaş ve cinsiyetten bağımsız bir birleştirme sağlanmıştır. Atama sonucunda özelliğin 3 alt kümesinin erkeklerde dağılımı sırası ile: % 9.8, 29.3, 60.9 ve kadınlarda % 10.2, 30.7, 59.1 olarak ortaya çıkmıştır. Koşul olarak öngörülmemesine karşın bu dağılım cinsiyetler arasında anlamlı fark bulunmayacak şekilde ($p=0.7129$) belirmektedir.

45-64 yaş (n=450) ve 65-80 yaş (n=1550) kümelerine dağılım açısından ise 3. özel niteliğin altkümeler dağılımı sırası ile % 6.7, 30.9, 62.4 e karşın % 11, 29.7, 59.3 olarak ve anlamlı fark göstererek ortaya çıkmıştır ($p=0.027$).

Binomial değişkenlikte toplumsal oranın randomizasyondaki durumunu irdelemek üzere, eşit oranlarda düzenlenmiş olan erkek/kadın benzetiminin yanı sıra 2 değişken durumu daha tasarlanmıştır. Bunların ilkinde % 5 diğ erinde % 20 sıklıkta oluşturulan iki durum bulunacaktır.

Benzetimlerin üretimi

3 (+2) özelliğin kurgulandığı bu önkabulsel yapı ve rastgele atama özellikleri çerçevesinde hazırlanmış olan 2000 olguluk toplumd an, 20, 50 ve 100 er olguluk rastgele örneklem kümeleri çekilmiş ve işlem 100 benzetim dizisi ile yinelenmiştir. Bu amaçla bizim dışımızda bir kaynaktan rastgele üretim için yararlanılmıştır (<http://www.assumption.edu/users/avadum/applets/applets.html>). Buradan edinilen rastgele sayı dizileri ile toplum içinden sıra sayıları elde edilerek, bu sırada yer alan olgu, özellikleri ile örneklem kümesinde yer almıştır.

Her denemenin yeni bir üretimle bağımsız olması yerine 2000 olgu toplamının öngörülen örneklem sayısı kadar bağımlı parçalara ayrılması, farklılıkları belirgin açığa çıkartma açısından uygun görülmüştür. 20 şer olguluk kümeler için, her denemede birbirlerinden tamamen farklı 50 çift örneklem kümesi sağlanabilmektedir ve bu özellikte 4 dizi üretilmiştir. Bu çerçevede 50 şer olguluk örneklem için 10 dizi ve 100 olguluk örneklem için 20 dizi üretilmiştir. Böylece her tip örneklem içeriği için gerekenin 2 katı küme türetilmiş olmaktadır. Bunların ilk ve son çeyreklik kısımlarında bulunan ve izleyen kümeler olarak ortaya çıkmış örneklem değerlendirmeye alınmış ve cinsiyetlere dağılım, yaş düzeyi, özgün nitelik altkümelerine dağılım açlarından, 2 küme kıyaslaması olarak irdelenmişlerdir. Binomial

dağılımların durumunu irdelemek amacı ile oluşturulan % 5 ve % 20 sıklıkla rastgele oluşturulmuş nitelikler için de aynı süreçler uygulanmıştır.

Biyoistatistiksel değerlendirme

İrdelenen ilk özellik olan cinsiyetin dağılımı iki kıyas kümesinde beklenen değerlerin düzeyine göre Fisher kesin olasılık testi veya ki-kare testi ile değerlendirilmiştir. Yaşların kıyaslanması için varyans eşdüzenliliği gözetilerek Student t testi kullanılmıştır. 3 altküme içeren özelliğin kıyaslanmasında ise yine olgu sayılarının dağılımına göre ki-kare veya Kolmogorov-Smirnov testi kullanılmıştır.

Yargılama çift yönlü olarak düşünülmüş ve anlamlılık sınırı olarak $\alpha = 0.05$ alınmıştır. Elde edilen $0.05 < p < 0.10$ düzeyindeki sonuçlar da saptanmış ve tek yönlü yargılamada anlamlı çıkan sonuçların da belirtilmesi sağlanmıştır.

Sonuçlar

100'er benzetim sonunda, üretilmiş olan 20, 50 ve 100'er olgu içerikli örneklem kümelerinin, katman özellik değişkenleri açısından ikili kıyaslarında, kaç değerlendirmede – ve hangi düzeyde – anlamlılıkla karşılaşıldığı aşağıdaki tablolarda sunulmaktadır.

Tablo 1. Yaş özelliği açısından benzetim kümelerinde kıyaslama sonuçlarının durumu

Küme Olgu Sayısı	Sonuç Anlam Düzeyi	Benzetim (N)		
		25	50	100
20	$p < 0.05$ (n, %)	1 (4)	1 (2)	5 (5)
	$p < 0.10$ (n, %)	3 (12)	5 (10)	10 (10)
50	$p < 0.05$ (n, %)	1 (4)	2 (4)	5 (5)
	$p < 0.10$ (n, %)	2 (8)	4 (8)	9 (9)
100	$p < 0.05$ (n, %)	1 (4)	1 (2)	5 (5)
	$p < 0.10$ (n, %)	2 (8)	3 (6)	10 (10)

Nicel özellik açısından, benzetimlerde, her 3 örneklem hacim tipinde de, $p < 0.05$ düzeyinde anlamlılık gösteren durum sıklığının % 5'i aşmaması dikkat çekicidir.

Tablo 2. 3 alt özellikli nitelik açısından benzetim kümelerinde kıyaslama sonuçlarının durumu

Küme Olgu Sayısı	Sonuç Anlam Düzeyi	Benzetim (N)		
		25	50	100
20	p < 0.05 (n, %)	1 (4)	2 (4)	4 (4)
	p < 0.10 (n, %)	2 (8)	4 (8)	9 (9)
50	p < 0.05 (n, %)	3 (12)	6 (12)	10 (10)
	p < 0.10 (n, %)	5 (20)	9 (18)	19 (19)
100	p < 0.05 (n, %)	0 (0)	0 (0)	4 (4)
	p < 0.10 (n, %)	0 (0)	2 (4)	9 (9)

Bu nitel özellik açısından, benzetimlerde, 20 ve 100'er örneklem hacmine sahip kümelerin kıyaslamasında, p<0.05 düzeyinde anlamlılık gösteren durum sıklığı % 5'i aşmamışken 50 şer olgulu kümelerin kıyasında dengesiz durumların sıklığı % 10'u geçmektedir.

Binomial değişken tasarımlarımızdan ilki olan ve % 50-50 oluşturulan niteliğin yanı sıra % 5 ve % 20 sıklıkla tasarlanmış değişkenlere ait sonuçlar aşağıdaki tabloda sunulmaktadır.

Tablo 3. Belli sıklıktaki (%x) Binomial nitelik(ler) açısından benzetim kümelerinde kıyaslama sonuçlarının durumu

Küme Olgu Sayısı	Sonuç Anlam Düzeyi	Benzetim (N)								
		25			50			100		
		% 50	% 5	% 20	% 50	% 5	% 20	% 50	% 5	% 20
20	p < 0.05 (n, %)	0 (0)	0 (0)	1 (4)	1 (2)	0 (0)	1 (2)	1 (1)	0 (0)	3 (3)
	p < 0.10 (n, %)	1 (4)	0 (0)	1 (4)	2 (4)	0 (0)	1 (2)	4 (4)	0 (0)	5 (5)
50	p < 0.05 (n, %)	1 (4)	0 (0)	0 (0)	3 (6)	0 (0)	3 (6)	5 (5)	2 (2)	6 (6)
	p < 0.10 (n, %)	1 (4)	0 (0)	2 (4)	3 (6)	0 (0)	6 (12)	5 (5)	3 (3)	10 (10)
100	p < 0.05 (n, %)	1 (4)	1 (2)	0 (0)	3 (6)	2 (4)	0 (0)	6 (6)	3 (3)	3 (4)
	p < 0.10 (n, %)	1 (4)	2 (4)	1 (4)	5 (10)	4 (8)	1 (2)	11 (11)	7 (4)	7 (7)

Farklı sıklıktaki binomial özellikler açısından, benzetimlerde, her 3 örneklem hacim tipinde de, p<0.05 düzeyinde anlamlılık gösteren durum sıklığının az da olsa yine 50 olgulu kümelerde gözlenmesi dikkat çekmektedir.

Bir diğer önemli konu da yukarıda ki değerlendirmelerin her değişken için bağımsız olarak ele alınmış olmasıdır. Sanal toplumumuzda cinsiyet dağılımı ile yaş dağılımını belli ölçüde bağıntılı olarak kurgulamamıza karşın her koşulda yaş+cinsiyet dengesi kombinasyonunun karşılanamayacağı açıktır. 3. değişkenimiz ise zaten yaş kümeleri açısından bağımsız olarak belirmiştir ve tüm değişkenlerin birlikte dengeli olacağı koşulların ortaya çıkma olasılığı daha da düşmektedir. Gerçekten de farklı küme hacimleri için oluşturulan benzetimlerde, sadece bir, 2 ve her 3 değişken açısından anlamlı farkların bulunduğu ($p<0.05$) durumlarla, aşağıdaki tabloda sunulduğu üzere karşılaşılmaktadır.

Tablo 4. Benzetim kümelerinde konu değişken sayısına göre anlamlı sonuç alınan durumların dağılımı

Küme Olgu Sayısı	Benzetim (N)								
	25			50			100		
	Tek	İki	Üç	Tek	İki	Üç	Tek	İki	Üç
20	2 (8)	0 (0)	0 (0)	3 (6)	0 (0)	0 (0)	9 (9)	0 (0)	0 (0)
50	5 (20)	0 (0)	0 (0)	12 (24)	0 (0)	0 (0)	20 (20)	0 (0)	0 (0)
100	2 (8)	0 (0)	0 (0)	4 (8)	0 (0)	0 (0)	13 (13)	1 (1)	0 (0)

Görüldüğü üzere bu koşullarda dengesiz oluşan kümelerin sayısı birikimli olarak artmış bulunmaktadır. 3 değişkenin birden anlamlı fark gösterdiği hiçbir duruma rastlanmamıştır. – Tek bir örnek dışında 2 özelliğin birden farklı bulunduğu duruma da rastlanmamış sayılabilir.

Tartışma

Günümüz karmaşık tedavi koşullarında, pek çok klinik araştırma ancak kısıtlı sayıda hasta temini ile yapılabilmektedir üstelik genellikle sonuca etkisi bilinen oldukça çok sayıda karıştırıcı-prognostik etken de bulunabilmektedir. Dolayısı ile yüksek örneklem sayısı ile dengelenmesi umularak tamamen basit randomizasyona bırakılacak yaklaşımlar hem pek olanaklı değildir hem de güvenli değildir. 3 özellikli benzetimimizde de, değişkenler tekil olarak ele alındığında oldukça tatminkar dengelerle karşılaşılmaktadır ve bu düzen gerek küme örneklem hacmine gerekse de benzetim sayısına bağlı gözükmemektedir. Buna karşılık konu edilen katmanların çeşitli kombinasyonlarının aynı anda dengeli geldiği örneklem

durumlarına çok daha kısıtlı sayıda rastlanmıştır. Dolayısı ile rastgele örnekleme, katman sayısı arttıkça, kıyasal kümelerin en azından bir katmanda dengesiz ortaya çıkmaları kesinlikle beklenmelidir.

Bir diğer nokta da örneklem kümeleri arasında fark bulunmamakla birlikte, her kümedeki durumun aslında toplumsal düzeyden farklı olabileceğidir. Basit randomizasyonu irdeleyen bir çalışma 10000 olguluk kuramsal bir toplumdaki % 15 sıklıktaki bir özelliğin, 50 olguluk örneklemelerin % 24 ünde, 100 olgulukların ise % 10'unda, toplumsal sıklığa göre en az % 10 fark olduğunu göstermiştir (1). Bizim çalışmamızda da, erkek/kadın eş oranını yargılayan ve kümeler arasında anlamlı fark bulunmamasına karşın, % 50 ye göre en az % 10 fark gösteren en az bir kümenin bulunduğu durum sayıları aşağıda verilmiştir.

Tablo 5. % 50 binomial değişkenin, kıyas kümeleri arasında fark bulunmamakla birlikte, en az bir kümede de $> \% 60$ veya $< \% 40$ oranlarında gözlemlendiği durumlar

Küme Olgu Sayısı	Anlamlı Olmayan Aşırı Değerde Durum Sayısı	Anlamlı Olmayan Durumlar İçinde Sıklık (%)
20	44	44,4
50	10	10,5
100	3	3,2

Bu durum, aslında gerçek toplumdan farklılığın yansıması olup kümeler arasında fark bulunmasa da toplumsal uygulamadaki sonuçların çalışma sonuçlarından farklı olabileceğinin göstergesidir. Örneklem hacminin artması bu yöndeki dengesizliği oldukça ortadan kaldırma eğilimi oluşturmaktadır.

Elde ettiğimiz gözlemlere göre basit rastgele örnekleme dengelerin sağlanması açısından sorunlar taşımaktadır ve daha güvenilir dengeleyici bir çözüm gerekmektedir. Katman kotalarının önceden belirlenerek, olgu akışı içinde bu değerlere uygun, basit veya blok rastgele örnekler alınması en verimli çözüm olarak görülse bile gerçek uygulamada sorunlarla karşılaşmaktadır (1, 2). Bunlardan en basiti beklenen olgunun bulunamaması nedeni ile ortaya çıkan eksikliklerdir, ayrıca çok sayıda katmanın kombinasyonu ile ortaya çıkacak pek çok özgün alt küme için olgu bulunabilmesi sorunu da ortadadır (1, 2, 3). Minimizasyon gibi uyarılma atamalı yaklaşımlar etkin olarak böylesi sorunları gidermek amacı ile kurgulanmışlardır ve denge sağlamadaki güvenilirlikleri tartışılmıştır (1, 5, 7, 10, 12). Bu bağlamda ortaya çıkan temel sorun kullanılacak katman sayısı olarak da belirlemektedir (1, 13).

Bir benzetim araştırması, 2-3 alt özellikli katmanlar içeren çalışmalarda, katmanlar kombinasyon sayısı, örneklem sayısının yarısı düzeyine erişince, $(n/2)$, herhangi bir tekil değişken açısından bile dengeleme olasılığının % 50 ye düştüğünü saptamıştır (13), buna karşılık minimizasyon ile 10-20 farklı katmana kadar güvenilir sonuçlar alındığı bildirilmektedir. Küçük hacimli çalışmalarda ($n < 200$), 3'den fazla prognostik değişken için bile minimizasyon katmanlı rastgele örneklemeden daha güvenilir sonuçlar üretmiştir (14). Ancak minimizasyonun da kendine özgü değerlendirme sorunları olması gözden kaçırılmamalıdır.

Katmanlamaya karşı bir diğer seçenek çözüm de çokdeğişkenli çözümlere (post-stratifikasyon) gidilmesidir. 100 den fazla toplam olgu içeren iki küme tabanlı çalışmalarda katmanlama ile, basit rassalamalı çok değişkenli çözümlerin eşdeğer güvenilirlikte sonuçlar verdiği saptanmıştır (15). Bu yaklaşımda, yargılamaya katılacak özelliklerin fazlası ile araştırmacının isteğine kalması ve bunların belirlenmesindeki ölçütlerin esnekliği kuşku oluşturan noktalardır.

Tüm bunlar, koşulların uygun olması durumunda katmanlı yaklaşımı öne çıkarmaktadır; bu yöndeki bir kılavuz temel ilkeler olarak şu değerlendirmeleri bildirmektedir (1):

- İnterim veya alt-küme çözümlerinin de protokolde yer aldığı, tedavi tip kümesi başına 200 olgudan az örneklem düşen ve bir tedavinin diğerine göre üstünlüğünü kanıtlamayı amaçlayan çalışmalarda katmanlı yaklaşıma gidilmelidir.
- İkincil çözümlerin öngörülmediği ve tedavi tip kümesi başına 200 den fazla olgunun düştüğü çalışmalarda katmanlı yaklaşıma gerek görülmemektedir.
- Tedavi tipi eşitliklerinin irdeleneceği çalışmalarda örneklem sayısına bakılmaksızın katmanlı yaklaşım uygulanmalıdır.
- Katman özellikleri olarak sonuç açısından geniş farklılıklar oluşturabileceği kanıtlı değişkenler seçilmelidir ve bunların sayısı kısıtlı tutulmalıdır.

Sonuç olarak basit randomize örnekleminin, kıyas kollarındaki dengeyi sağlamakta kısıtlamalar taşıdığı anlaşılmaktadır. İlk çözümsel seçenek olarak ortaya çıkan katmanlı randomize yaklaşım, uygun koşullarda gerçekleştirildiğinde güvenilir sonuçların alındığı, Tip I ve II hatayı azaltmakta yetkin bir örnekleme yöntemi olarak klinik araştırma tasarımında yerini korumaktadır. Amaçsal özelliğe, kıyas kol örneklem hacimlerine göre, minimizasyon izleyen seçenek yöntemdir. Yine de bu özgün ve uygulanma dizgesi gerektiren çözümlerin, tıpsal açıdan tüm uygulama koşullarına uyumunun sağlanması doğal olarak olanaklı değildir – acil girişim koşullarının yer aldığı ve anlık hasta özelliklerinin rassal seçime göre öncelik taşıdığı

alanlar-, böylesi durumlar için de çok değişkenli istatistiksel çözümleme yaklaşımları kullanılacaktır.

Kaynakça

1. Walter N. Kernan, Catherine M. Viscoli, Robert W. Makuch, Lawrence M. Brass, Ralph I. Horwitz . Stratified Randomization for Clinical Trials, J Clin Epidemiol, 52(1):19–26, 1999.
2. Friedman L.M., Furberg C.,D., DeMets D.L. Fundamentals of Clinical Trials. Springer-Verlag, New York, 1998.
3. Bridgman S., Dainty K., Kirkley A., Maffulli N. Practical Aspects of Randomization and Blinding in Randomized Clinical Trials, Arthroscopy, 19(9):1000-1006, 2003.
4. Colditz GA., Miller JN., Mosteller F. How study design affects outcomes in comparisons of therapy, Statistics in Medicine, 8(4):441-454, 1989.
5. Scott NW., McPherson GC., Ramsay CR., Campbell MK. The method of minimization for allocation to clinical trials:a review, Controlled Clinical Trials, 23:662–674, 2002.
6. Zelen M. The randomization and stratification Of patients to clinical trials, J Chron Dis 27:365-375, 1974.
7. Taves DR. Minimization: A new method of Assigning Patients to Treatment and control groups. Clin Pharmacol Ther, 15:443-453, 1974.
8. Begg CB., Iglewicz B. A treatment Allocation Procedure for Sequential Clinical Trials, Biometrics, 36:81-90, 1980.
9. Pike PR., Armitage P., Breslow NE., Cox DR., Howard SV., Mantel N., McPherson K., Peto J., Smith PG. Design and analysis of Randomized Trials requiring prolonged observations of each patient, British journal of Cancer, 34:585-612, 1976.
10. Hallstrom A., Davis K. Imbalance in treatment assignments in stratified blocked randomization, Controlled Clinical Trials, 9(4): 375-382, 1988.
11. Sylvan B. Green and David P. Byar. The effect of stratified randomization on size and power of statistical tests in clinical trials, Journal of Chronic Diseases, 31(6-7):445-454, 1978.
12. Amery WK., Dony J. The use and limitations of a stratified randomization in surgical adjuvant studies of cancer, European Journal of Cancer, 15(1):1-5, 1979.
13. Terry M. Therneau. How many stratification factors are "too many" to use in a randomization plan? Controlled Clinical Trials, 14(2):98-108, 1993.

14. Pocock S.J., Simon R. Sequential Treatment assignment with balancing for prognostic factors in the controlled clinical trial, *Biometrics*, 31:103-115, 1975.
15. McHugh R., Matts J. Post-stratification in the randomized clinical trial, *Biometrics*, 39:217-225, 1983.

RANDOMİZE KONTROLLÜ KLİNİK DENEMELERDE TEDAVİ ETKİNLİĞİNİN SAPTANMASINDA NNT'NİN (NUMBER NEEDED TO TREAT) KULLANIMI: GÜVEN ARALIĞI SORUNU

E. Arzu Kanık, Bahar Taşdelen

*Mersin Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Mersin
e-mail:arzukanik@mersin.edu.tr*

Özet

Bu çalışmanın konusunu, randomize kontrollü klinik denemelerde tedavinin etkisini belirlemek amacıyla hesaplan NNT (number needed to treat) istatistiğinin kullanım alanları, avantaj ve dezavantajları ile NNT'nin hesaplanması için kullanılacak alternatif yöntemler oluşturmaktadır. Kontrollü klinik denemelerde, tedavi etkinliğinin değerlendirilmesinde veya tedavilerin etkinliklerinin karşılaştırılmasında yorumlanması ve hesaplanması oldukça kolay olmasından dolayı, NNT son yıllarda yaygın olarak kullanılmaktadır. Önceleri, sadece randomize kontrollü denemelerde (Randomized Controlled Trials) kullanılması önerilen NNT'nin diğer alanlarda ve Meta analizinde kullanımı yaygınlaşmıştır. Bir kontrol ve bir tedavi grubundan oluşan bir denemede, uygulanan tedavinin etkisini belirlemenin en kolay yolu, tedavi ve kontrol grubundaki iyileşme oranları arasındaki farkın tahminidir. Bu fark atfedilen nispi risk (ARR) (absolute relative risk) olarak adlandırılır ve bu farka ait nokta tahminin güven aralığı da kolayca tahmin edilebilir. $1/ARR$ olarak tanımlanan NNT'nin hesaplanması basit olmasına karşın hesaplanmasında ve güven aralığında oldukça önemli sorunlar vardır. ARR sıfır değerini aldığı anda yani tedavi ve kontrol grubu arasında fark olmadığından NNT hesaplanamamaktadır. ARR yardımıyla bulunan güven aralıkları ise bazen negatif değer içermektedir. Bu durumda, NNT de negatif olacaktır ki bu değer yorumlanması sorunludur. Ayrıca NNT için bulunan güven aralığının nokta tahminini içermediği durumlarla da karşılaşılabilir. Tüm bu sorunlar nedeniyle, literatürde NNT'nin kullanımı konusunda farklı görüşler yer almaktadır. Çözüm olarak NNT'nin güven aralığının hesaplanmasında ARR'ye alternatif olarak RR'den yararlanılması önerilmektedir.

Bu çalışmada, tüm avantaj ve dezavantajları göz önünde bulundurularak, NNT'nin güven aralığını hesaplamak için bir simülasyon çalışması yapılmıştır. Kontrol ve tedavi gruplarında 50'şer vaka bulunduğu durumda

kontrol grubunda iyileşme oranının % 10 olduğu ve tedavi grubunda (% 10- % 90) arasında olduğu durumlar için 1000 tekrarlı olarak NNT değerleri ve güven aralıkları hesaplanmıştır. Simülasyon sonuçlarına göre, tedavi grubunda iyileşme oranının kontrol grubundaki iyileşme oranın en az 3 katı olduğu durumlarda, NNT'nin güven aralığının hesaplamasında RR kullanılmasının avantajlı olduğu görülmüştür. Bunun yanında kontrol grubunda iyileşme oranının % 10 ve tedavi grubunda iyileşme oranının kontrol grubundaki iyileşme oranın 1-2 katı olduğu durumlarda avantajlı olmadığı ve hesaplanan NNT değerinin dışında kalan güven aralığı oranın % 10 olduğu görülmüştür. Ancak bu koşullarda ARR kullanılarak hesaplanan güven aralıklarında bu oran çok daha yüksek % 25 olarak hesaplanmıştır.

**KLİNİK ARAŞTIRMALARDA “KÖRLÜK”:
KAVRAM, SINIRLAR, UYGULAMALAR ve
DEĞERLENDİRME**

**Mustafa Şenocak, Engin Baykal, Necdet Süt, Ömer Uysal,
Ahmet Dirican**

*İstanbul Üniversitesi, Cerrahpaşa Tıp Fakültesi, Biyoistatistik BD.
e-mail: necdetsut@yahoo.com*

Özet

Klinik araştırmalar, konu edinilen tıpsal girişim dışında, ortalama özellikleri olabildiğince benzer olgulardan oluşan kümelerin kıyaslanabilmesi için gerçekleştirilir. Girişim ve denetim kümeleri arasındaki, saptanabilen ve saptanamayan karıştırıcı etkenlerin dağılımını dengelemek ve seçim biaslarını engellemek için gerçekleştirilen randomizasyon bu yoldaki ilk adımdır. Ancak tek başına randomizasyon bu dengeyi sağlamaya yetmez ve içinde “körleme” yaklaşımının da bulunduğu ek yöntemler, her tür sonuç açısından çalışma çıktılarının en doğru şekilde eldesini sağlamak için mutlak gereklidirler.

En basit tanımı ile körleme, çalışmaya katılanlardan, kümelere dağılım özelliklerinin saklanmasıdır. “Katılanlar” kavramı hastaları ve girişimi uygulayanları içerdiği gibi sonuç verileri derleyenleri, yargılayanları hatta veri çözümlenmelerini yapanları bile içerebilir. Gerçekten de, çalışmaların özelliğine göre, bu yaklaşım listesindeki eksikliklerin biaslara, dolayısı ile de güvenilmez sonuçlara yol açabileceği gösterilmiştir..

Körleme yöntemi ilginç ve zengin bir tarihsel altyapıya sahiptir ve yaklaşık 2 asırdır kullanılmaktadır ancak bu gün bile araştırmacıların ve yayınları izleyenlerin çoğu körlemenin sınırlarını diğer yöntem bilimsel önlemlerle karıştırabilmektedirler. Körleme yöntemleri ve terminolojisi konusunda büyük anlaşmazlıklar olduğu ve tanımda bile ortak anlayış eksikliği bulunduğu araştırmalarla gösterilmiştir. Bu çerçevede, körleme, biasları azaltmak amacı ile pek çok araştırma tasarımında yer alabilmesine karşın sıklıkla salt randomize kontrollü araştırmalarla bağdaştırılmaktadır.

Çalışmamızda, körleme yaklaşımının kavramsal sınırlarını, farklı araştırma tiplerindeki yöntemsel özelliklerini, yayınlarda bildirim sorunlarını ve ölçümü ile ilgili yöntemleri gözden geçirmeyi amaçladık.

Sonuçta, körlemenin tanımı, sınırları, uygulanması açılarından ne yazık ki karışıklık yaşandığı saptanmaktadır. Açık tanımlar ve bilgilendirmeler, biasların azaltılması ve çok önemli bu yaklaşımın yanlış kullanımına bağlı güvenilmez sonuçların engellenebilmesi açısından acilen gereklidir.

Anahtar Kelimeler: Körlük, Kör, Klinik Araştırma, Yöntembilim.

İKİ RENKLİ MICROARRAY DENEME DÜZENLERİNİN PLANLANMASI

Seval Kul, E. Arzu Kanık

Mersin Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı

E-mail: sevalkul@mersin.edu.tr

Özet

Bu çalışmada farklı amaçlarla oluşturulmuş iki renkli Microarray analizlerinde yaygın olarak kullanılan deneme düzenleri ele alınmıştır.

Bir insanda nükleotidlerin oluşturduğu yaklaşık 20–25 bin gen vardır (1). Günümüzde DNA microarrayler sayesinde aynı anda binlerce genin analiz edilmesi mümkündür (2). Genetik veri analizindeki bu gelişmeler beraberinde microarray denemelerinin planlanması ve elde edilen büyük veri setlerinin istatistiksel analizi sorununu gündeme getirmiştir. Bilimsel bir deney düzenleme; yetenek, maliyet, donanım ve doğruluğun dengeli bir şekilde düşünülme sanatıdır (3). Microarray analizleri oldukça maliyetlidir. Bu nedenle microarray denemelerinin diğer denemeler gibi oldukça iyi planlanması gerekmektedir. Deneme düzeninin planlanması açık bir amaç ile başlar, ardından uygun örnek seçimi ve uygun deneme düzeni seçimi ve verilerin uygun istatistiksel analizle yorumlanması ile devam eder (4).

Microarray denemeleri birçok amaçla oluşturulabilir. Bunların çoğu üç temel kategorinin içine girer; Sınıf karşılaştırmaları (Daha önceden belirlenmiş sınıflardan elde edilen örnekler arasındaki gen ekspresyon profillerinin farklılık gösterip göstermediği), Sınıf Tahmini (Oluşturulan bir istatistik model yardımıyla yeni örneklerin ekspresyon profillerine göre hangi sınıfa ait olduğunun tahmin edilmesi) ve Sınıf Keşfi (Populasyon içinden elde edilen örneklerin yeni alt kümelerinin tanımlanması) (5, 6, 7).

Microarray analizlerinde iki renkli arrayler kullanıldığında örneklerin eşleştirilmesi ve etiketlenmesi için öncelikle uygun deneme düzeninin belirlenmesi gerekir. Bu tür çalışmalarda kullanılan üç farklı deneme düzeni vardır; Referans Deneme (Reference) Düzeni, Dengeli Blok Deneme (Balanced Block) Düzeni ve Döngü (Loop) Deneme Düzeni (5, 6, 7, 8). Bu çalışmada bahsedilen deneme düzenleri tartışılmış ve bu deneme düzenlerine ait optimum teknik ve biyolojik tekrar sayılarının hesaplanması üzerinde durulmuştur.

Anahtar Kelimeler: Microarray, referans deneme düzeni, dengeli blok deneme düzeni, döngü deneme düzeni, teknik ve biyolojik tekrar.

Kaynaklar

1. http://www.ornl.gov/sci/techresources/Human_Genome/project/20to25K.shtml/ Erişim Tarihi: 06.07.2005.
2. <http://www.stat.berkeley.edu/users/hhuang/STAT141/STATC141-lecture10.pdf/> Erişim Tarihi: 06.07.2005
3. <http://discover.nci.nih.gov/microarrayAnalysis/Experimental.Design.jsp/> Erişim Tarihi: 06.07.2005
4. Simon RM at all, Design and Analysis of DNA Microarrays Investigations, Statistics for Biology and Health, Springer, 2004.
5. Kerr MK, Churchill GA, Statistical Design and the Analysis of Gene Expression Microarray Data. Genet. Res., 2001: 77(2):123-8.
6. Simon R, Radmacher MD, Dobbin K, Design of Studies Using DNA Microarrays, Genetic Epidemiology 2002: 23: 21–36.
7. Simon RM, Dobbin K, Experimental Design of DNA Microarray Experiments, BioTechniques 2003: 34:26–21.
8. Dobbin K, Simon RM, Sample Size Determination in Microarray Experiments for Class Comparison and Prognostic Classification. Biostatistics. 2005: 6(1):27–38.

TEKRARLI ÖLÇÜMLERİN EĞRİ ALTINDA KALAN ALAN YARDIMIYLA ANALİZİ

Umur Özbek, Derya Öztuna, Beyza Dođanay, S.Kenan Köse

Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı

[e-mail:ozbekumut@yahoo.com](mailto:ozbekumut@yahoo.com)

Özet

Uzunlamasına çalışmalar, bireylerin belli bir izlem süresi içinde gözlemlendiđi ve çeşitli zaman aralıklarında gözlemlerin yinelenip; verilerin elde edildiđi çalışmalardır. Sağlık alanında sıklıkla kullanılan bu çalışmalara örnek olarak; hormonlar arası ultradian (kısa) ve sirkadyan (uzun) deđişimlerin tahmin edilmesi ve belirli bir zaman aralığındaki tümel salgılamalarının deđerlendirilmesi [1], sağlıklı ve hasta bireylerde aspirin emilimlerinin farklı zaman aralıklarında karşılaştırılması [2] verilebilir.

Tekrarlı ölçümler, bireysel farklılıkların ve gruplar arasındaki farklılıkların ortaya çıkarılmasında etkin bir tasarım olmasının yanında, zamana bađlı olarak alınan ölçüm sayısı fazla olduđunda istatistiksel analizin yapılmasında bazı problemlerle karşılaşılmaktadır. Her bir zaman ölçümü için ayrı ayrı gruplar arası karşılaştırma yapılması, istatistiksel sonuçların geçersiz olmasının yanı sıra, elde edilecek yorumların da yanlı olmasına neden olur. Çünkü bu tür bir analiz ile tekrarlı ölçümler arasındaki bađımlılık yapısı göz ardı edilmiş olmaktadır.

Tekrarlı ölçümler ve diđer deđişkenler arasındaki olası ilişkilerin ortaya çıkarılmasında, zamana bađlı tekrarlı ölçümleri birleştirmek için eğri altında kalan alan yaygın olarak kullanılmaktadır. Bu yöntemde, her bir birey için tekrarlı ölçümler üzerinden eğri altında kalan alan hesaplanarak, elde edilen deđerler yardımıyla gruplar arası farklılık karşılaştırması yapılır. **Eđri altında kalan alanın hesaplanması, istatistiksel analizi kolaylaştırır ve tekrarlı ölçümlerde bilgi kaybına neden olmaksızın testin gücünü artırır.**

Çalışmamızda, eğri altında kalan için yaygın olarak kullanılan iki farklı formül tanıtılacak ve kullanım alanları, Glukoz Tolerans Testi ile ilgili bir çalışmanın verileri üzerinde uygulanacak ve tartışılacaktır.

Anahtar Kelimeler: Uzunlamasına çalışmalar, tekrarlı ölçümler, eğri altında kalan alan.

Kaynaklar

1. Pruessner J.C., Kirschbaum C., Meinlschmid G., Hellhammer H.D. (2003). Two Formulas For Computation Of The Area Under The Curve Represent Measures Of Total Hormone Concentration Versus Time-Dependent Change. *Psychoneuroendocrinology*, 28, 916-931.
2. Matthews S.N.J., Altman G.D., Campbell J.M., Royston P. (1990). *Analysis of Serial Measurements in Medical Research*. *British Medical Journal*, 300, 230-235.
3. Davidian M. (2001). *Applied Longitudinal Data Analysis Lecture Notes*.
4. Kenward G. M. (1987). *A Method for Comparing Profiles of Repeated Measurements*. *Applied Statistics*, 36 (3), 296-308.

BAYESİAN OLASILIK KURAMININ TIPTA KULLANILMASI

Nazan Bilgel

*Uludağ Üniversitesi Tıp Fakültesi Aile Hekimliği Abd
e-mail: nazan@uludag.edu.tr*

Özet

Bu derleme yazıda bayesian olasılık kuramının tıpta kullanımı ile ilgili olarak yapılan çalışmalardan örnekler verilmekte ve son yıllarda tıbbi araştırmalarda kullanımı yaygınlaşan bu görüşün önemi vurgulanmaya çalışılmaktadır.

Bayesian olasılık kuramının tıpta kullanılması

Hekimlerin, hemşirelerin ve sağlık alanında çalışanların meslek yaşamları karar verme anlarıyla doludur. İşte bunlardan bazı örnekler ¹:

- Hastada, A hastalığı mı yoksa B hastalığı mı var?
- C hastalığının olmadığını ispatlamak için bu tanı testini yaptırmalı mıyım?
- Bu hastanın verdiği tıbbi öykü ne derecede güvenilirdir?
- Eğer hastada B hastalığı varsa onu X ilacı ile mi yoksa Y ilacı ile mi tedavi etmeliyim?
- Yeni bir ilaç olan Z'nin bilinen ilaçlar olan X ve Y'ye üstünlükleri nelerdir?
- Hastayı hangi yan etkiler konusunda uyarmalıyım?
- Bu hastayı bir uzmanın görmesine gerek var mıdır?
- Bu hastanın yaşama şansı nedir?
- Hastalığın uzun süredeki prognozu nasıl olacaktır?
- Bu hasta ne zaman kontrole gelmelidir?

Bu soruların kaç tanesi anlamlılık testleri kullanılarak cevaplandırılabilir? Çok fazlası değil. O halde niçin tıptaki istatistik eğitiminde zamanın yarısından çoğu anlamlılık testlerine ayrılmaktadır? Günümüzdeki istatistik eğitiminde en önemli aykırılık, klinik ve araştırmayla ilgili çoğu girişimlerin karar vermeyi sağlamak üzere başlatılması ancak karar verme aşamasının genellikle hala informal ve ad hoc olarak yapılmasıdır. Burada istatistikçinin

rolü veri özetlerini ampirik kanıt olarak sunmaktır. Bunların nasıl kullanılacağı ve kaynaklar arasından toplanılan bilgilerin nasıl birleştirileceğinin kararı ise klinisyenler ve araştırmacılara bırakılmıştır. İstatistiksel sentez ve karar verme arasındaki bu kopukluk doğal değildir ve olması istenmez. Çünkü böyle yapıldığında kantitatif bilginin etkisi göz ardı edilmektedir². Karar vermenin teorik yöntemleriyle ilişkili olan Bayesian istatistiğini kullanmak bu aykırılığı önlemede doğal bir yoldur. Ancak çok az derecede matematik bilgisi olanlara Bayesian etkileşimi öğretmek zordur³. Ayrıca:

- Bayesian teknikler az kullanıldığından
- Bayesciler standartlarını henüz tam olarak oluşturamamış olduklarından
- Bayesian sorgulamayı kavramak için şartlı olasılık kavramının bilinmesi gerektiğinden ve bunun da yeni başlayanlar için kafa karıştırıcı olabileceğinden

Bayes felsefesi ve tekniği tıp alanında klasik frekantist yöntemler kadar yoğun bir yer edinmemiştir.

Bayesian olasılık kuramı matematiksel istatistik kuramının bir dalıdır. Bu kuram belirsizlik taşıyan herhangi bir durumun modelini oluşturmak, bu durumla ilgili evrensel doğruları ve gerçekçi gözlemleri kullanarak sonuçlar üretmek amacıyla kullanılır⁴. Bu kuramı ilk olarak ortaya atan kişi 1702-1761 yılları arasında yaşamış bir İngiliz din adamı ve matematikçi olan Thomas Bayes'dir. Geçmişteki olasılık tahminlerine dayanarak yeni olguların tahminini yapmış ve bu görüşleri ölümünden sonra 1763 yılında "essay towards solving a problem in the doctrine of chances" isimli kitabıyla yayımlanmıştır. Olasılık kuramında en önemli kavramlardan biri "koşullu olasılıktır". $P(X=x|Y=y)=r$ koşullu olasılığının ifadesi şöyledir: "Y=y" nin doğru olması durumunda "X=x" olma olasılığı "r"dir. X ve Y'nin alabileceği değerlerin kombinasyonları için koşullu olasılıkları belirleyen tabloya koşullu olasılık dağılımı denir ve $p(X|Y)$ şeklinde ifade edilir. Koşullu olasılık "çarpım kuralını" belirlemede önemlidir. Çarpım kuralı iki olayın birden oluşma olasılığını tanımlar ve $p(A \cap B)$ ile ifade edilir. Bu durumda $p(A \cap B) = p(A|B)p(B) = p(B|A)p(A)$ olur.

Bu eşitliğin yeniden düzenlenmesi ile "bayes kuramı" elde edilir⁴:

$$P(A|B) = \frac{P(B/A)P(A)}{P(B)}$$

Burada:

A: Belirsizlik Taşıyan Önerme

B: Kanıt

$P(A|B)$: A'nın B kanıtından sonraki olasılığı (POSTERIOR)

$P(A)$: A'nın B kanıtından önceki olasılığı (PRIOR)

$P(B|A)$: B kanıtının A olayının gerçekleşmesi için oluşma olasılığı (LIKELIHOOD)

$1/P(B)$: Normalizasyon etmenidir ve hesaba katılmayabilir.

Bayesian istatistikte sık kullanılan terimler:

Önsel olasılık veya dağılım (prior): Herhangi bir gözlem yapmadan önce modelin gerçek olma olasılığıdır. Bu amaçla objektif veriler de subjektif görüşler de kullanılabilir.

Sonsal olasılık veya dağılım (posterior): Gözlemler dikkate alındığında modelin gerçek olma olasılığı olabilirlik (likelihood) belli bir modeldeki verilerin koşullu olasılık durumudur. Buna bayes faktörü adı da verilmektedir.

Bayesian ve klasik istatistik kuramları arasındaki en önemli fark olasılıkların değerlendirilmelerinde yatmaktadır⁵. Örneğin A ve B tedavilerini karşılaştıran bir çalışmada klasik istatistik kuramı ile a tedavisinin daha üstün olduğu sonucuna ulaşılabilir çünkü değerlendirme sonucunda tedaviler arasında gözlenen aşırı farklılığın gerçekte birbirine eşdeğer olma olasılığının çok düşük olacağı sonucuna varılmıştır. Bayesian istatistik ise işe gözlenen farklılık ile başlar ve A tedavisinin B ye üstün olma olasılığını değerlendirir. Diğer bir ifade ile klasik istatistikte saptanan belli bir duruma (ki burada iki tedavi arasında fark yoktur durumudur) gözlemlenen değerlerin uygun olma olasılığı araştırılırken, bayesian istatistikte gerçek durumun var olma olasılığı verilerden elde edilen belirsiz bir durum için (ki burada A tedavisinin B'ye üstün olması durumudur) değerlendirilir.

Klasik görüşte A ve B tedavileri arasındaki farkın önemli olduğu ($p=0,002$) söylenir. Bunun anlamı A ve B arasındaki böyle aşırı bir farkın gözlenme şansının 1000'de 2 olmasıdır. Aynı veriler için bayesian istatistik kullanıldığında ise A tedavisinin B'ye göre 0,999 kez üstün olduğu sonucuna varılır. Her iki değerlendirmede de A tedavisinin üstünlüğüne dair kuvvetli bir kanıt söz konusudur. Günümüzde hekimler elde ettikleri sonuçları p değerleri ile ifade etmeyi daha uygun bulmaktadır. Ancak çoğu kişi de bayesian olasılığı daha doğal ve iyi olarak görmektedir ve birçok hekim de p değerlerini hatalı olarak bayesian olasılığı için geçerli olan ifadeler biçiminde yansıtmaktadır. Tıp öğrencileri arasında yapılan bir sınavda $p<0,05$ 'in anlamı sorulmuş ve öğrencilerin çoğu A tedavisinin B'ye göre üstün olma olasılığının % 5'ten küçük olduğu cevabını işaretlemiştir.

Eğer bayes istatistiği daha doğal ise niçin daha çok tercih edilmemektedir? Bunun cevabı bayesian istatistikte önsel olasılığın bilinmesi gerekliliğinde yatmaktadır. A tedavisinin B'ye üstün olması olasılığının

belirlenmesinde bir önsel dağılımın yani çalışma başlamadan önce A tedavisinin B'ye üstün olma olasılığını gösteren bir kanıtın var olması gerekir. Bu kanıt ise her zaman var olmayabilir veya subjektif değerlendirmelere, bireysel görüşlere bağlı olarak yaratılmış bir önsel olasılık var olabilir. Klasik istatistiksel görüşü yani frekantist yöntemleri kullananlar bu noktada bayesian istatistiği subjektif oluş ile suçlamaktadır. Buna karşılık, bayesian kuramı savunanlar, frekantistlerin de subjektif yaklaşımlar sergilediklerini söylemekte ve örneğin niçin önemlilik düzeyi olarak $p < 0,05$ 'i seçtiklerini sorgulamaktadırlar. Frekantistler ile bayesianlar arasında bu tartışmalar sürerken bazı araştırmacılar da bayesian yöntemleri yoğun biçimde kullanmaktadır.

Bayesian istatistiğin tanı testlerinde kullanılması hekimler ve epidemiyologlar tarafından bilinen bir olgudur ve bayesian görüşün tıp alanındaki ilk uygulamalarından biridir⁶. Sensitivite, spesifite, pozitif ve negatif prediktif değerler, yalancı pozitiflik, yalancı negatiflik kavramları bayesian görüşün tıp alanındaki en yaygın kullanım alanlarındandır. Hastalıkların nüks etmesi veya tedavilerindeki başarısızlığı değerlendirme amacıyla da bayesian yöntemler kullanılmaktadır. Belisizlik durumlarında karar vermeyi sağlamak için de bu yöntem kullanılmaktadır. Farklı durumlardaki olasılıklar bayesian yöntem kullanılarak hesaplanmakta ve bu duruma uyan hallerde belirli kararların alınması sonucunda elde edilecek yararların neler olabileceği ortaya konulabilmektedir. Böylece elde edilebilecek yararların en üst düzeye çıkarılabilmeleri sağlanmaktadır. Detsky⁷, hangi tür klinik çalışmaların maliyet –etkinlik bakımından en yararlı olduğunu, Eddy⁸, çeşitli kanser türleri için toplumda yapılacak tarama testlerinin yararlılığını bayesian yöntemleri kullanarak saptamıştır.

Tıpta bayesian yöntemlerin uygulanmasına karşı direncin en kuvvetli olduğu alanlar klinik deneyler ve epidemiyolojik çalışmalardır. İlaçların etkilerini, dozlarını ve güvenilirliklerini saptamaya yönelik olarak yapılan faz I ve faz II çalışmalarında bayesian yöntemlerin kullanılmasına karşı olan direnç daha azdır. Faz III çalışmaları ve epidemiyolojik çalışmalarda ise bayesian yöntemlerin kullanılmasına karşı olan direnç devam etmektedir⁶. Ancak son yıllarda epidemiyolojik çalışmalarda bayes kuramının kullanıldığı da gözlenmektedir. Bunlardan biri de meme kanseri için bir risk oluşturan BRCA1 ve BRCA2 gen mutasyonları ile ilgili çalışmalardır⁹. Ailesinde bu genlerin var olduğu bilinen bir kadının meme kanseri olma olasılığı bayesian yöntemlerle hesaplanabilir. Bunun için önsel olasılık olarak toplumdaki mutasyon sıklığı, mutant gen taşıyıcısı olan ve olmayan kişilerdeki yaşa özel meme kanseri sıklığı bilgilerine ihtiyaç vardır. Ancak genotipi ilgilendiren bu bilgilerin yokluğunda bile epidemiyolojik verilerin bayesian yöntemle değerlendirilmesi sonucunda genetik etkiler ve gen çevre etkileşimlerini saptamak mümkündür. Iversen ve arkadaşları bu yöntemle meme kanseri olan BRCA1 ve BRCA2 taşıyıcıları ile meme kanseri olan ve taşıyıcı

olmayan kadınlardaki yaşam süresi arasındaki farkı hesaplamışlardır. Bir hastalık için bir kişinin sonsal olasılığını saptamak ile epidemiyolojik verilere bayesian yöntemlerin uygulanması arasında temelde bir farklılık yoktur. Tanıya varma amaç olduğunda hekim hastadan elde ettiği verilere dayanarak o hastalık için bir önsel olasılık saptar. Bu önsel olasılık, uygun tanı testlerinin sonuçları kullanılarak güncelleştirilir ve sonsal olasılık elde edilir. Önsel olasılık ve tanı testlerinin sonuçlarının olabirlikleri (likelihood) bayes kuramına yerleştirilerek sonsal olasılık hesaplanır. Aynı durum epidemiyolojik çalışmalar için de söz konusudur⁶. Bayes kuramında sonsal olasılık (posterior), önsel olasılık (prior) ile olabirliğin (likelihood) bir fonksiyonudur. Yani önsel olasılıktan etkilenir ancak, örnek büyüklüğü arttıkça bu etkilenme giderek azalır ve bu fonksiyonda olabirlik (likelihood) daha önem kazanmaya başlar. Sonsal olasılık (veya dağılım), önsel olasılık (veya dağılım) ve mevcut verilere bağlı olarak bir bilinmeyenle ilgili olan bilgileri birleştirerek özetler (maruz kalanlardaki odds ratio gibi) epidemiyolojik çalışmalarda bayesian yöntem kullanılmasıyla karıştırıcı faktörlerin (confounding) kontrolü de daha güçlü olarak sağlanmış olur. Örneğin bebek ölümleri ile ilgili epidemiyolojik çalışmalarda annenin sigara içmesinin ölümleri hafifçe arttırdığı bilinmektedir (odds ratio= 1,3) klasik lojistik regresyon modelinde olası karıştırıcı faktörler, sigara içme, beden kütle indeksi, yaş, ırk gibi modelde yer alır. Ancak eğer çalışma yeterince büyük değilse, bebek ölümüne etkili olduğu bilinen bazı faktörlerin likelihood ratio testine göre anlamsız olduğu görülebilir. Araştırmacılar bu durumda confounding katsayısı anlamsız olan karıştırıcı faktörleri modelden çıkararak, önceki çalışmaların verilerine dayanarak ortalama sabit bir katsayıyı modele ekleyerek veya küçük ve orta çaplı araştırmaları düzenlemekten kaçınarak çözüm bulmaya çalışırlar.

Bayesian analizde ise önceki çalışmaların bilgileri (regresyon katsayıları) kolaylıkla modele dahil edilebilir. Önceki kısıtlamalar, bilinmeyenlerin olası değerlerine eklenebilir veya önsel dağılım önceki çalışmaların istatistiklerine bağlı olarak oluşturulur. Örneğin önsel dağılım daha önce yapılan ve sigara içmenin bebek ölümleri üzerindeki etkisini gösteren odds ratio'yu 1,3 olarak bulan çalışmanın bu değerine göre düzenlenebilir ve bu önsel dağılımda 1'den küçük değerler için 0 önsel olasılık değeri atanır. Böylece maruz kalanlardaki odds ratioyu tahmin etme frekantist multiple logistik regresyon analizine göre daha iyi bir biçimde yapılmış olur¹⁰. Bayesian istatistikte gözlem sayısı arttıkça, sonsal dağılımın önsel dağılımdan etkilenmesi giderek azalır ve odds ratio intervalleri daha çok gözlemlenen değerlerden etkilenmeye başlar.

Son yıllarda Bayesian yöntemlerin yaygın olarak kullanılmaya başlamasının en önemli nedeni bayesian hesaplamalar için monte carlo markov zincirleri algoritmalarının geliştirilmiş olmasıdır. Farklı modellere uygulanabilecek şekilde geliştirilmiş yazılımlar da mevcuttur. Bunlardan biri de ücretsiz olan ve internette indirilebilen Win Bugs isimli (www.mrc-

bsu.am.ac.uk/bugs) programdır. Bayesian hiyerarşik modeller şimdiye kadar pek çok epidemiyolojik çalışmada kullanılmıştır. Belli aralıklarla toplanan verilere dayanarak kronik bir hastalığın doğal hikâyesinin belirlenmesi çalışmasında, farklı zamanlardaki hastalık hızlarının karşılaştırılmasında, epidemiyolojik çalışmalarda ölçüm hatalarının saptanmasında, meme kanseri ile ilgili çalışmalarda, genetik çalışmalarında, fertilité ile ilgili çalışmalarda olduğu gibi. Bu çalışmalar bayesian görüşün sonuçları açıklamak ve maruziyet-hastalık ilişkisini ortaya çıkarmak için frekantist görüşe göre daha sağlam ve gerçekçi olduğunu ortaya koymuştur¹¹.

Randomize kontrollü klinik deneylerde bayesian yöntemlerin kullanılmasıyla elde edilen sonuçların daha güvenilir olacağını gösteren çalışmalar da mevcuttur. Bunlardan biri de gusto araştırmasıdır. Bu çalışmada kalp krizi geçiren 41.000 acil hastaya streptokinaz ve doku doku plazminojen aktivatörü t-PA verilmiş ve t-PA verilen hastalar arasında yaşayanların streptokinaz verilenlere göre daha fazla olduğu, aradaki farkın 0.001 düzeyinde anlamlı bulunduğu saptanmıştır. Bu araştırma sonucunda daha pahalı olan yeni ilacın yani t-PA'nın kalp krizi geçirenlerde daha etkili olduğu sonucuna varılmış ve FDA bu ilacı onaylamıştır. Brophy ve Lawrence¹² 1995 yılında aynı verileri bayesian görüşü ve yöntemleri kullanarak çalışmayı yeniden değerlendirmişler ve iki ilacın birbirine eşdeğer olduğunu bulmuşlardır. Özellikle kanıta dayalı tıp uygulamalarında bayesian görüşün kullanımı giderek yaygınlaşmaktadır. Randomize klinik çalışmalarda bayesian ve frekantist yöntemlerin kısa bir karşılaştırması aşağıdaki tabloda gösterilmiştir¹³.

Konu	Frekantist Yöntemler	Bayesian Yöntemler
Araştırmadan önce var olan bilgi	Araştırma düzeninde informal olarak kullanılır	Önsel dağılım oluşturularak formal biçimde kullanılır
İlgilenilen parametrenin sunumu	Sabit bir değer olarak	Bir olasılık dağılımına sahip bilinmeyen bir kantite olarak
Temel soru	Veriler belli bir parametre değerine ne ölçüde uygunluk gösterir?	Belli bir parametre değeri verilere ne ölçüde uygunluk gösterir?
Sonuçların sunumu	Güvenilirlik aralığı, p değeri, likelihood fonksiyonları	Parametrenin sonsal dağılım plotu, sonsal olasılık hesapları ve sonsal dağılımın karar analizinde kullanılması

Bayesian görüş pek çok bilimsel çalışmaya rahatça uygulanabilir ve bilgisayar olanaklarının gelişmesiyle birlikte giderek daha fazla kullanım

alanı bulmuştur. Isı veritabanında 1990-1998 yılları arasında yapılan bir taramada 4000 civarında araştırmanın bu yöntemle değerlendirildiği saptanmıştır¹³. Medline veri tabanında yapılan taramada ise bayesian görüşü veya bayes kuramı elementlerini kullanan araştırma sayısı aşağıdaki tabloda gösterilmiştir.

Yıllar	Araştırma sayısı
1969-1979	79
1980-1989	414
1990-1999	1382
2000-2005	3006

Bu sayılar tıbbi araştırmalarda bayesian kuramın kullanılmasına olan ilginin giderek arttığını göstermektedir. Bu veri tabanında bayesian yöntem ile yapılmış türkiye ait bir tıbbi araştırmaya rastlanılmamıştır.

Sağlık konuları günümüzde daha karmaşık hale gelmiştir. Çok sayıda etkenin sağlık üzerindeki etkilerinin belirsizliği giderek daha çok ilgi uyandırmaktadır. Bayesian yaklaşım bu etkilerin gerçek değerini saptamamızda ve karar vermemizde bize çok yararlı olacaktır. Çünkü bu yöntemle önceden sahip olunan bilgilerin tümünün yeni verilere eklenmesi ve bu yeni verilerle birlikte analizi mümkün olmaktadır. Bu nedenle tıp eğitiminde ve araştırmacıların yetiştirilmesinde bayesian görüşe de yer verilmesi ve bayesian yöntemi bilen istatistikçilerle işbirliği içinde çalışılması büyük önem taşımaktadır.

Kaynaklar

1. Berry, D.A.(1996). Teaching elementary bayesian statistics with real applications in science. The american statistician, 51, 241-246.
2. Stangl D. (2002). From testing to decision making: changing how we teach statistics to health care professionals. Icots6. [Http://www.bayesian-initiative.com](http://www.bayesian-initiative.com)
3. Moore, D.S. (1997). Bayes for beginners? Some reason to hesitate. The american statistician, 51, 254-261.

4. Bullard F (2001). A brief introduction to bayesian statistics. Nctm <http://www.bayesian-initiative.com>
5. Editorial (1996). Bayesian statistical methods. Bmj, 313, 569-570.
6. Dunson D.B. (2001) Commentary: Practical advantages of bayesian analysis of epidemiologic data. American journal of epidemiology, 153(12), 1222-1226.
7. Detsky, A.S. (1985) Using economic analysis to determine the resource consequences of choices made in clinical trials. J chron dis, 38, 753-765.
8. Eddy, D.M., Hasselblad, V., Shachter R.A. (1990) Bayesian method for synthesising evidence: the confidence profile method. Int j technol assess health care, 6. 31-55.
9. Berry DA, Parmigiani, G., Sanchez, J. Et al. (1997). Probability of carrying a mutation of breast- ovarian cancer gene brca1 based on family history. J natl cancer ins, 89, 227-238.
10. Goodman S.N. (1999). Toward evidence based medical statistics. The bayes factor. Annals of internal medicine, 130(12), 1005-1013.
11. Goodman S.N. (1999). Toward evidence based medical statistics. The p value fallacy. Annals of internal medicine, 130(12), 995-1004.
12. Brophy J.M. and Joseph, I. (1995). Placing trials in context using bayesian analysis: gusto revisited by reverend bayes. Journal of the american medical association, 273, 871-875.
13. Spiegelhalter D.J, Myles J.P., Jones D.R., Abrams K.R. (1999). Methods in health service research. An introduction to bayesian methods in health technology assessment. Bmj, 319, 508-512.

İKİ METROPOL ŞEHİRDE HALKA VERİLEN SU KALİTESİ VE BELEDİYELERİN SU POLİTİKALARI

Ömer Satıcı

Dicle Üniversitesi Tıp Fakültesi Diyarbakır

Özet

Amacımız iki mega kentin su hizmetlerini karşılaştırmaktır. Su, hayatın devamını sağlamanın çok daha ötesinde işleve sahiptir. İki metropol kentimizde de sular idaresi vatandaşın içme suyunu çeşmelerden sağlmasını tavsiye ediyor. İSKİ ve DİSKİ'nin müşterilerine ait tanımlayıcı istatistikler kullanıldı. Diyarbakır Merkez Kent Nüfusu 800.000 civarında olup, 133070 abonesi vardır. Halen günde ortalama olarak 200.000 metre-küpün üzerinde su üretimi yapılmaktadır. Buna rağmen kentin bazı bölgelerinde susuzluk problemi yaşanmaktadır. DİSKİ laboratuvarlarında 18 farklı değişken analiz edilmektedir. Diyarbakır'da meskenlerde su fiyatı 20 m³/ayda 1 YTL dir. Kayıp su oranı % 67'dir. Müşteri adedi 1999 yılında 90,844 iken bu sayı 2004'te 133,070'e çıkmıştır. Yıllara göre müşteri adedi artışı 2000 de % 8.08 iken 2001 de % 14.52 ve azalarak 2004'te % 6'ya inmiştir.

İstanbul'un nüfusu ise 12 milyonu aşmaktadır. Bugün her eve ilin Su Kanalizasyon İdaresi'nin (İSKİ) dünya standartlarındaki suyu ulaştırıyor. İstanbul'da günlük olarak ortalama 2 milyon m³ içme suyu tüketiliyor. Toplam 3.558.000 m³/gün kapasiteye sahip arıtma tesislerinde dünyanın ileri teknolojileri kullanılarak ozon ve aktif karbon sistemiyle arıtma işlemi gerçekleştiriliyor. Şehre verilen içme suları arıtma tesisi çıkışından ve İstanbul'un 350-400 farklı noktasından günlük olarak alınan örnekler, son teknolojilerle donatılmış İSKİ laboratuvarlarında 37 farklı değişkenle analiz ediliyor. İstanbul'da meskenlerde su fiyatı 25 m³/Ay'a kadar 1,5 YTL'dir. Kayıp su oranı % 25-30 civarındadır. Müşteri adedi 1994 yılında 1,715,820 iken bu sayı 2004'te 3,482,752'ye çıkmıştır. Yapılan analiz neticeleri İstanbul'un musluk suyunun Türk Standartları Enstitüsü (TSE 266), Avrupa Birliği (EC), Dünya Sağlık Teşkilatı (WHO), ABD Çevre Koruma Ajansı (EPA)'nın ön gördüğü kriterlerde olduğunu gösteriyor. Yorum: Su maliyetinin kırsal kesimlerde daha pahalı olması gerekmele birlikte asıl maliyet mesafe, alt yapı ve dağıtım hizmetlerine gitmektedir. DİSKİ, İstanbul'a göre suyu ucuza vermektedir. İstanbul gibi Diyarbakır da çok göç almıştır. Diyarbakır'a

gelen göç dalgası daha çok köylü kökenli olduğu için suyun bedava olduğu anlayışı gözlenmektedir. İSKİ ve DİSKİ'nin çalışmaları birlikte incelendiğinde politikalarla ilgili ilginç sonuçlar elde edilebilir.

Anahtar Kelimeler: Metropol kentler, içme suyu kayıpları, politikalar.

1- Giriş

Su, size ne sağlayabilir? Su gereklidir. Su, hayatın devamını sağlamanın çok daha ötesinde işleve sahiptir. İyi su gerçekten –sağlığımızı-zindeliğimizi –görünüşümüzü geliştirebilir. Su olmazsa ne olur? Vücut suyunda % 1'lik bir azalma bizi susatır. % 5'lik azalma hafif ateşe neden olur. % 8'lik bir azalmada tükürük bezleri tükürük üretemez, deri mavileşir. % 10'luk bir azalmada dil şişer, böbrekler işlev göremez, kaslarda spazm başlar, yürümek mümkün değildir. Suyun vücuttaki işlevi nedir? Vitamin ve besinleri hücrelere, dokulara ve organlara taşır. Gözlerimizi, ağzımızı ve burun kanallarımızı nemlendirir. Vücut ısını dengeler. Hücrelere oksijen taşır. Su size ne sağlayabilir? Su, darbeleri ve organlara karşı zararı en aza indirgeyen, şok emici işlevi görür. Eklem için kayganlığı sağlayıp, kolay hareket etmelerini sağlar. Böbrekleri yıkayarak, vücudu zehirli maddelerden arındırır. Kan basıncını kontrol eden elektrolitleri dengeler. Daha da ötesi suyun size sağladığı, Vücuttaki kan miktarını muhafaza ederek, enerjinizi korur. Yeterli miktarda su alınımı, konsantrasyonunuzu artırır, reflekslerinizi geliştirir. Yaktığınız kalori miktarını artırır. İlacın vücuda daha çok ve daha çabuk nüfuz etmesini sağlar. Soğuk algınlığından kansere kadar birçok hastalık riskini azaltır. İçme suyu sağlık ve zindeliğinizi geliştirebilir. Kendinizi daha tok hissedersiniz. Egzersiz esnasında daha çok enerji verir ve yakılan enerji miktarını artırır. Vücutunuzdaki yağ depolarının azalmasına yardımcı olabilir. Cildi nemlendirerek, pürüzsüz ve yumuşak olmasını sağlar. Su, iyi su ise iyidir.

1.1 Su size ne yapabilir?

İçme suyum temiz mi? Temiz gözüküyor. Temiz kokuyor. Temiz tadı var. Temiz olduğunu düşünüyorsunuz. Su, partikül (tortu -çamur- asbest) içeriyor mu? Kimyasallar açısından (inorganik-organik) temiz mi? Suda kist-virus-bakteri gibi mikro organizmalar var mı? İçtiğim suda mikrobiyolojik kirlenme var mı?(1, 2, 3)

1.2 İçme suyumuz nasıl kirlenir?

Kurşun borular, lehimlenmiş borulardaki mikrobiyolojik oluşum. Paslanan boru ve su depolarından partiküller. PVC borulardan Vinil Klorid,

Toprak üzerinden sızan endüstriyel atıklar. Tüm bunlara benzer artıklar ve şehir suları, klorlama işlemi, kloroform gibi maddelerin oluşumuna neden olabilmektedir (3, 4, 5).

1.3 Suyun sağlığa etkileri?

Kurşun; anemiye, düşük kiloda doğumlara, erişkinlerde yüksek tansiyona neden olabilmektedir. Civa; böbrek hasarına yol açabilir. Radon; mide kanseri riskini artırabilir. EPA (Environmental Protection Agency) araştırmaları; mesane kanser oranının % 2 -% 10'u kirlenmiş içme suyundan kaynaklanabileceğini göstermektedir (5, 6, 7, 8, 9).

1.4 Su, iyi su ise iyidir?

İyi suyu nereden bulabiliriz? Bu soru insanlık var olalı beri sorulmuştur. Şişe suyu, büyük şehirlerde avantajları ve dezavantajları ile insan hayatına girmiştir. Suyun kalitesinin yanında maliyeti ve zahmetli oluşu da incelenmesi gereken bir konudur. Yaz sıcaklarıyla birlikte serinlemek için markete ya da büfeye koşarak aldığımız pet şişe suların da çevreye ödediği ciddi bir bedel var. Worldwatch'ın araştırması, dünya genelinde şişe suyu tüketiminin her yıl yüzde 12 oranında artış gösterdiğine dikkat çekiyor. Dünya çapında tüketiciler, şişe suları için yılda 35 milyar dolara yakın harcama yapıyor. Şişe suyu üretiminin çevreye maliyeti konusunda en büyük kaygı ise, su talebindeki artışın mevcut su kaynaklarında sıkıntı yaratması. Şişe Geri Dönüşüm Enstitüsü, su şişelerinde kullanılan plastik türü olan saf reçine PET (polietilen tereftalat) satışlarının, 1990-1999 yılları arasında iki katına ulaştığını ve 738 milyon kiloyu bulduğunu belirtiyor. 1 kilo PET üretmek içinse 17.5 kilo su tüketiliyor! Ayrıca işlem sırasında 40 gram hidrokarbon, 25 gram sülfür oksit, 18 gram karbon monoksit, 20 gram nitrojen oksit ve 2.3 kilogram karbondioksit havaya karışıyor (9, 10, 11).

1.5 İyi suyu nereden bulabiliriz?

Şehrin musluklarından akan suyu çekinmeden içebilecek miyiz? Kaynatmakla suyun dezavantajları ve avantajları nelerdir? Kaynatırsak suyun kokusu, rengi, tadı değişecek mi? Sudaki partiküller, organik ve inorganik maddeler azalmaz mı? Bu zahmete gerek var mıydı? Belediyemin bana verdiği suyun avantajları nedir? Kloru azaltmış mı? Suyun tadını, kokusunu, rengini geliştirmiş mi? Organik bileşikleri azaltmış mı? İçtiğim suyun bana dezavantajları nedir? İnorganik bileşikleri, bakteri ve virüsleri fazla mı? Kloru, Trihalometan'ları ve pek çok organik bileşiği azaltmış mı (9,

10, 11)? Küçük partikülleri derinlemesine filtre etmiş mi? Pek çok inorganik bileşiğin azaltılmasını yapmış mı? Bakteri ve virüslerin azaltılmasında etkili midir? Şehre su veren, su fabrikası, yeterli teknolojik imkanları olan bir yer mi?

2. Materyal ve Metot

Çalışma iki mega kentimize ait su ve kanalizasyon idarelerinin çalışmalarından elde edilen veriler üzerinden yapıldı. İki mega kent İstanbul ve Diyarbakır'dır. Diyarbakır Merkez kent nüfusu 800.000 civarında olup, kişi başına ihtiyaç duyulan su miktarı 150 lt. gibi yüksek bir değer alındığında kentin günlük su ihtiyacı 120.000 metreküptür. Halen Gözeli ve Arıtma Tesisinden günde ortalama olarak 200.000 metreküp üzerinde su üretimi yapılmaktadır. Alman teknolojisi kullanan DİSKİ, klorlama yöntemi ile arıtma işlemini gerçekleştiriyor. 2005 yılının ilk 6 ayında kaçak su kullanan 947 aboneye 159 milyar TL ceza verilmiş olup, tekrarı halinde yasal yollara başvurulacağı bildirilmiştir. Diyarbakır'da kayıp su oranı % 67 kadardır. İstanbul'da nüfus 12 milyonu aşmaktadır. İstanbul'da bugün her eve ilin Su Kanalizasyon İdaresi'nin dünya standartlarındaki suya ulaşıyor. İstanbul'da günlük olarak ortalama 2 milyon m³ içme suyu tüketiliyor. Toplam 3.558.000 m³/gün kapasiteye sahip arıtma tesislerinde dünyanın ileri teknolojileri kullanılarak ozon ve aktif karbon sistemiyle arıtma işlemi gerçekleştiriliyor. Şehre verilen içme suları arıtma tesisi çıkışından ve İstanbul'un 350-400 farklı noktasından günlük olarak alınan örnekler, son teknolojilerle donatılmış İSKİ laboratuvarlarında 37 farklı değişkenle analiz ediliyor. Yapılan analiz neticeleri İstanbul'un musluk suyunun Türk Standartları Enstitüsü (TSE 266), Avrupa Birliği (EC), Dünya Sağlık Teşkilatı (WHO), ABD Çevre Koruma Ajansı (EPA)'nın ön gördüğü kriterlerde olduğunu gösteriyor. Kayıp su oranı % 25-30 civarındadır. Çalışmada, her iki kente ait yıllara göre su kayıpları, müşteri adeleri ve artış yüzdeleri, yıllara göre müşterilerin yapmış olduğu itirazlar incelendi. Ayrıca, yıllara göre tahakkuk ve tahsilat oranları verildi. İki mega kentin su fiyatlarına ait çalışmaları değerlendirildi. İSKİ ve DİSKİ'nin kendi vatandaşlarına verdikleri suyun kalitesi karşılaştırıldı.

3. Bulgular

İstanbul'da ve Diyarbakır'da nüfus arttıkça, teknoloji ilerledikçe suyun daha fazla tüketilmesi doğaldır. Köyünde, kentinde yaşayan insanlar için, doğuda kişi başına su tüketimi günde 15-60 litre yetiyordu. 1990'lı yıllarda yaşanan göç, nüfusu hızla hareketlendirip, değiştirmiştir. Otomatik yıkayıcılar ve pek çok yeni cihazların kullanımı ile su giderek yetmez olmuştur. Tarımda, sulamada daha fazla suyun kullanılmasını gerektirmiştir.

3.1 İSKİ ve DİSKİ' nin su fiyatları ve su kalitesine ilişkin bulgular aşağıdadır.

İstanbul İSKİ ve Diyarbakır DİSKİ için su satış tarifeleri aşağıda verildiği gibidir.

Tablo 1. İSKİ ve DİSKİ'nin SU SATIŞ TARİFELERİ 01-01-2005'den Geçerli

Tarife	İSKİ Fiyatı (YTL/M ³)	DİSKİ Fiyatı (YTL/M ³)
1-1 Konut 1(Konut başına ilk 25 m3/ay kullanım)	1.5	0.72
1-2 Konut 2(Konut başına ilk 26 m3/ay kullanımdan +)	3.0	0.82-1.83
2- İşyeri	3.0	2.47
3- Konutlarda müşterek kalorifer	1.5	
4- Konut-İşyeri bahçe ve çevre sulama	1.5	
5- 1Resmi okullar	1.3	2.26
5-2 Resmi Hastaneler-Özel Hastaneler	2.5	0.92
5-3 Genel ve Katma Bütçeli Kuruluşlar	2.5	0.92
6- Sanayi ve Şantiye	3.0	1.01
7-Toplu Su	1.0	
8-1 Tanker, Kuyu suyu 1(0-25 m3/ay a kadar kullanım)	0.75	
8-2 Tanker, Kuyu suyu 1(26 m3/ay dan fazla kullanım)	1.5	
9- Tanker ve Kuyu suları	1.5	
10-1 Tanker ve Kuyu suları bahçe sulama	0.75	
11-1 Turizm teşvik Belgeli Kuyu (0-25 m3/ay a kadar)	0.75	
11-2 Turizm teşvik Belgeli Kuyu (26 m3/ay dan fazla)	1.5	
12- Endüstriyel KÖP	0.2	

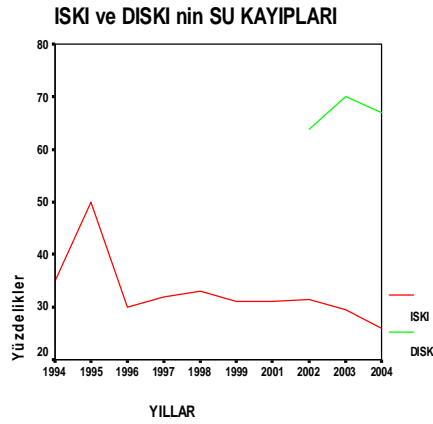
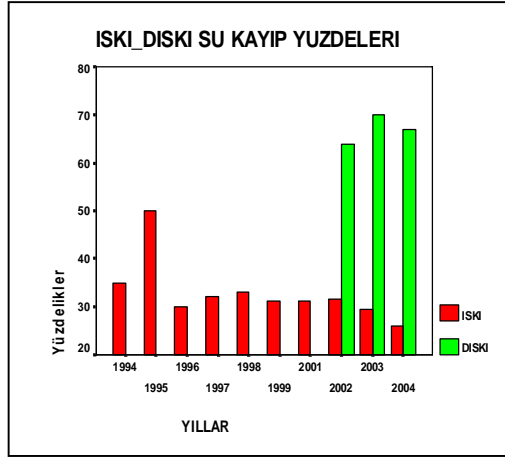
Açıklama: İstanbul'da Tarihi Hamamlara 3 YTL, Spor etkinlikleri 1. Kademe için 1,5 YTL, 2. kademe için (26 m3/ay'dan fazla kullanım) 3 YTL den su verilmektedir. Ticari gayesi olmayan Dernek Vakıflar için 1. kademe 1,5 YTL, Turistik olmayan ve Turistik Teşvik Belgeli Oteller için 1,5 YTL fiyatlar uygulanmaktadır. Diyarbakır'da ise, DİSKİ tüm tarife gruplarına 3 aylık periyotlarda DİE 'nin Tüketici Fiyat Endeksi (TÜFE)'yi yansıtıyor.

3.2 İSKİ ve DİSKİ' nin yıllara göre su kayıpları incelendiğinde aşağıdaki sonuçlar gözlenmektedir.

D.S.İ uzmanlarınca yapılan hidrojeolojik etüt sonuçlarına göre; Türkiye'nin yıllık su rezervi yaklaşık 175 milyar m³tür. Bunun yaklaşık 9.5 milyar m³ kadarı yeraltı suları, 165.9 milyar m³ kadarı da yerüstü su kaynakları tarafından oluşturulur. Potansiyelin, yaklaşık % 35 kadarı sosyal ve ekonomik hayatın hizmetine sunulabilmiştir. Örneğin nüfusumuz, 2010 yılında 100 milyona yaklaşacağı hesaplanmaktadır. O halde, batı standartlarında bir toplumsal hayat düzeyine ulaşması gereken Türkiye için bu potansiyel, ileri sürüldüğü kadar fazla değildir. Mevcut su kayıpları miktarlarının bulguları aşağıda verildiği gibidir.

Tablo 2. İSKİ ve DİSKİ nin yıllara göre su kayıpları oranları

Yılı	İSKİ % si	DİSKİ % si
1994	35	
1995	50	
1996	30	
1997	32	
1998	33	
1999	31	
2001	31	
2002	31,5	64
2003	29,5	70
2004	26	67
Ortalama	32±6,42	67±3



İSKİ'nin su kayıpları % 26 ile % 35 arasında değişirken, DİSKİ'nin su kayıpları % 64 ile % 70 arasında değişmektedir. DİSKİ'nin su kayıp oranları İSKİ'ye göre daha yüksek gözlenmektedir ($p < 0.01$). İSKİ'de yıl ve su kaybı yüzdesi değişkenleri arasındaki korelasyon r yıl, su yüzdesi = $-0,63$ dir ($p < 0.05$). Regresyon Denklemi; Y (Su Kayıp % si) = $40,3 - 1,345x(\text{Yıl}) + e$ dir ($p < 0.05$). Katsayılar önemli bulunmuştur ($p < 0.01$)

3.3 İçme suları ile ilgili yönetmeliğin amacı

İnsani tüketim amaçlı suların teknik ve hijyenik şartlara uygunluğu ile suların kalite standartlarının sağlanması, kaynak suları ve içme sularının üretimi, ambalajlanması, etiketlenmesi, satışı, denetlenmesi önemlidir (12).

Tablo 3. İstanbul ve Diyarbakır Su Kalitesi Raporu Mayıs 2005

Değişkenler	Türk Standartlar TSE 2866 1997	Dünya Sağlık Teşkilatı (WHO)	ABD Çevre Koruma Ajansı (EPA) 2002	Avrupa Birliği (EC) 1998	Tesis Ortalama Kalite Değerleri			
					B. Çekmece	İkithane	Kağıtkhane	Diyarbakır
Bulanıklık	25	5	1	1	0,2	0,3	0,2	0,12
Birincil Standartlar Mikrobiyolojik EMS/100 ml								
Koliform Bakteri	<1	0	<1	0	0	0	0	0
Birincil Standartla Dezenfeksiyon yan ürünleri µg/l								
Top.Trihalometanlar	--	460	80	100	26	15	29	--
Bromat	--	25	10	10				--
Birincil standartlar (İnorganik Kimyasallar) µg/l								
Alüminyum	0,20	0,20	0,20	0,20	0,046	0,067	0,037	0,01
Arsenik	0,05	0,05	0,05	0,01	0,000	0,000	0,000	0
Baryum	0,3	0,7	1	-	0,037	0,024	0,018	
Kadmiyum	0,005	0,005	0,01	0,005	0,000	0,000	0,000	
Krom (Toplam)	0,05	0,05	0,05	0,05	0	0	0	0
Florür	1,5	1,5	0,7-2,4	1,5	0,13	0,07	0,06	
Kurşun	0,05	0,05	0,05	0,01	0,0003	0,001	0,001	
Civa	0,001	0,001	0,002	0,001	0	0	0	
Nitrat (NO ₃)	50	50	45	50	0,85	1,6	0,6	0,2
Selenyum	0,01	0,01	0,01	0,01	0	0	0	
Gümüş	0,01	-	0,05	-	0	0	0	
Antimon	0,01	0,005	0,006	0	0	0	0	
Berilyum	-	-	0,004	-	0	0	0	
Birincil Standartlar (Radyolojik) pCi/l								
Gross Alfa	1	2,7	15					
Gross Beta	10	27	50					

İkincil standartlar (Estetik)								
Klorür	600	250	250	250	66	30	35	13
Renk birimi	20	15	15	-	3,1	2,5	2,5	0
Bakır	3	-	1	2	0,002	0,002	0,002	0
Deterjanlar	0,2	-	0,5	-				Yok
Demir	0,2	-	0,3	0,2	0,003	0,006	0,000	0,02
Mangan	0,05	0,5	0,05	0,05	0,003	0,005	0,001	0,006
Koku yapan madde								
Ph	6,5-9,	6,5-8	6,5-8	6,5-9	7,4	7,4	7	7,7
Sülfat	250	250	250	250	80	47	34	
Top.Çöz.Madde								
Çinko	5	-	5	-	0,004	0,014	0,003	0,19
İlave değişkenler, µg/l								
Kalsiyum	200	-	-	-	45	49	45	125
Sertlik (CaCO ₃ olarak)	-	500	-	-	166	146	160	170
Magnezyum	50	-	-	-	12	4	4	
Potasyum	12	-	-	-	3,89	2,53	1,81	
Sodyum	175	200	-	200	38,87	14,87	12,34	
Serbest Klor.	0,5	5	-	-	1,2	1,4	1,3	0,68
Amonyum	0,5	1,5	-	0,5	<0,05	<0,05	<0,05	0

İSKİ 37 değişkenle, DİSKİ ise 18 değişkenle su kalitesini ölçmektedir.

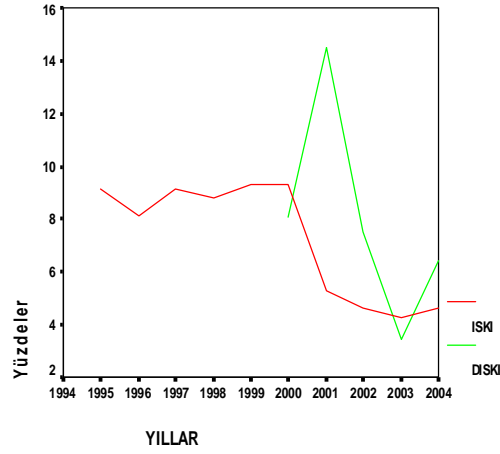
İSKİ ve DİSKİ'nin vatandaşa vermiş olduğu su kalitesi yukarıdaki değişkenlerle incelenmektedir. Yetkili mercilerce suyun kalitesinin doğrudan ya da dolaylı olarak tüketici sağlığını etkilemediğinin tespit edildiği durumlarda kullanılan sulara, sağlıklı su denir(12). Bu tanıma uygun olarak, suyun kalitesini incelerken, ölçülen tüm değişkenler açısından İSKİ ve DİSKİ'nin su ölçümleri sağlıklı su ölçümlerini vermektedir.

3.4 İSKİ ve DİSKİ'nin hizmet verdiği müşteri sayısı olarak durum incelendiğinde sonuçlar aşağıdaki gibidir.

Su canlı insana hitap eden, birinci derecede sağlıkla ilgili bir maddedir. Su hayattır. Suyun herhangi bir şekilde kirlenmesi sonucunda veya suyun niteliği nedeniyle oluşabilecek olumsuz etkiler hakkında tüketici nüfus haberdar edilerek insan sağlığının korunmasını sağlayacak tavsiyelerde bulunmak ve gerekli önlemler almak gerekir.

Tablo 4. İSKİ ve DİSKİ nin Yıllara göre Müşteri adetleri ve artış yüzdeleri

Yılı	İSKİ		DİSKİ	
	Müşteri Adedi	Artış % si	Müşteri Adedi	Artış %si
1994	1,715,280			
1995	1,886,003	9,16		
1996	2,031,908	8,1		
1997	2,211,945	9,14		
1998	2,404,331	8,80		
1999	2,631,400	9,30	90,844	
2000	2,877,404	9,30	98,180	8,08
2001	3,032,502	5,30	112,436	14,52
2002	3,179,118	4,60	120,893	7,52
2003	3,320,728	4,26	125,037	3,42
2004	3,482,752	4,65	133,070	6,42

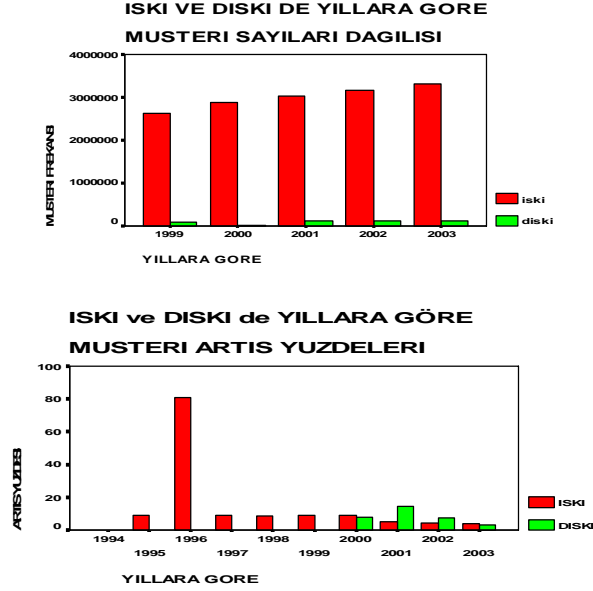
İSKİ ve DİSKİ MUSTERİ ARTIŞ YUZDESİ

NOT: İSKİ'nin 2004 yılı verileri 10 aylıktır.

İstanbul'da sadece su aboneli alan insan sayısı 3,482,752 mekan bulunmaktadır. Bu rakam Diyarbakır için 133,070 mekan olarak gözlenmektedir. Bu sonuçlar oranlandığında, Diyarbakır kentinin İstanbul'un % 3.8'i kadar olduğu görülür. İstanbul tek başına Türkiye nüfusunun % 17'sidir. Bu büyüklükte kentleri kontrol etmek zordur. İstanbul'un 12 milyonluk nüfusu

baz alındığında ve su abone oranına göre hesaplandığında Diyarbakır'ın nüfusunun 459 bin dolayında olduğu görülmektedir.

İSKİ ve DİSKİ' nin yıllara göre müşteri sayılarına ve müşteri artış yüzdelere ilişkin grafikler aşağıda verildi.



Grafik 1. İSKİ ve DİSKİ 'de Yıllara Göre Müşteri Sayıları ve Artış Yüzdeleri

3.5 İSKİ ve DİSKİ'nin müşteri hizmetlerinden aldıkları şikayet sayısı sonuçları aşağıda verildi.

İstanbul ve Diyarbakır halkının kullandığı içme ve kullanma suları hakkında bilinçlendirilmesi gerekmektedir. Bu nedenle ilgili kamu ve sivil toplum örgütlerinin bu konuda çalışmalarda bulunması yararlı olur. Müşterilerin yaptıkları şikayetlerle ilgili veriler aşağıda sunulmuştur.

İSKİ de basılan faturalara yapılan itirazlar binde 3 ile binde 9 arasında değişim gösterir iken, bu oranlar DİSKİ de yüzde 1,66 ile 2,46 arasında değişmektedir.

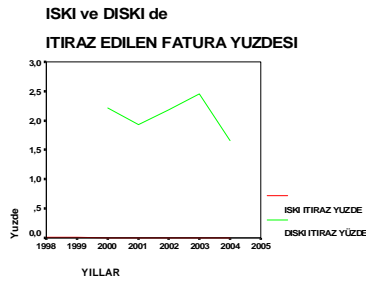
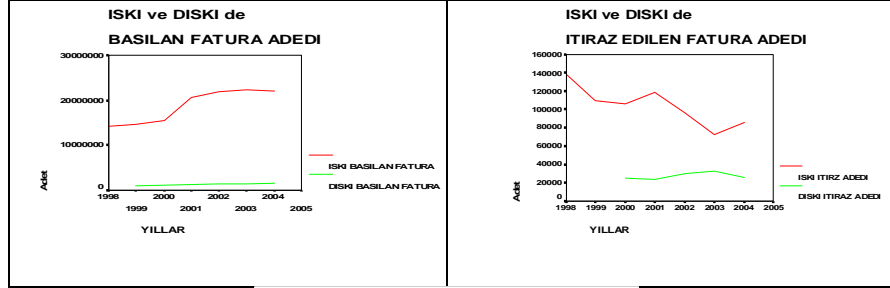
3.6 DİSKİ de Yıllara Göre Kaçak Su Miktarları ve Tahakkuk Eden Ceza Tutarı

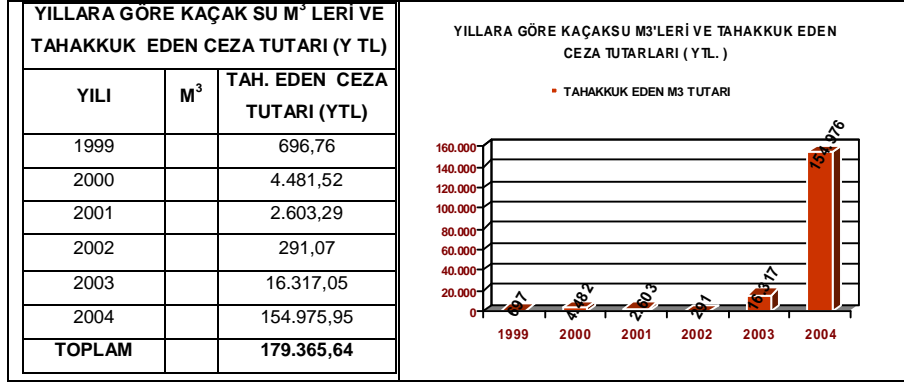
DİSKİ, İstanbul'a göre suyu ucuza vermektedir. İstanbul gibi Diyarbakır da çok göç almıştır. Diyarbakır'a gelen göç dalgası daha çok köylü

kökenli olduğu için suyun bedava olduğu anlayışı gözlenmektedir. Vatandaş suyun ücretini ödemek istememektedir. Yıllara göre ceza kesilen miktarlar YTL olarak aşağıdadır.

Tablo 5. İSKİ ve DİSKİ'nin yıllara göre müşterilerin idareye yapmış olduğu itiraz sonuçları

Yılı	İSKİ			DİSKİ		
	Basılan Fatura	İtiraz Adedi	İtiraz % si	Basılan Fatura	İtiraz Adedi	İtiraz % si
1998	14,259,296	137,974	0,009			
1999	14,658,196	109,943	0,007	1,022,794		
2000	15,438,481	106,249	0,005	1,101,897	24,468	2,22
2001	20,614,601	118,786	0,005	1,195,921	23,118	1,93
2002	21,918,821	96,253	0,004	1,394,577	29,490	2,19
2003	22,389,722	72,948	0,003	1,330,913	32,787	2,46
2004	22,009,176	85,644	0,003	1,525,745	25,370	1,66
2005						

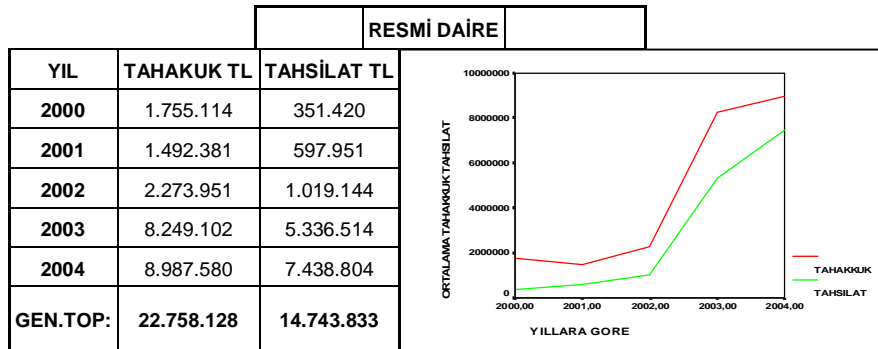


Tablo 6. DİSKİ de Yıllara Göre Kaçak Su Miktarları ve Tahakkuk Eden Ceza Tutarı

Diyarbakır'da Yıllara göre tahakkuk eden ceza tutarları 1999 yılında 696,76 YTL iken, bu miktarlar düzenli artarak 2004 yılında 154,975,95 YTL ye kadar çıkmıştır. Altı yılda toplam olarak vatandaşa kesilen ceza miktarı 179,365 .64 YTL para cezasıdır.

3.8 DİSKİ De Yüksek Tüketim Müşteriler Müdürlüğü Resmi Daireler Ait Tahakkuk Tahsilat Raporu

Diyarbakır'da sadece vatandaş değil resmi kurum ve kuruluşlar da su borçlarını ödememektedir. Kırsal kesimden gelen vatandaşların su ücretini ödememe mantığının anlaşılır sebepleri olabilir. Ama Diyarbakır şehrinde yaşayan resmi kurum yönetici ve bürokratlarının bu davranışını anlamak zordur.

Tablo 7. DİSKİ de Yüksek Tüketim Müşteriler Müdürlüğü Resmi Daireler Ait Tahakkuk Tahsilat Raporu

Resmi daireler su ücretlerinin az bir kısmını ödemektedir.

4. Tartışma

Vücudumuzun su ihtiyacı nasıl karşılanır? Günde ne kadar su içmeliyiz? Hijyen, sanitasyon, temizlik nedir? Dezenfeksiyon nedir? Patojen mikroorganizma nedir? Sağlıklı su nedir? İnsan sağlığında suyun önemi nedir? Tüketicie ulaştırılan suların kalitesi nelere bağlıdır? Sular nasıl kirlenir? Suların dezenfeksiyonu nedir? İçme suyunun dezenfeksiyonunun amacı nedir? Dezenfeksiyonun insan sağlığındaki rolü nedir? Suyun dezenfeksiyonu nasıl yapılır? Suların sertliği neye bağlıdır? Sularda sertliğe yol açan başlıca iyonlar nelerdir? İnsani tüketim amaçlı su nedir bu kapsamdaki sular nelerdir? pH nedir? Kaynamış su arınmış mıdır? Klor sağlığa zararlı mıdır? Suda hangi analizler yapılmalıdır? Sular ile bulaşan hastalıklar nelerdir? Sağlık Bakanlığının tüketiciye sağlıklı su ulaştırmadaki görev ve sorumluluğu nedir? Su ile ilgili tesislerin denetiminin insan ve çevre sağlığı açısından önemi nedir? Gibi soruları çoğaltabiliriz. Fiyatlar incelendiğinde, İSKİ'nin koymuş olduğu su tarifesi ortalama 3 YTL dolaylarında iken, DİSKİ'nin su fiyatları 0,72 YTL den başlamakta 2.5 YTL ye kadar çıkmaktadır. Büyük şehirlerde su sorunu nasıl çözülecek? Şehirlerdeki en büyük fabrikalar şehre su sağlayan fabrikalardır. Elektronik izleme sistemleri ile karbon temizliği ile ozon ile yüzlerce kirletici maddeyi temizleyen teknolojiler kullanılmaktadır. Gereksiz partiküller temizlenir. Organik bileşikler kurşun ve civa'ya, benzer maddeleri yok eder. Büyük karbon filtreler aracılığıyla suyu temizler. Yüzlerce kirleticiyi yok ederek, tat/koku giderir. Organik ve inorganik temizlik yapılır, berraklık sağlanır. İçme suyunda hastalıklara neden olan bakteri ve virüsleri % 80-% 90'dan fazla bir oranda yok edilir. Bakteri-mantar-alg toksini-kist-virus gibi mikroorganizmalar yok edilir. Bunu kim denetler? Temizlenen ve şehre verilen suyun bağımsız denetleme kurum ve kuruluşlarınca denetlenmesi gerekmektedir. Devlet Su İşleri(DSİ), Üniversiteler, sivil toplum kuruluşlarıdır. Uluslar arası standartlar nedeniyle taraf olan, NSF International **Su artımında dünya çapında tarafsız, bağımsız üçüncü taraf örgütü** dür. Ayrıca, *Water Quality Association Gold Seal Certification Program: *Su Kalite Derneği, kaliteli su arıtma sektörünü temsil eden, kar amacı gütmeyen uluslararası ticaret derneğinin* önereceği standart belediyeçilik hizmeti için gerekenler şunlardır.*

1. *Daha temiz, daha duru ve daha iyi tadı olan bir su*
2. *Daha iyi su kalitesi*
3. *Kanıtlanmış performans*
4. *Bilgi ötesi çağın, ileri teknolojisi ve vizyon sahibi olmak*

5. *Yıllar boyu süren araştırmaların sonucu kabul edilmiş uluslar arası standartlarda su sağlamak,*
6. *Vatandaşa büyük kolaylık sağlamak*
7. *Vatandaşı, daha az endişelendirecek bilgi vermek.*
8. *Daha güvenilir, daha değerli hizmet vermektir.*

Günümüzde artık bilgi çok ucuzladı. Belediyelerimizin sahip olması gereken şey şudur: Vizyon.

Kentinden, köyünden göç eden insanlar geldikleri şehirlere ayak uydurmalıdırlar. Ya da şehir yöneticileri kentin idare ve yönetim kurallarını kesin olarak uygulamalıdırlar. Yüzylerce yıl süren feodal dönem çok gerilerde kaldı. Kapitalizm iki yüz yıl sürdü, emperyalizm devam ediyor. Dünya bilgi çağını 1985 yılında yakaladı. 15 yıl süren bir bilgi çağı da artık gerilerde kaldı. Şimdi zaman bilgi ötesi çağını gösteriyor. Globalizm, ahlak ve etiğini beraber getirmiştir. Ahlak toplumu tarif ederken, geçmişi ve özelemlerimizi hatırlatmaktadır. Eskiden herkes köyünde temiz suyunu içerdi. Etik ise, birey ve insan haklarını anlatmaktadır. Canlıdır. İnsandır ve yaşıyor. Kural olarak susayan bir insana bir bardak suyu çok göremezsiniz. Etik budur. İnsana sağlıklı su vermeniz gerekir. Toplumlar su gibi birleşmiştir. Okyanuslara dönüşmüştür. Artık mahalli değerler ve özelemler nostaljik kavramlar olarak kalmaktadır. Gökyüzünde uçan kuşlar gibi bireyler de sınır ve toprak tanımamaktadır. Her vatandaş ben de bireyim diyor. Temiz su ve temiz bir yaşam istiyor. Bu durumda belediyelerin yönetimlerine talip olacak insanların artık yeniyi temsil etmesi yetmiyor. Bilgili olacak, vizyon sahibi olacak, çağımız bunu gerektiriyor. Siz bilgiye çok kolay ulaşabilirsiniz. Bir internet kafeye gidiniz istediğiniz bilgi ve tanımları bulabilirsiniz. Siz yeniyi temsil etmezseniz, hala eski usul ve yöntemlerle düşünmeye devam ederseniz, vatandaşınıza su (hayat) sunamazsınız. Şehirlerde toplu ve yaygın hastalıklarla karşılaşabilirsiniz. Artık vatandaş suyunu kim kirletti demiyor. Niçin bana temiz su vermiyorsunuz diye soruyor. Suyun nasıl kirlendiği, nereden geldiği, nasıl temizlendiği ile kimse ilgilenmiyor. Niçin sorusunu sormanın zamanıdır. İstanbul NİÇİN KALABALIK OLUYOR? Toprağından kopan, İstanbul'a geliyor. İstanbul'da SU NİÇİN PAHALIDIR? İSKİ'nin yeni etkinlik alanları genişletiliyor. Neden? NİÇİN yeni göçler devam ediyor? Yeni semtler hangi maliyetlerle İSKİ'nin bütçesine yansiyacaktır. Fiyatlar incelendiğinde, İSKİ'nin koymuş olduğu su tarifesi ortalama 3 YTL ile birim ölçümler uygun iken, DİSKİ'nin su fiyatları 0,72 YTL ile düşük fiyatta kalmaktadır. Buna rağmen vatandaşın su faturalarına itirazı devam etmektedir. Su ücretini ödemek vatandaşın ağır gelmektedir. Benzer davranış kamu kurum yetkililerinde de gözlenmektedir.

Vatandaş gibi kamu kurum ve kuruluşları da su borçlarını ödemektedir. DİSKİ yöneticilerinin feryatları yakındır. Dört tarafı su olan

İSTANBUL su bulmakta güçleniyor. İSTANBUL 3,482,752 su abonesi ile insan hayalini zorlayacak büyüklüktedir. Ayda abone başına 10 YTL alınsa, 30,482,752 YTL cinsinden gelir elde edilir. Abone başına ayda 100 YTL alınsa, 300,482,752 YTL lik gelir elde edilir. DİSKİ için en az ayda 1.330.700 YTL gelir elde ediliyor. Bu konuda çok spekülasyon yapılabilir. Gerçek anlamda İSTANBUL ve DİYARBAKIR KENTLERİ ile ilgili araştırma yapılması gerekmektedir. Köylülük İSTANBUL şehrine de kimliğini-damgasını basmıştır. Yeniden "BARBAR TOPLUM" olma toplumsal dönüşümü bizi çağırılmaktadır. İSTANBUL ve DİYARBAKIR'ın çok sayıda suçlu barındırmasının bir nedeni de göçtür. İSKİ laboratuvarlarında 37, DİSKİ'de ise 18 farklı değişkenle suların analiz edilmesi, mukayese edildiğinde DİSKİ nin yetersizliği gibi gözükabilir. Asıl dikkat edilmesi gereken İstanbul nüfusunun Türkiye'nin bütçesini yutan bir canavar haline geldiği gerçeğidir. Yapılan analiz neticeleri İstanbul'un musluk suyunun Avrupa Birliği (EC), Dünya Sağlık Teşkilatı (WHO), ABD Çevre Koruma Ajansı (EPA)'nın ön gördüğü kriterlerde olduğunu gösteriyor. Ama ozonla dezenfekte edilen İSTANBUL suları çok iyi bir altyapıyı gerektirmektedir. Her ne kadar şimdilik kayıp su oranı % 25-30 civarında ise de yakın bir gelecekte nasıl olacağını kestirmek zordur. Kendimizi evimizde şişe suyu satıcılarının kucağında bulabiliriz. Yakın bir gelecekte, su dezenfeksiyonu yapan küçük ev aletleri tüm varlıklı İstanbulluların hayatına girecektir. İstanbul büyük kent olmanın verdiği avantajlarla bu gün için su kaçaklarını azaltmıştır. Yakın gelecekte en çok mücadele edeceği konu, su kaçakları olacaktır. Alt yapı eksikliği, yeni yerlerin katılmasının yarattığı yetersizlikler, kullanılan ileri teknoloji İstanbul'u çok erken eskitecektir. İSKİ ve DİSKİ gerçek anlamda gevşek davranmaktadır (Slow-business): şimdilik su satışlarından para kazanıyor olabilirler. Ama yakın bir gelecekte yeni yatırımları karşılayamayacak duruma düşebilir. Böyle devam ederse, ya İstanbul büyüyecek TÜRKİYE olacak, yada İstanbul'da insanlık çözülecektir. 1995'ten 2000 yılına kadar her yıl İstanbul'da abone sayısı % 10 artmıştır. 2000 yılından sonra da her yıl abone sayısı % 5 artmıştır. Devlet büyüklerinin bu artışı dengelemelerinin zamanı çoktan geçmiştir. Makro plan çözümsüzlük olamaz. Bu devasa büyüme bizi rahatsız etmeyebilir. Bu büyümeden hoşlanabiliriz de, Türkiye ekonomisi 2004 yılında % 9.9 büyümüş olabilir, hatta Çin'i de sollamış olabilir. Yıllık gelir kişi başına 4,172 \$ a da çıkmış olabilir. Faiz oranları % 70'ten, % 8'e de inmiş olabilir. Dünyadaki global değişimin asla unutmamamız gerekir. Artık temel doğal yasaları tanımamızın zamanı gelmiştir. Ölçüm ve değerlendirmeleri birimler üzerinden yapmamız gerekiyor. Hiçbir ekonomik durum ve gelişme, insan ile insan ilişkisini ve insan ile doğa ilişkisini koparmaz. Paris'in nüfusu yıllardır sabittir. Florida'nın da yıllardır sabittir. Stocholm'un da yıllardır nüfusu sabit kalmıştır. Lonra'yı Madrid'i Münih'i inceleyin böyle bir olay göremezsiniz. İstanbul NİÇİN büyüyor? İlgililerin, topluma bu sorunun cevabını vermesi gerekiyor? Toplum, insanlarımız kendine bunu soruyor. Diyarbakır şehir nüfusu 1990 yılında

330.000 kişiden on yılda nasıl olur da 1.300.000 kişiye çıkıyor? (Oysa abone durumları incelendiğinde nüfus oranları bir birini tutmamaktadır.) Eğer biri, topluma şunu açıklayabilirse, çok çarpıcı bir açılama olacaktır. İstanbul ucuz iş gücü temin etmiş olabilir. Globalizmin sonucu olarak "pazar-liberalleşebilir". Sonuç olarak, şirketler güç kaybeder, kişiler aktif hale de gelebilir. Aşırı nüfus, pazarın ana itici gücü olabilirler. İyilik ve para bir harmoni içinde uyum da sağlayabilir. Su tüketimi ve ihtiyacı, hatta su dağıtım ve pazarlama işi yeni iş kollarını da beraberinde getirebilir. Global çözümleri hep beraber göreceğiz. Devasa büyük bir fakir nüfus-hemen yanında zengin bir azınlık genişleyen İstanbul şehri. Ulaşım, konut, eğitim benzeri problemler İstanbul'un yakasını bırakmayacaktır. Bu tempolu koşu daha uzun yıllar İSTANBUL-DİYARBAKIR kentlerini meşgul edecektir. Dahası, sorumluluğu sadece ekonomiye, politikaya yükleyemeyiz. Kültürel kayıplar-yozlaşmalar olacak, piyasalarda manken insanlar çoğalacaktır. Kentlerimizin büyüme hızına ve yaratacağı problemlere vatandaşlarımız yetişemeyecektir. Bireyler çok acımasız bir rekabet ortamında kalacak ve acımasızlaşacaktır. Toplum İslam kökenlidir. Belki yakın gelecekte, insanlarımızın inançlarında olmayan intihar olayları çoğalabilir. Toplumsal stres artmıştır. On beş yıldan bu yana bu iki kentin nüfusuna gelip katılan insanlar, kendi köylerinde oturuyordu. Bölgesel kent kültürü ve üretim ilişkileri içinde idiler. Ekmeğini, etini, sütünü kendileri karşılıyordu. Gelenlerin büyük çoğunluğunun eğitim ve bilimsel seviyeleri eski İstanbul-Diyarbakır kökenlilerin çok gerisindedir. Bu toplumsal katılım, ilerleten değil, topluma ve şehre yük bindiren bir sonuçtur. Umarım, şehir yöneticileri su ve suyla ilgili hizmetleri hiç unutmazlar. "Bu benim işim değil" "bana ne diyecek", bir anlık bir unutkanlık, İstanbul ve Diyarbakır kentleri için yıkım olacaktır. İki kentimizde de vatandaş, eskimiş dahili tesisatlarını yenileyerek ve apartman depolarını temizleterek Dünya ve Avrupa Birliği kalitesinde şehirlere verilmekte olan içme suyunu güvenle içebilirler.

Kaynaklar

1. Paula Schmittiel Eureka Milles Eureka,Utah Risk Assessment (USA Environmental Protection Agency EPA Remedial Project Manager 1-800-227-8917 ext. 6861) FACT SHEET November 2001.
2. King, WD; Marrett, LD, Case-control study of bladder cancer and chlorination by-products in treated water (Ontario, Canada). Source Cancer Causes and Control, 1996, 7, 596-604.
3. Bull RJ; Birnbaum LS; Cantor KP; Rose JB; Butterworth BE; Pegram R; Tuomisto J **Water chlorination: essential process or cancer hazard?** Source Fundam Appl Toxicol, 1995 Dec, 28:2, 155-66.

4. McGeehin MA; Reif JS; Becher JC; Mangione EJ **Case-control study of bladder cancer and water disinfection methods in Colorado.** Source Am J Epidemiol, 1993 Oct 1, 138:7, 492-501.
5. Kramer MD; Lynch CF; Isacson P; Hanson JW, **The association of waterborne chloroform with intrauterine growth retardation.** Source Epidemiology, 1992 Sep, 3:5, 407-13.
6. Bove FJ; Fulcomer MC; Klotz JB; Esmart J; Dufficy EM; Savrin JE, **Public drinking water contamination and birth outcomes.** Source Am J Epidemiol, 1995 May 1, 141:9, 850-62.
7. Williamson SJ, Epidemiological studies on cancer and organic compounds in U.S. drinking waters. Source Sci Total Environ, 1981 Apr, 18:, 187-203.
8. Vanderslice RR; Orme J; Ohanian EV; Sonich-Mullin C, **Problems in assessing the risks of mixtures of contaminants in drinking water.** Source Toxicol Ind Health, 1989 Oct, 5:5, 747-55.
9. Cantor KP; Hoover R; Mason TJ; McCabe LJ, **Associations of cancer mortality with halomethanes in drinking water.** Source J Natl Cancer Inst, 1978 Oct, 61:4, 979-85.
10. Marienfeld CJ; Collins M; Wright H; Reddy R; Shoop G; Rust P, **Cancer mortality and the method of chlorination of public drinking water: St. Louis City and St. Louis County, Missouri.** Source J Environ Pathol Toxicol Oncol, 1986 Sep-Dec, 7:1-2, 141-57.
11. P.W. Mc Randle, Consider ist Lifecycle: Bottled Water. The Gren Guide, (<http://www.keepmedia.com/pubs/GreenGuide/2004/03/01/512686?extID=10026>) Sub 28 2004.
12. Resmi Gazete: 17 Şubat 2005 - 25730 (<http://www.ttb.org.tr/mevzuat/modules.php?name=News&file=article&sid=265>).

YARIŞAN RİSKLER İÇİN COX REGRESYON MODELİ

Durdu Sertkaya, Birce Ünal

Hacettepe Üniversitesi Fen Fakültesi İstatistik Bölümü

e-mail: birceunal@excite.com

Özet

Bu çalışmada, yaşam çözümlemesinde önemli olan yarışan riskler ve sebebe-özel hazard fonksiyonu konuları ele alınmıştır. İki başarısızlık türü olduğu durumda, iki farklı yöntem ele alınarak Cox regresyon modeli incelenmiş ve bir uygulama yapılmıştır.

Anahtar kelimeler: Yarışan riskler, Yaşam çözümlemesi, Cox regresyon, Başarısızlık türü, Durdurulmuş.

TÜRKİYE'DE BEKLENEN YAŞAM SÜRESİ: 1980–2001 YILLARI ARASINDA GERÇEKLEŞEN DEĞİŞİM

Nilgün Moralı, Esra Turan

Yaşar Üniversitesi, İstatistik Bölümü
e-mail: nilgun.morali@yasar.edu.tr

Özet

Bu çalışmada, 1980-2001 DİE ölüm istatistiklerine dayanarak hazırlanan hayat tablolarından elde edilen beklenen yaşam süreleri incelenerek Türkiye'de beklenen yaşam süresinin 20 yıl içinde nasıl ve ne kadar değiştiği araştırılmıştır. Bu değişiklik, 2001 yılı ile 1980 yılındaki beklenen yaşam süreleri arasındaki fark alınarak hipotez testleri ile hesaplanmıştır. Bu testlerin sonucunda, Türkiye'de 1980 yılından 2001 yılına kadınlarda, erkeklerde ve her iki grup birlikte düşünüldüğünde beklenen yaşam süresinin artmış olduğu bulgusu ortaya çıkmıştır. Ayrıca yıllara göre beklenen yaşam süresinin hangi yaşlarda arttığı da belli bir yaş için ardışık yılların farkları alınarak hipotez testleri uygulanmış, 0 yaş için beklenen yaşam süresinin 1980 yılından 2001 yılına doğru giderek arttığı, fakat diğer yaşlarda yıllara göre böyle bir artış görülmediği gözlemlenmiştir.

Anahtar Kelimeler: Demografi, Yaşam Analizi.

Giriş

Son yüzyılda gerçekleşen hızlı teknolojik değişime paralel olarak sağlık sektöründe yaşanan gelişmeler ve yaşam standartlarının yükselmesi genel olarak tüm dünyada insan ömrünün bir önceki yüzyıla oranla uzamasına neden olmuştur. Birleşmiş Milletler Kalkınma Programı'nın 2003 tarihli İnsani Gelişim Endeksi verilerine göre, doğumda beklenen ortalama yaşam süresinin dünya ortalaması 66,7 olduğu hesaplanmıştır. (UNDP-Human Development Index, 2003). Aynı raporda, 2003 yılı için Türkiye'de ortalama yaşam süresinin 70.1 ile dünya ortalamasının üzerinde seyrettiği ortaya konmuştur. Literatürde ülkemizde ortalama yaşam süresinin gelecekte de artacağı yönünde çeşitli çalışmalar bulunmaktadır (Türkiye İşveren Sendikası Konfederasyonu (TİSK), 2003). Devlet Planlama Teşkilatı'nın (DPT) 2002 tarihli bir çalışmasında, Türkiye'de ortalama yaşam süresinin

1995 yılında 67,4 olarak belirlenirken; 20 yılda 5-6 yaş artarak 2020 yılında 73 yaşa çıkacağı hesaplanmıştır (DPT, 2002).

Kocaman, ülkemizde ortalama yaşam süresinin artarak 1995–2000 arası dönemde 68,55'e ulaştığını belirtmiştir; fakat bu çalışmada ortaya çıkan bulguların bir hipotez testi ile desteklenmediği göze çarpmaktadır (Kocaman, 2004).

Bu çalışmanın amacı, Türkiye'de ortalama yaşam süresinin nasıl değiştiğini ve istatistiksel yöntemler kullanarak ortaya koymaktır. Bu konuda çeşitli gazete ve dergilerde sık sık yazılar yayınlansa da bunlar bilimsel açıdan yetersiz kalmakta ve bu çalışma için önemli bir referans değeri taşımamaktadır.

Veriler

Bu çalışmada, Devlet İstatistik Enstitüsü'nün (DİE) 1980–2001 arası dönemi kapsayan yıllar için yaş ve cinsiyete göre ölüm istatistiklerinden yararlanılmıştır (DİE, 1980–2001 Türkiye Ölüm İstatistikleri).

Yöntem

Çalışmamız üç aşamadan oluşmaktadır. Öncelikle 1980–2001 DİE ölüm istatistiklerine dayanarak her yıl için ayrı ayrı hayat tabloları hazırlanmıştır. Bu hayat tabloları her yıl için kadınlar tablosu, erkekler tablosu ve toplam olarak üç tablodan oluşturulmuştur (Moralı, 1997:59-89).

İkinci aşamada, elde edilen hayat tablolarından 1980 ve 2001 arası dönemdeki tüm yıllarda kadınlar, erkekler ve toplam için beklenen yaşam süreleri bulunmuş ve beklenen yaşam süresinde ki değişiklikler öncelikle grafikler çizilerek incelenmiştir (Menge ve Fischer, 1985: 13).

Çalışmamızın son aşamasında, ortalama ömrün ne kadar değiştiğini bulabilmek için 2001 yılındaki beklenen yaşam sürelerinden 1980 yılındaki beklenen yaşam süreleri arasındaki fark alınarak t testi uygulanmıştır. Ayrıca yıllara göre beklenen yaşam süresinin hangi yaşlarda arttığı, belli bir yaş için ardışık yılların farkları alınarak t testi ile bulunmuştur.

Bulgular

Türkiye'de toplam nüfus için 1980–2001 arası dönemi kapsayan yıllarda belirli bir yaştaki birey için hesaplanan hayatta kalma süreleri incelendiğinde, 0 yaşındaki bir bireyin 1980 yılındaki beklenen ömrü yaklaşık 43 yıl iken 2001 yılında yaklaşık 59 olmaktadır (Ek-1). 1980 yılında

kadınların doğuştan hayatta kalma süresi 46, erkeklerde 42 olarak görülürken; bu süre 2001 yılında kadınlarda 62'ye erkeklerde 57'ye yükselmiştir. Bu bulgular ışığında, bebek ölümlerinin 1980'den 2001'i kapsayan bir dönemde azaldığı söylenebilir. Doğumda beklenen hayatta kalma süresi büyük artış göstermesine rağmen diğer yaşlar incelendiğinde bu kadar büyük artış görülememektedir (Ek-2).

Türkiye'de ortalama ömrün arttığını gösterebilmek için, 1980 ve 2001 yıllarında beklenen yaşam sürelerinin aynı olduğu hipotezine karşıt olarak 2001 yılında beklenen yaşam süresinin 1980'den daha fazla olduğu hipotezi öne sürülmüştür. Bu hipotez test edildiğinde % 95 güvenle 2001 yılında beklenen yaşam süresinin 1980 yılından fazla olduğu sonucuna varılır (Tablo 1).

Bu bulgudan sonra ortalama ömrün ne kadar arttığının da saptanması gerekir. Bu bulguyu bulabilmek için 2001 yılında beklenen yaşam süresinin 1980'den 1 yıl, 2 yıl, 3 yıl, ... daha fazla olduğu alternatif hipotezleri öne sürülmüştür. Bu hipotezlerin testi sonucunda 2001 yılında beklenen yaşam süresinin 1980 yılından 1 yıl daha uzun olduğu söylenilebilir.

Tablo 1. Ortalama ömrün artmasına ilişkin hipotez testleri (Toplam)

Hipotez	t değeri	P değeri	Sonuç
$d = 0.000$ vs $d > 0.000$	8,21	*0,000	$d > 0$
$d = 1.000$ vs $d > 1.000$	2,95	*0,002	$d > 1$
$d = 2.000$ vs $d > 2.000$	-2,31	0,988	$d > 2$ değil

Yıllara göre bir kadının beklenen yaşam süresi hayat tabloları yardımıyla hesaplanmıştır (Ek-3). Bu veriler kullanılarak Türkiye'de kadınlar için beklenen yaşam süresinin ne kadar arttığı bulabilmek için, 2001 yılında beklenen yaşam süresinin 1980'den 1 yıl, 2 yıl, 3 yıl, ... daha fazla olduğu alternatif hipotezleri öne sürülmüştür. Bu hipotezlerin testi sonucunda kadınların 2001 yılında beklenen yaşam süresinin 1980 yılından 1 yıl daha uzun olduğu söylenilebilir.

Tablo 2. Ortalama ömrün artmasına ilişkin hipotez testleri (Kadın)

Hipotez	t değeri	P değeri	Sonuç
$d = 1.000$ vs $d > 1.000$	2,48	*0,008	$d > 1$
$d = 2.000$ vs $d > 2.000$	-2,55	0,994	$d > 2$ değil

Yıllara göre bir erkeğin beklenen yaşam süresi hayat tabloları yardımıyla hesaplanmıştır (Ek-4). Bu veriler kullanılarak, Türkiye’de erkekler için beklenen yaşam süresinin ne kadar arttığı bulabilmek için, 2001 yılında beklenen yaşam süresinin 1980’den 1 yıl, 2 yıl, 3 yıl, ... daha fazla olduğu alternatif hipotezleri öne sürülmüştür. Bu hipotezlerin testi sonucunda kadınların 2001 yılında beklenen yaşam süresinin 1980 yılından 1 yıl daha uzun olduğu söylenilebilir.

Tablo 3. Ortalama ömrün artmasına ilişkin hipotez testleri (Erkek)

Hipotez	t değeri	P değeri	Sonuç
$d = 1.000$ vs $d > 1.000$	3,31	0,000	$d > 1$
$d = 2.000$ vs $d > 2.000$	-1,04	0,85	$d > 2$ değil

Çalışmada beklenen yaşam süresinin yıllara göre hangi yaşlarda arttığını bulabilmek için 0 yaş, 10 yaş, 20 yaş, ... ında olan bir bireyin yıllara göre beklenen yaşam süresi artış göstermemiştir hipotezine karşılık arttığı hipotezi öne sürülmüştür. Bu hipotez testlerinin sonucunda 0 yaşında kadın, erkek ve toplam için yıllara göre beklenen yaşam süresi artmaktadır. Buna karşılık diğer yaşlarda yıllara göre böyle bir artış görülmemektedir.

Tablo 4. 0 yaş için yıllara göre ortalama ömrün artmasına ilişkin hipotez testleri

	Hipotez	t değeri	P değeri	Sonuç
Toplam	$d = 0$ vs $d > 0$	4,98	0,000	$d > 0$
Kadın	$d = 0$ vs $d > 0$	4,35	0,000	$d > 0$
Erkek	$d = 0$ vs $d > 0$	5,43	0,000	$d > 0$

Sonuç

Bu çalışma da 1980–2001 yılları arası Devlet İstatistik Enstitüsü’nden elde edilen ölüm istatistikleri kullanılarak Türkiye’de ortalama ömrün nasıl değiştiği araştırılmıştır. Çalışmada DİE nüfus ve doğum istatistiklerini de kullanarak hesaplar yapsaydık, hayat tabloları daha geniş olacak ve değerler daha güvenilir olacaktı.. Buna karşılık çalışmada amaçlanan ortalama ömüre ilişkin sabit bir değer bulmak değil, bulunan verilerin yıllara göre tutarlı olmasını sağlamaktır. Nüfus ve doğum istatistikleri verilerinde süreklilik bulunmamaktadır ve çalışmanın amacına hizmet etmemektedir. Türkiye’de 1980 yılından 2001 yılına kadınlarda, erkeklerde ve her iki grup birlikte düşünüldüğünde beklenen yaşam süresinin 1 yıl artmış olduğu bulguları ortaya çıkmıştır.

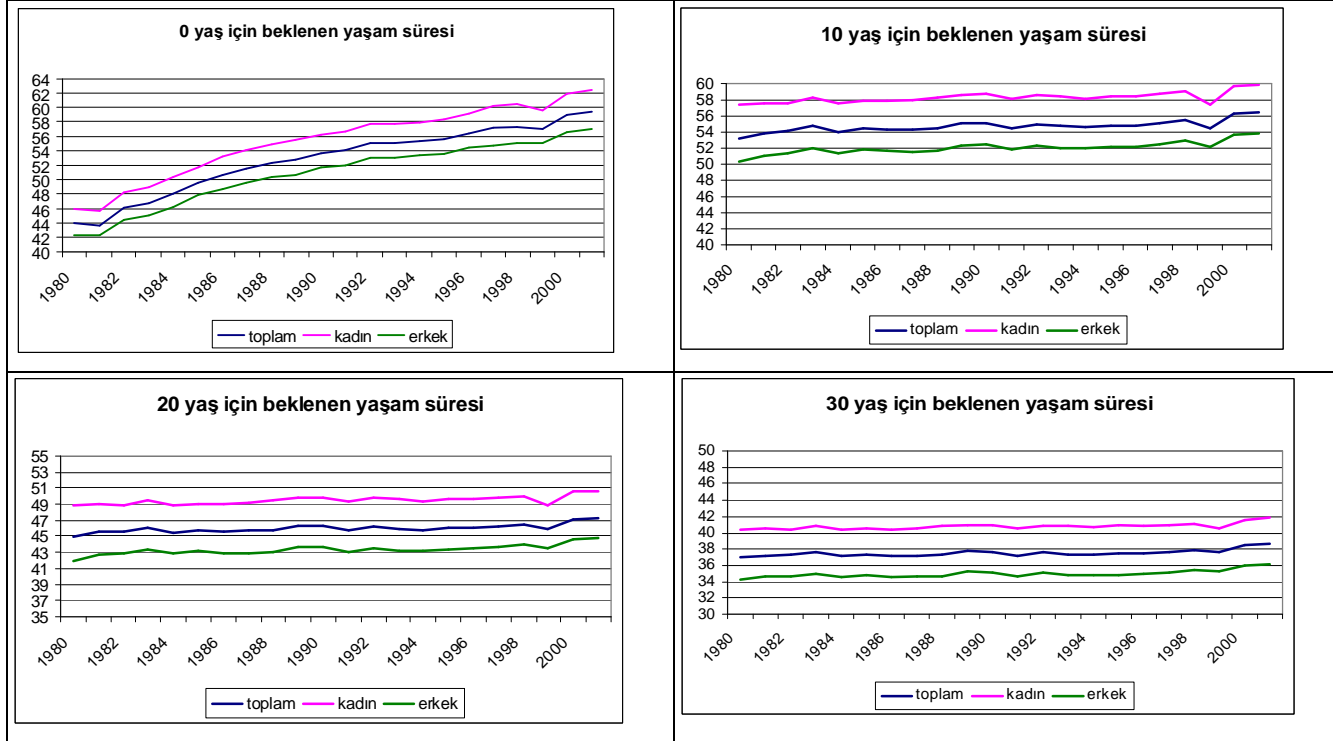
Bulunan diğler bulgular ise, 0 yaş için beklenen yaşam süresinin 1980 yılından 2001 yılına doğru giderek artar, ama diğler yaşlarda yıllara göre böyle bir artış görülmemektedir. Bu bulgular ışığında bebek ölümlerinin azaldığı sonucuna varılabilir.

EK 1

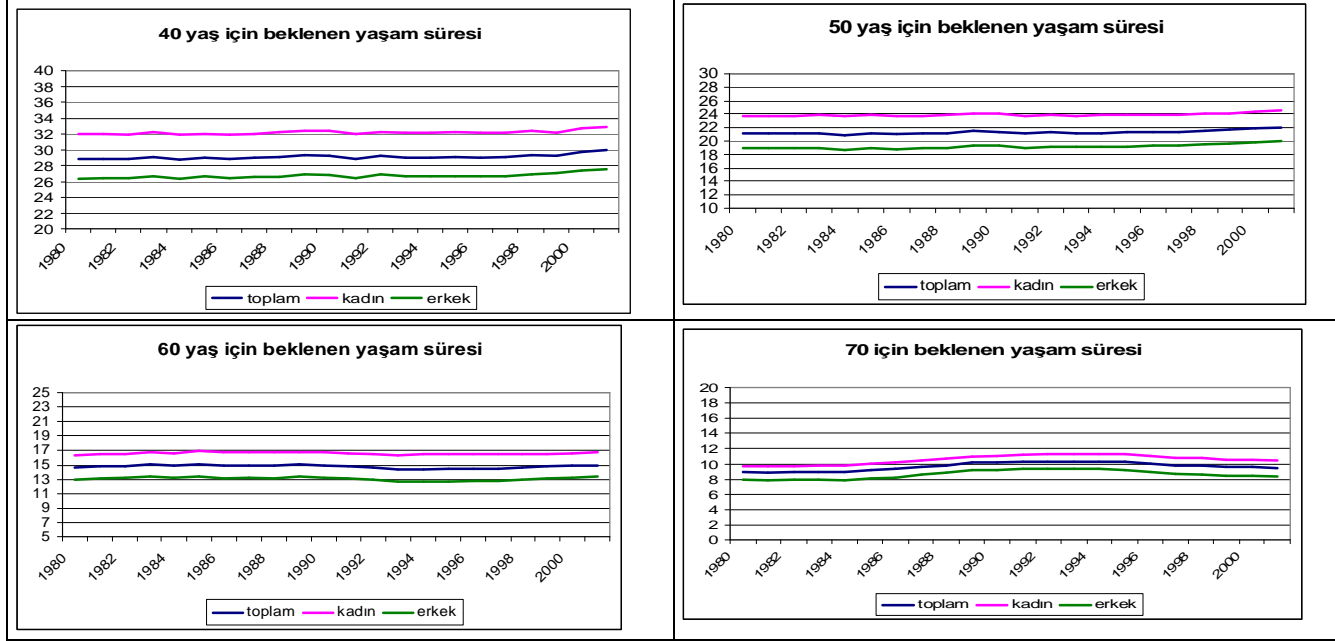
TÜRKİYE DE 1980–2001 YILLARI ARASINDA YAŞLARA GÖRE BEKLENEN ÖMÜR

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
0	43,94	43,69	46,13	46,73	48,04	49,51	50,65	51,55	52,29	52,75	53,65	54,05	55,08	55,09	55,27	55,60	56,43	57,09	57,38	57,00	58,92	59,38
5	57,12	57,62	58,17	58,66	58,04	58,46	58,49	58,38	58,67	59,27	59,42	58,86	59,34	59,16	58,95	59,28	59,31	59,65	60,01	58,68	60,80	61,03
10	53,30	53,91	54,19	54,74	54,03	54,41	54,32	54,28	54,51	55,04	55,14	54,53	54,99	54,77	54,57	54,84	54,83	55,13	55,51	54,39	56,25	56,46
15	48,94	49,66	49,85	50,31	49,70	50,04	49,89	49,89	50,05	50,58	50,63	50,07	50,48	50,28	50,09	50,30	50,27	50,54	50,91	50,01	51,57	51,80
20	44,94	45,49	45,62	46,01	45,45	45,74	45,58	45,67	45,78	46,28	46,32	45,76	46,18	45,93	45,77	45,98	46,09	46,23	46,55	45,88	47,19	47,33
25	40,98	41,33	41,49	41,83	41,32	41,54	41,36	41,44	41,56	41,99	41,96	41,46	41,92	41,70	41,57	41,79	41,79	41,92	42,19	41,75	42,83	42,99
30	36,96	37,19	37,30	37,58	37,08	37,30	37,09	37,22	37,30	37,72	37,66	37,20	37,59	37,36	37,30	37,46	37,46	37,57	37,82	37,58	38,42	38,58
35	32,90	33,03	33,12	33,36	32,90	33,09	32,93	33,06	33,15	33,49	33,42	32,98	33,34	33,12	33,11	33,21	33,22	33,26	33,47	33,33	33,98	34,15
40	28,81	28,87	28,90	29,14	28,73	28,99	28,82	28,95	29,06	29,33	29,26	28,89	29,20	29,01	28,98	29,07	29,01	29,07	29,28	29,23	29,72	29,89
45	24,83	24,86	24,94	25,10	24,76	24,96	24,79	24,95	25,02	25,20	25,24	24,96	25,18	25,02	25,04	25,08	25,07	25,13	25,31	25,32	25,66	25,81
50	21,15	21,12	21,17	21,24	20,85	21,18	21,04	21,17	21,21	21,48	21,42	21,17	21,28	21,16	21,23	21,30	21,30	21,38	21,57	21,62	21,86	22,05
55	17,95	17,83	17,88	17,97	17,72	17,92	17,79	17,82	17,73	17,86	17,89	17,72	17,82	17,59	17,71	17,67	17,69	17,78	17,97	18,09	18,30	18,48
60	14,63	14,71	14,84	15,03	14,87	15,08	14,87	14,95	14,92	15,01	14,91	14,78	14,65	14,29	14,38	14,46	14,50	14,52	14,62	14,73	14,91	15,00
65	11,30	11,41	11,59	11,91	11,94	12,26	12,37	12,47	12,59	12,62	12,59	12,42	12,30	12,20	12,13	12,11	12,03	11,79	11,71	11,69	11,89	11,91
70	8,90	8,84	8,89	8,98	8,93	9,17	9,27	9,60	9,85	10,13	10,24	10,29	10,35	10,36	10,37	10,34	10,05	9,84	9,77	9,56	9,56	9,44
75	6,81	6,96	7,16	7,18	7,08	7,15	7,10	7,15	7,28	7,35	7,48	7,62	7,87	8,12	8,30	8,47	8,50	8,38	8,39	8,14	8,03	7,72
80	5,09	5,11	5,26	5,27	5,22	5,40	5,52	5,66	5,66	5,74	5,65	5,67	5,63	5,70	5,75	5,98	6,12	6,27	6,55	6,61	6,64	6,59
85	4,09	4,19	4,39	4,40	4,13	4,11	4,04	4,03	4,09	4,24	4,39	4,47	4,59	4,56	4,46	4,48	4,51	4,34	4,49	4,43	4,57	4,74
90	3,13	3,27	3,68	3,57	3,42	3,37	3,38	3,32	3,37	3,47	3,33	3,23	3,36	3,39	3,36	3,49	3,71	3,59	3,60	3,47	3,40	3,37
95	1,68	1,97	2,39	2,29	2,13	2,21	2,16	2,09	2,16	2,27	2,17	2,15	2,25	2,31	2,14	2,10	2,16	2,16	2,25	2,13	2,25	2,28

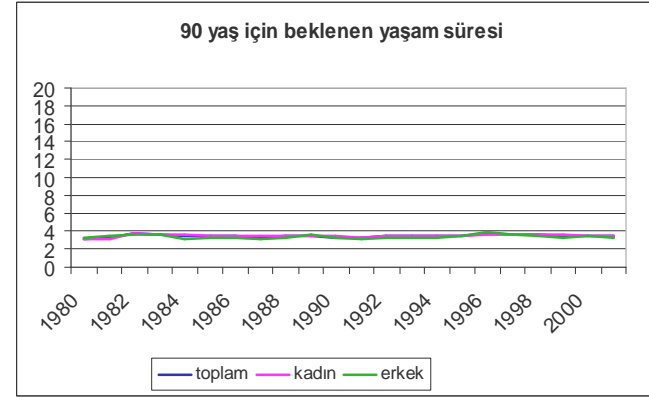
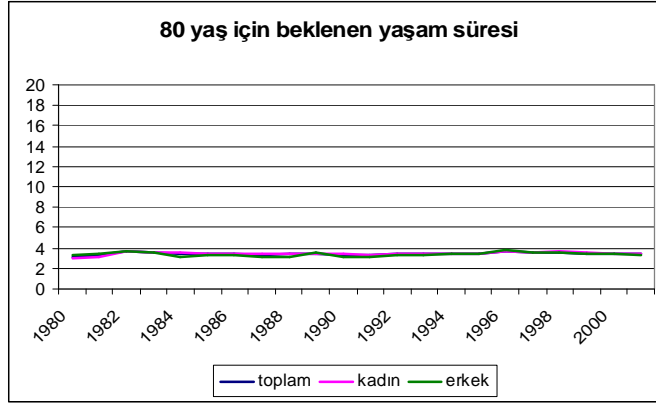
EK-2



EK-2



EK-2



EK-3

TÜRKİYE DE 1980-2001 YILLARI ARASINDA KADINLARIN BEKLENEN YAŞAM SÜRESİ																						
	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
0	46,00	45,57	48,17	48,90	50,34	51,72	53,25	54,11	54,94	55,46	56,25	56,68	57,78	57,75	57,88	58,38	59,18	60,22	60,46	59,53	61,93	62,43
5	61,07	61,16	61,51	62,19	61,59	61,87	62,04	62,03	62,47	62,88	63,03	62,49	62,99	62,88	62,52	62,93	62,81	63,25	63,54	61,61	64,19	64,34
10	57,33	57,56	57,55	58,25	57,60	57,84	57,86	57,97	58,31	58,65	58,79	58,19	58,66	58,50	58,20	58,53	58,39	58,73	59,01	57,37	59,64	59,79
15	52,92	53,31	53,23	53,80	53,25	53,44	53,41	53,54	53,85	54,17	54,27	53,72	54,13	54,02	53,71	53,98	53,83	54,12	54,41	53,01	54,95	55,14
20	48,79	49,04	48,92	49,44	48,93	49,07	49,02	49,22	49,50	49,78	49,86	49,30	49,73	49,59	49,31	49,58	49,64	49,77	49,98	48,88	50,51	50,65
25	44,63	44,73	44,68	45,16	44,65	44,77	44,66	44,88	45,15	45,36	45,42	44,93	45,32	45,19	45,05	45,29	45,20	45,36	45,51	44,73	46,09	46,25
30	40,40	40,47	40,38	40,83	40,35	40,47	40,31	40,56	40,79	40,99	41,01	40,57	40,85	40,75	40,69	40,90	40,79	40,93	41,07	40,54	41,60	41,77
35	36,17	36,22	36,11	36,53	36,08	36,24	36,11	36,31	36,50	36,67	36,66	36,23	36,50	36,39	36,38	36,54	36,43	36,50	36,65	36,23	37,11	37,24
40	31,94	31,97	31,86	32,22	31,89	32,05	31,89	32,04	32,24	32,37	32,37	32,02	32,23	32,11	32,08	32,26	32,12	32,17	32,32	32,09	32,71	32,88
45	27,80	27,79	27,74	28,01	27,75	27,88	27,72	27,82	27,99	28,05	28,16	27,85	28,01	27,88	27,96	27,99	27,95	27,97	28,15	28,00	28,46	28,60
50	23,77	23,75	23,71	23,94	23,65	23,89	23,72	23,81	23,96	24,07	24,07	23,77	23,86	23,76	23,86	23,97	23,87	23,92	24,10	24,04	24,36	24,53
55	20,12	20,03	19,99	20,18	20,03	20,17	20,09	20,10	20,19	20,21	20,26	20,05	20,10	19,88	20,01	20,00	19,92	19,97	20,15	20,21	20,46	20,59
60	16,35	16,42	16,45	16,73	16,62	16,85	16,68	16,70	16,83	16,79	16,78	16,61	16,52	16,24	16,40	16,46	16,39	16,39	16,46	16,47	16,66	16,72
65	12,64	12,71	12,84	13,18	13,20	13,55	13,61	13,70	13,90	13,91	13,94	13,78	13,70	13,60	13,57	13,57	13,45	13,23	13,25	13,18	13,30	13,32
70	9,71	9,64	9,68	9,83	9,85	10,12	10,15	10,47	10,73	10,96	11,12	11,17	11,25	11,30	11,31	11,34	11,00	10,84	10,85	10,55	10,54	10,42
75	7,34	7,47	7,65	7,69	7,62	7,68	7,58	7,63	7,80	7,88	8,05	8,18	8,45	8,69	8,88	9,07	9,06	9,02	9,10	8,78	8,69	8,38
80	5,37	5,40	5,51	5,59	5,55	5,73	5,80	5,88	5,94	5,96	5,87	5,90	5,87	6,00	6,09	6,32	6,44	6,66	6,95	7,02	7,02	6,99
85	4,14	4,22	4,45	4,51	4,35	4,29	4,26	4,22	4,31	4,36	4,49	4,61	4,68	4,66	4,51	4,56	4,58	4,50	4,65	4,70	4,80	4,98
90	3,06	3,20	3,70	3,57	3,53	3,40	3,42	3,39	3,43	3,44	3,39	3,29	3,39	3,47	3,36	3,49	3,65	3,61	3,64	3,52	3,39	3,40
95	1,59	1,97	2,39	2,30	2,09	2,21	2,19	2,15	2,14	2,14	2,15	2,16	2,28	2,37	2,17	2,11	2,15	2,16	2,25	2,15	2,23	2,27

EK-4

TÜRKİYE DE 1980-2001 YILLARI ARASINDA ERKEKLERİN BEKLENEN YAŞAM SÜRESİ																						
	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
0	42,35	42,21	44,43	45,03	46,28	47,83	48,66	49,58	50,28	50,69	51,66	52,06	53,06	53,10	53,35	53,53	54,40	54,79	55,08	55,05	56,60	57,05
5	54,14	54,89	55,43	55,94	55,36	55,89	55,78	55,60	55,83	56,55	56,69	56,15	56,64	56,43	56,36	56,58	56,75	57,00	57,37	56,42	58,19	58,49
10	50,27	51,10	51,44	52,02	51,35	51,83	51,62	51,48	51,67	52,32	52,40	51,78	52,27	52,03	51,94	52,12	52,22	52,48	52,88	52,09	53,63	53,92
15	45,94	46,85	47,08	47,61	47,02	47,47	47,20	47,11	47,21	47,87	47,89	47,33	47,78	47,53	47,45	47,57	47,66	47,90	48,29	47,69	48,96	49,24
20	42,00	42,73	42,90	43,34	42,81	43,22	42,94	42,95	42,99	43,62	43,63	43,08	43,54	43,23	43,18	43,30	43,50	43,62	43,96	43,56	44,61	44,78
25	38,17	38,67	38,84	39,21	38,77	39,07	38,79	38,79	38,83	39,41	39,33	38,83	39,38	39,10	39,02	39,17	39,28	39,36	39,68	39,43	40,30	40,48
30	34,28	34,61	34,72	35,01	34,57	34,88	34,58	34,62	34,64	35,19	35,08	34,63	35,12	34,81	34,79	34,88	34,98	35,06	35,35	35,27	35,93	36,12
35	30,32	30,50	30,59	30,84	30,43	30,67	30,44	30,52	30,57	31,03	30,92	30,47	30,94	30,65	30,67	30,69	30,82	30,81	31,03	31,05	31,53	31,74
40	26,30	26,38	26,40	26,66	26,26	26,61	26,39	26,49	26,57	26,95	26,82	26,45	26,86	26,64	26,63	26,63	26,67	26,71	26,93	26,98	27,34	27,54
45	22,41	22,47	22,52	22,73	22,38	22,66	22,44	22,62	22,65	22,93	22,91	22,66	22,97	22,78	22,79	22,81	22,84	22,92	23,07	23,16	23,40	23,58
50	18,93	18,90	18,92	18,98	18,58	18,99	18,82	18,98	18,96	19,35	19,25	19,03	19,20	19,08	19,14	19,15	19,25	19,33	19,51	19,62	19,78	20,01
55	16,02	15,87	15,89	16,00	15,72	16,00	15,79	15,84	15,63	15,87	15,87	15,73	15,91	15,67	15,81	15,71	15,84	15,94	16,12	16,28	16,43	16,66
60	12,99	13,07	13,20	13,39	13,22	13,44	13,17	13,29	13,14	13,36	13,19	13,08	12,97	12,56	12,62	12,67	12,83	12,85	12,96	13,14	13,28	13,43
65	9,92	10,06	10,22	10,58	10,63	10,95	11,06	11,18	11,23	11,29	11,20	11,02	10,91	10,79	10,73	10,65	10,65	10,38	10,22	10,25	10,48	10,54
70	7,94	7,89	7,90	7,98	7,88	8,10	8,25	8,59	8,83	9,18	9,22	9,26	9,32	9,29	9,31	9,20	8,99	8,72	8,57	8,46	8,45	8,34
75	6,09	6,27	6,46	6,49	6,36	6,45	6,45	6,51	6,58	6,68	6,75	6,89	7,13	7,38	7,56	7,66	7,78	7,54	7,46	7,30	7,16	6,87
80	4,61	4,64	4,82	4,76	4,68	4,90	5,08	5,31	5,22	5,41	5,32	5,30	5,27	5,25	5,25	5,46	5,66	5,69	5,95	5,98	6,03	5,99
85	3,98	4,13	4,28	4,20	3,72	3,77	3,65	3,70	3,70	4,05	4,20	4,23	4,43	4,37	4,38	4,33	4,39	4,06	4,20	3,96	4,15	4,32
90	3,29	3,45	3,63	3,59	3,17	3,31	3,29	3,16	3,22	3,54	3,22	3,10	3,31	3,25	3,36	3,48	3,83	3,57	3,50	3,36	3,42	3,31
95	1,87	1,96	2,39	2,26	2,25	2,22	2,06	1,95	2,22	2,51	2,20	2,13	2,20	2,18	2,07	2,10	2,17	2,14	2,26	2,10	2,30	2,31

KAYNAKÇA

1. DİE (1980), 1980 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
2. DİE (1981), 1981 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
3. DİE (1982), 1982 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
4. DİE (1983), 1983 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
5. DİE (1984), 1984 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
6. DİE (1985), 1985 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
7. DİE (1986), 1986 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
8. DİE (1987), 1987 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
9. DİE (1988), 1988 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
10. DİE (1989), 1989 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
11. DİE (1990), 1990 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
12. DİE (1991), 1991 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
13. DİE (1992), 1992 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
14. DİE (1993), 1993 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
15. DİE (1994), 1994 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
16. DİE (1995), 1995 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
17. DİE (1996), 1996 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler

18. DİE (1997), 1997 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
19. DİE (1998), 1998 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
20. DİE (1999), 1999 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
21. DİE (2000), 2000 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
22. DİE (2001), 2001 Türkiye Ölüm İstatistikleri, Yaş ve Cinsiyete göre Ölümler
23. DPT (2002), 8. Beş Yıllık Kalkınma Planı(2001-2005), 2002 Yılı Programı Destek Çalışmaları, Ekonomik ve Sosyal Sektördeki Gelişmeler, Ankara, s.187
24. Kocaman T., Bulunan Yaşa göre Hayatta Kalma İhtimalleri, Planma Dergisi, Özel Sayı, DPT'nin kuruluşunun 42. yılı
25. Menge W. O. And H. C. Fischer (1985), The Mathematics of Life Insurance, s.31
26. Moralı N. (1997), Aktüeryal Teknikler, s.59-89
27. UNDP (2003), Human Development Index
28. Türkiye İşveren Sendikası Konfederasyonu (2003), İstihdamı Artırmada Geç Emekliliğin Rolü, Mayıs 2003 Aylık Dergi

YAŞAM SÜRELERİNDEKİ FARKLILIKLARIN GINI KATSAYISI İLE İNCELENMESİ

Pınar Özdemir Geyik, A. Kerem Uludağ, Erdem Karabulut,
Osman Saraçbaşı

Hacettepe Üniversitesi Tıp Fak. Biyoistatistik AD
e-mail: pozdemir@hacettepe.edu.tr

Özet

Bu çalışmanın amacı, ekonomi alanında çok uzun süredir kullanılmakta olan Gini Katsayısı'nın sağlık alanındaki kullanımını yaşam tabloları temelinde göstermektir. Günümüzde ortalama yaşam süresi seviyesi bir çok ülkede oldukça yüksek olmakla birlikte yaşam süresi kavramı kapsamındaki değişkenlik ölçüleri giderek daha fazla ilgi çekmektedir. Bu nedenle, yaşam süresi dağılımdaki eşitsizliğe ilişkin çeşitli ölçüler geliştirilmiştir. Bu ölçülerden bazıları Lorenz Eğrisi temeline dayanır. Lorenz eğrisi ekonomi alanında, ilgilenilen popülasyonda gelir dağılımını tanımlamak için kullanılan en yaygın yöntemdir. Bu eşitsizliğin derecesini bir sayı olarak verebilmek için eşitsizlik ölçülerinden yararlanılır. Bu ölçülerden birisi de Gini katsayısıdır. Bu katsayı 0 ile 1 arasında değişir ve 0 tam eşitliği 1 ise tam eşitsizliği gösterir. Gini Katsayısı ekonometride bireyler arası ve haneler arası gelir eşitsizliğini ölçmek üzere standart bir ölçüm olarak kullanılmaktadır. Diğer yandan, yaşam süresindeki eşitsizliği ölçmek için de kullanılabilir bir katsayıdır. Bu katsayı bireylerin yaşam sürelerindeki eşitsizliği bireysel temelde ölçerek, bu eşitsizliğin derecesinin zamana ve ülkelere (bölgelere) göre evrensel olarak karşılaştırılabilmesini sağlar. Bu çalışmada farklı ülkelerin yaşam tabloları kullanılarak Gini katsayıları saptanmış ve ülkeler arası karşılaştırmalar yapılmıştır.

Anahtar Kelimeler: Gini Katsayısı, Lorenz Eğrisi, Yaşam Tablosu, Eşitsizlik Ölçüleri.

Giriş

Günümüzde ortalama yaşam süresi seviyesi birçok ülkede oldukça yüksek olmakla birlikte bazı ülkelerde de düşüktür. Beklenen yaşam süresi düşük olan bir ülkede bireyler arasında beklenen yaşam süresi açısından değişkenlik ya da eşitsizlik, beklenen yaşam süresi yüksek olan ülkedeki

bireylerden daha mı yüksektir? Beklen yaşam süresi eşit olan iki ülkede bireylerin yaşam sürelerindeki eşitsizlik de benzer midir? Bu ve benzeri sorulardan dolayı yaşam süresi kavramı kapsamındaki değişkenlik ölçüleri giderek daha fazla ilgi çekmektedir. Bu nedenle, yaşam süresi dağılımdaki eşitsizliğe ilişkin çeşitli ölçüler geliştirilmiştir. Bu eşitsizlik ölçüleri eşitsizliğin ya da değişkenliğin derecesini farklı yollarla açıklamaya çalışır. Bu ölçülerden bazıları Lorenz eğrisi temeline dayanır. Lorenz eğrisi temeline dayalı olan, en çok bilinen ve sıklıkla kullanılan eşitsizlik ölçülerinden birisi Gini katsayısıdır. Gini katsayısı İtalyan istatistikçi Corrado Gini tarafından geliştirilmiş bir eşitsizlik ölçüsüdür [1].

Lorenz eğrisi, alt-gruplar hiyerarşik olarak artan sıklıkta sıralandığında elde edilen yığılımlı frekansların eğrisidir. Bu ölçü, dağılımda üst ve alt aşırı değerlerin olmasından etkilenmeme gibi bir avantaja sahiptir [2]. Lorenz eğrisi sağlıkla ilgili değişkenler için oluşturulduğunda x eksenini popülasyonun yığılımlı yüzdesini, y eksenini ise ilgilenilen sağlık değişkeninin yığılımlı yüzdesini temsil eder. İlgilenilen olayın toplumda eşit dağılımı köşegenden geçen doğru ile belirtilir. Lorenz eğrisinin eşitlik doğrusundan uzaklaşması ilgilenilen değişkenle ilgili toplumdaki eşitsizliğin arttığını gösterir [3]. Bu eğri ilgilenilen değişkene bağlı olarak köşegen doğrusunun altında ya da üstünde olabilir. İlgilenilen değişken toplumun yararı ile ilgili bir değişkense, eğri köşegen doğrusunun altında olacaktır. Eğer ilgilenilen değişken toplumun zararı ile ilgili bir değişkense, eğri köşegen doğrusunun üstünde olacaktır.

Gini katsayısı sıklıkla kullanılan özetleyici bir eşitsizlik ölçüsüdür ve Lorenz eğrisi ile kolay tanımlanan bir ilişkiye sahiptir. Gini katsayısı Lorenz eğrisi ile köşegen (eşitlik doğrusu) arasında kalan alanın, üçgenin altında kalan alana oranıdır[2]. Bir başka deyişle Gini katsayısı Lorenz eğrisi grafiğindeki alanların oranı olarak hesaplanır. Tam eşitlik doğrusu ile Lorenz eğrisi arasında kalan alan A, ve Lorenz eğrisi altında kalan alan B olursa, Gini katsayısı $A/(A+B)$ olarak hesaplanır. Dolayısıyla, Gini katsayısı 0 ile 1 arasında değer alır. 0 tam eşitliği (herkes aynı yaşam süresine sahip) 1 ise tam eşitsizliği gösterir [1]. Gini katsayısının 0,20'nin altında olması düşük eşitsizliği, 0,20-0,50 arasında olması orta düzeyde eşitsizliği, 0,5'in üzerinde olması ise yüksek eşitsizliği gösterir [2].

Gini katsayısı aşağıdaki Brown Formülü yardımıyla hesaplanabilir:

$$G = \left| 1 - \sum_{k=0}^{k=n-1} (X_{k+1} - X_k)(Y_{k+1} + Y_k) \right|$$

Formülde

- G: Gini katsayısı
X: Populasyon değişkeninin yığılımlı oranı
Y: İlgilenilen sağlık değişkeninin yığılımlı oranı

Yaşam Süresi Dağılımına İlişkin Lorenz Eğrisinin Oluşturulması

Lorenz eğrisinde sıklıkla kullanılan gelir ve populasyon yerine sırasıyla doğumdan ölüme kadar yaşanan kişi yılları ve yığılımlı ölen kişi sayısını kullanarak, bir başka deyişle yaşam tablosu verilerinden yararlanarak Lorenz eğrisi oluşturulabilir [4]. Yaşam tablosunun fonksiyonları temelinde yoğunluk ve dağılım fonksiyonları aşağıdaki gibi tanımlanabilir:

$$f(x)=d(x)/l(0) \quad (1)$$

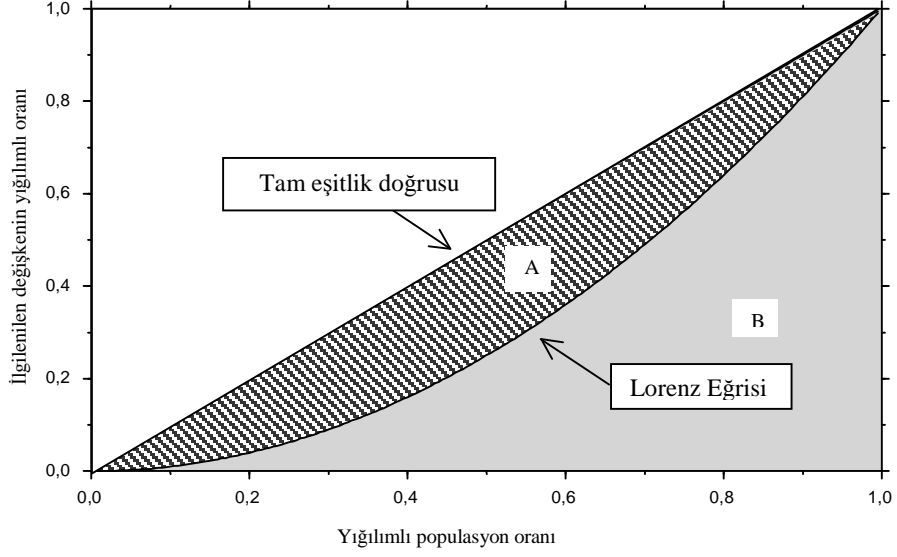
$$F(x)=1-l(x)/l(0) \quad (2)$$

$$\Phi(x) = \frac{1}{e(0)l(0)} \int_0^x td(t)dt \quad (3)$$

Lorenz eğrisi tamamlanmış yaşam tablosu temelinde, ω : yaşam tablosundaki en büyük yaş, $x= 0, \dots, \omega$, ve \bar{t} : t ve $t+1$ yaşları arasında ölen bireylerin yaş ortalaması olduğunda, yatay eksendeki değerler

$$F_x = \frac{\sum_{t=0}^{x-1} d_t}{\sum_{t=0}^{\omega-1} d_t} = 1 - \frac{l_x}{l_0} \quad \text{düşey eksendekiler ise} \quad \Phi_x = \frac{\sum_{t=0}^{x-1} d_t \cdot \bar{t}}{\sum_{t=0}^{\omega-1} d_t \cdot \bar{t}} = \frac{T_0 - (T_x + x l_x)}{T_0}$$

ile bulunur. Eğer tüm bireyler aynı e_0 yaşta ölürlerse, tam eşitlik doğrusunda sadece iki nokta olur. Bunlar $\forall x, x \neq e_0$ için $F(x)=0, \Phi(x)=0$ ve $x=e_0$ için $F(x)=1, \Phi(x)=1$ 'dir.



Şekil 1. Lorenz Eğrisi

$$G_0 = 1 - 2 \int_0^1 \Phi(p) dp, \quad p=F(x) \quad (4)$$

Hanada (1983) Eşitlik 4 ve Eşitlik 1-3'den yararlanarak aşağıdaki eşitliği elde etmiştir.

$$G_0 = 1 - \frac{1}{e(0)[l(0)]^2} \int_0^{\infty} [l(x)]^2 dx \quad (5)$$

Gini katsayısını Eşitlik 4'de verilen geometrik tanımından farklı bir şekilde ifade etmenin çeşitli yolları vardır. Bunların tümü birbirine eşittir (Anand, 1983). Kendall ve Stuart tarafından verilen tanım özellikle bu ölçünün yapısını anlamak için yardımcı olacaktır:

$$G_0 = \frac{1}{2\mu} \int_0^{\infty} \int_0^{\infty} |x - y| f(x) f(y) dx dy \quad (6)$$

Bu eşitlik gini katsayısının basit bir ifadeyle, ortalama yaşam süresine göreli olarak bireysel ölüm yaşlarındaki (yaşam süresindeki) mutlak

farkların ortalaması olması biçiminde tanımlanabilir. Eğer ilgilenilen nüfus l_0 bireyden oluşuyorsa gini katsayısı aşağıdaki eşitlik yardımıyla bulunabilir:

$$G_0 = \frac{1}{2(l_0)^2 e_0} \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j| \quad (6a)$$

Bu eşitlik standart yaşam tablosunun fonksiyonları ile aşağıdaki şekilde de yazılabilir:

$$G_0 = \frac{1}{2(l_0)^2 e_0} \sum_{x=0}^{\omega} \sum_{y=0}^{\omega} d_x d_y |\bar{x} - \bar{y}| \quad (6b)$$

Bu eşitlikteki \bar{x} ve \bar{y} sırasıyla $[x, x+1)$ ve $[y, y+1)$ başlangıç yaş aralıkları için ortalama ölüm yaşlarıdır.

Uygulama

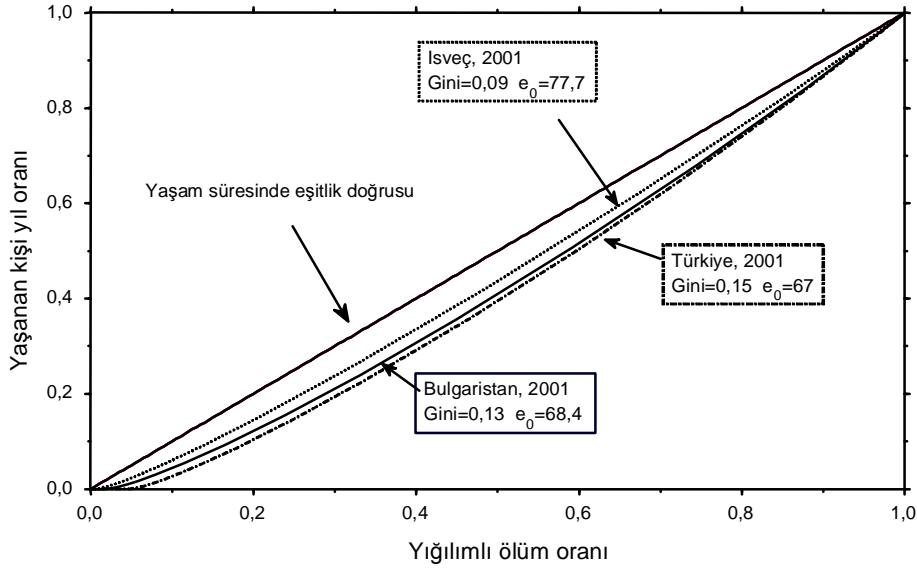
Bu uygulamanın amacı yaşam süresindeki eşitsizliğin gelişmiş ve gelişmekte olan ülkelerde zaman içerisindeki değişimi ve günümüz koşullarında aynı zamanda ülkeler arasındaki yaşam süresi eşitsizliğinin nasıl ortaya konulacağını göstermektir.

Bu çalışmada Türkiye, İsveç ve Bulgaristan'ın yaşam süresindeki eşitsizlik ölçüsünü belirlemek için bu ülkelerin erkeklere ait yaşam tabloları kullanılmıştır. Türkiye dışındaki iki ülkenin İsveç ve Bulgaristan olmasının nedeni: İsveç'in beklenen yaşam süresi en yüksek ülkelerden birisi olması ve gelişmiş bir ülke olması, Bulgaristan'ın ise beklenen yaşam süresinin Türkiye'ye yakın olması ve kullanılan veri tabanında yer alan gelişmişlik düzeyi kötü ülkelerden birisi olmasıdır.

Çalışmada İsveç'in 1900 ve Bulgaristan'ın 1947 yılından günümüze kadarki yaşam tablosu verileri Human Mortality Database (2005) verilerinden ve Türkiye'nin 2001 yılına ait yaşam tablosu Dünya Sağlık Örgütü'nün (2001) verilerinden sağlanmıştır.

WHO'nun 2001 yılına ait üç ülkenin yaşam tablolarından yararlanılarak, yığılımlı ölüm oranı ve yığılımlı kişi yıl oranları hesaplanmıştır. Bu iki orandan yararlanılarak üç ülkenin Lorenz eğrileri çizilmiştir (Şekil 2). 2001 yılında İsveç'in Lorenz eğrisinin eşitlik doğrusuna diğer iki ülkeye göre daha yakın olduğu, Türkiye'nin eğrisinin Bulgaristan'ın altında kaldığı görülmektedir. Bu eğrilerden yararlanarak Türkiye'deki yaşam süresi eşitsizliğinin İsveç ve Bulgaristan'dan fazla olduğu söylenebilir. Bu eşitsizliği sayısal

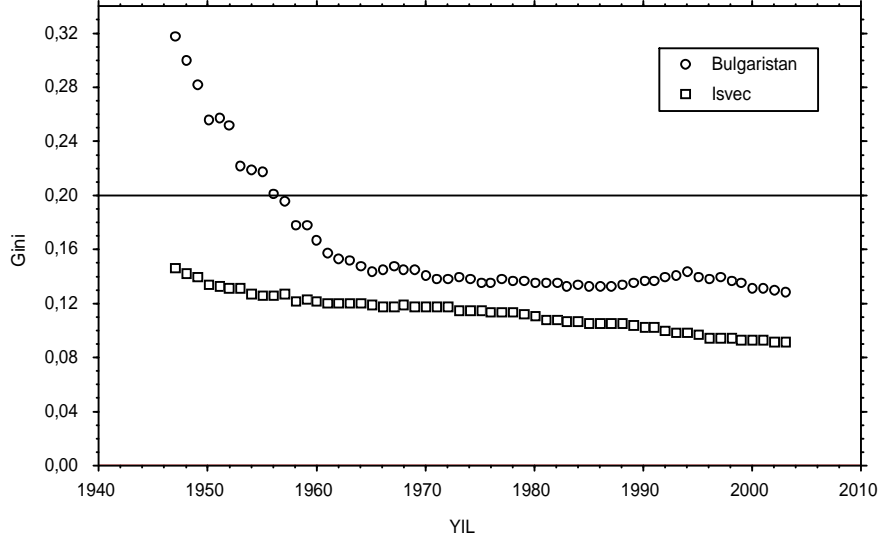
olarak ifade etmek için her ülke için Gini katsayıları hesaplanmıştır. Gini katsayısı İsveç için 0,09, Bulgaristan için 0,13 ve Türkiye için 0,15 olarak elde edilmiştir.



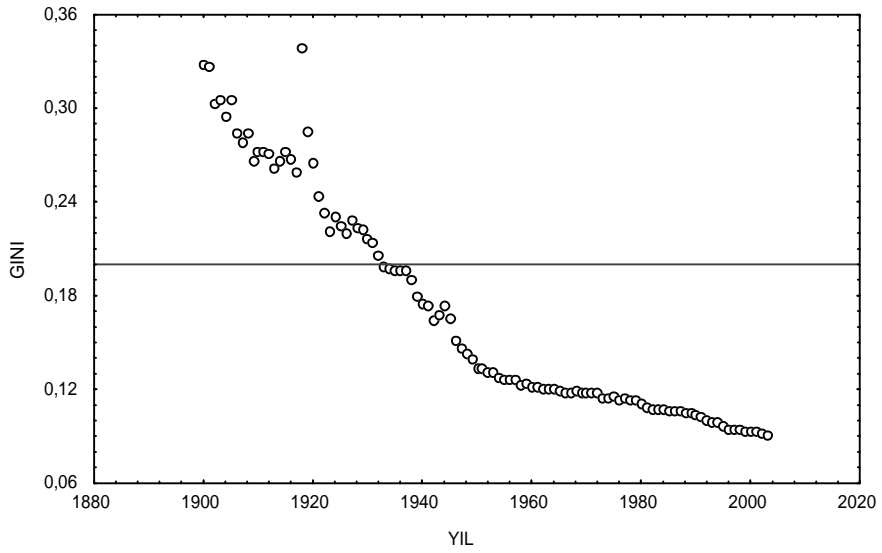
Şekil 2. Üç farklı ülkenin erkek popülasyonunda (2001) yaşam sürelerine ilişkin Lorenz eğrileri

Human Mortality Database veri kaynağından İsveç'in 1900 ve Bulgaristan'ın 1947 yıllarından sonraki yaşam tabloları alınarak, bu ülkelerin 1947 yılı ve sonrasındaki yaşam sürelerindeki eşitsizliğin değişimi her yıl için Gini katsayısı hesaplanarak incelenmiştir (Şekil 3). Bulgaristan'da Gini katsayı ancak 1957 yılı ve sonrasında % 20'lerin altına düşerken, İsveç 1947 ve sonrasında sürekli % 20'den az ve giderek düşen Gini katsayısına sahiptir.

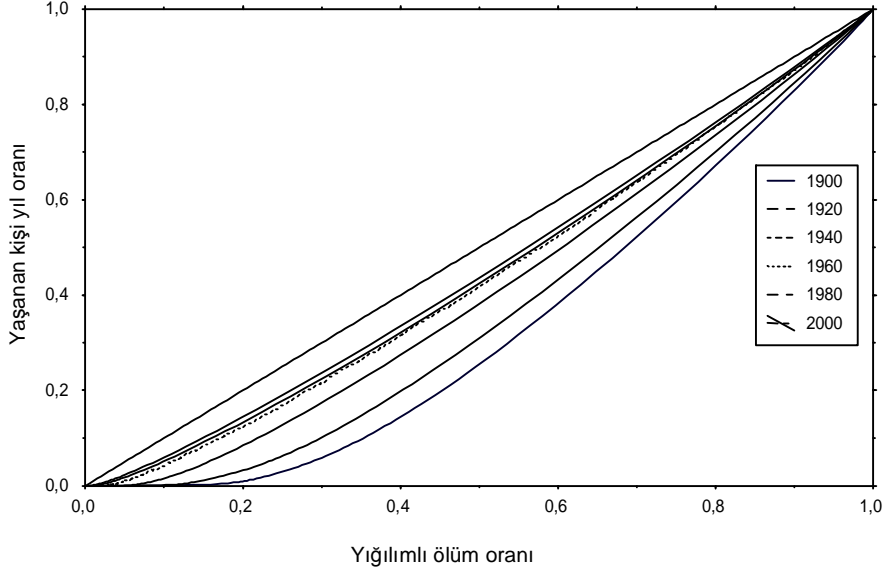
İsveç'in 1947 öncesindeki yaşam süresi dağılımının bireyler arasında ne ölçüde eşit olduğunu görmek için 1900 başlangıç yılı alınarak günümüze kadarki Gini katsayıları hesaplanmış ve 1933 yılı ve sonrasında ancak 0,20'lerin altına düştüğü bulunmuştur (Şekil 4). İsveç'in yaşam süresindeki eşitsizliği gidermedeki başarısı Lorenz eğrileri ile de 20'şer yıllık periyotlarda incelenirse, 1900 yılında eşitlik doğrusuna oldukça uzak olan Lorenz eğrisinin 1940'lardan sonra hızla eşitlik doğrusuna yaklaştığı söyle-nebilir (Şekil 5).



Şekil 3. 1947-2003 yılları arasında İsveç ve Bulgaristan popülasyonunda (erkek) Gini katsayıları



Şekil 4. İsveç erkek popülasyonunda Gini katsayısının zaman içindeki değişimi



Şekil 5. 1900-2000 yılları arasında İsveç erkek popülasyonunda yaşam süresi dağılımının Lorenz eğrileri

Tartışma ve Sonuç

Sağlık hizmetlerinin ve politikalarının temel amacı tüm bireylere sağlıklı ve mümkün olduğunca uzun bir yaşam süresi sağlamaktır. Bu amaca ulaşmanın en önemli ölçütlerinden birisi ise toplumdaki tüm bireylerin sunulan olanaklardan eşit şekilde faydalanabilmesidir. Toplumlardaki sağlık düzeylerinin belirlenmesinde ve bu doğrultuda sağlık politikalarının oluşturulmasında kullanılan temel ölçülerden birisi beklenen yaşam süresidir. Beklenen yaşam süresi ülkelerin sağlık düzeylerinin değerlendirilmesinde her ne kadar önemli ve değerli bir ölçü olsa da yaşam süresinin bireyler temelinde ne kadar homojen (eşit) olduğu hakkında bilgi vermemektedir. Dolayısıyla, bir ülkenin sağlık açısından gelişmişliğinin beklenen yaşam süresinin yanı sıra, bu yaşam süresinin o ülkedeki bireyler temelinde ne kadar dengeli dağıldığı ile değerlendirilmesi, daha ayrıntılı bilgi sağlayacaktır. Kişilere daha uzun ve benzer sürede yaşama fırsatı sağlamanın ne ölçüde başarılı olduğu beklenen yaşam süresi ve yaşam süresindeki eşitsizlik ölçüleri ile ortaya konulabilir. Bu yüzden, gini katsayısı beklenen yaşam süresiyle birlikte, önemli bir sağlık düzeyi göstergesi olarak kullanılabilir.

Kaynaklar

1. Gini Coefficient. http://en.wikipedia.org/wiki/Gini_coefficient
2. Haidich A, Ioannidis J. The Gini coefficient as a measure for understanding accrual inequalities in multicenter clinical studies. Journal of Clinical Epidemiology 2004; 57:341-348
3. Measuring Health Inequalities: Gini Coefficient and Concentration Index. http://www.paho.org/English/SHA/be_v22n1-Gini.htm
4. Shkolnikov VM, Andreev EE, Begun ZA. Gini Coefficient as a life table function: computation from discrete data, decomposition of differences and empirical examples. Demographic Research 2003;8:305-358.
5. Human Mortality Database, University of California, Berkeley, and Max Planck Institute for Demographic Research, www.mortality.org (son ziyaret: Temmuz, 2005)
6. Life Tables for 191 Countries, Mortality Data 2000.
7. <http://www.ho.int/countries/en/>

VEKİL SONLANIM DEĞİŞKENLERİNİN GEÇERLİLİĞİNİN BELİRLENMESİNDE HANGİ KLİNİK DENEMELER KULLANILMALI

Yaşar Sertdemir, Refik Burgut, İlker Ünal

Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana.

Giriş

Kanser, AIDS vb. gibi ölümcül olan hastalıkların tedavileri çok önemlidir. Bu gibi hastalıkların tedavisinde kullanılan ilaç ve tekniklerin geliştirilmesi çok uzun zaman ve para gerektirmektedir. Umut vaat eden tedaviler ise önce klinik denemelerde test edilmelidir. Bu test süreci yaşam süresinin (gerçek sonlanım) incelendiği kanser tedavilerinde 5-10 yıl kadar sürebilmektedir. Ancak yaşam süresi yerine nükse kadar geçen süre, hastalıksız sağ kalım veya tedaviye yanıt (vekil sonlanım) incelendiğinde bu 5-10 yıllık takip süresi kısalabilmektedir.

Ancak vekil sonlanım (VS) değişkenlerinin kullanılabilmesi için önce geçerlilikleri test edilmelidir. Bu amaçla Prentice bazı kriterler önermiş, daha sonra Freedman "Açıklanan Oranı" kullanmayı önermiştir. Bu açıklanan oran ile ilgili bazı sorunların olması nedeniyle Buyse ve ark. "Rölatif Etki"(RE) ile "Düzeltilmiş İlişki"(DI) kriterlerini önermişlerdir. Ancak tek bir araştırmadan elde edilen bilgilerin yeterli olmayacağını kabul ederek Buyse ve ark daha sonra RE'nin çok merkezli karşılığı olan R^2_{trial} ve DI'nin karşılığı olan R^2_{indiv} kriterlerini önermişlerdir.

Bu çalışmada VS değişkeninin geçerliliğinin belirlenmesinde kullanılan klinik denemelerin sonuçlarının R^2_{trial} kriterini nasıl etkilediği gösterilecektir.

Anahtar Kelimeler: Vekil sonlanım, Meta analizi, Kolon kanseri, Regresyon, Geçerlilik.

Materyal ve Metot

Literatür araştırmasından elde edilen kolon kanseri ile ilgili klinik deneme sonuçları tedavinin genel yaşam süresi (gerçek sonlanım) ve hastalıksız sağ kalım (vekil sonlanım) üzerine etkileri ve bunların standart hataları kaydedilerek incelenmiştir.

R^2_{trial} kriteri meta analitik yaklaşımda aşağıdaki formüllerle elde edilen alfa ve beta katsayılarının regresyon analizinde elde edilen R^2 değeridir. (S=VS, T=GS)

$$S_{ij} | Z_{ij} = \mu_{Sj} + \alpha_i Z_{ij} + \varepsilon_{Sij} \quad (1)$$

$$T_{ij} | Z_{ij} = \mu_{Tj} + \beta_i Z_{ij} + \varepsilon_{Tij} \quad (2)$$

Literatür araştırmasından elde edilen kolon kanseri ile ilgili klinik deneme sonuçları tedavinin genel yaşam süresi (gerçek sonlanım) ve hastaliksız sağ kalım (vekil sonlanım) üzerine etkileri ve bunların standart hataları kaydedilerek incelenmiştir.

Eldeki klinik denemeler tedavinin etkinliğine göre gruplandırılmıştır. $|\hat{\beta}/se_{\hat{\beta}}|^2 = z_b$ ve $|\hat{\alpha}/se_{\hat{\alpha}}|^2 = z_a$ olduğu kabul edilirse birinci grup için **z_b ve $z_a \geq 1.65$** üçüncü grup **z_b ve $z_a < 1.65$** ve bunların dışında kalanlar ikinci grubu oluşturmaktadır ($z_{1.65}$ için $p=0.200$).

Bulgular

Bu gruplama ile yapılan analizler sonucu birinci grup için $R^2_{\text{trial}} = 0.873$, ikinci grup için $R^2_{\text{trial}} = 0.565$ ve üçüncü grup için $R^2_{\text{trial}} = 0.750$ bulunmuştur.

Sonuç

Bu sonuçlar bize vekil sonlanım değişkenlerinin geçerliliğinin belirlenmesinde tedavinin her iki sonlanım için birbirine yakın derecede etkin olduğu klinik deneme verilerinin tercih edilmesinin daha doğru olacağını işaret etmektedir.

RASCH ANALİZİ HAKKINDA KISA BİR ÖZET

Atilla Halil Elhan, Yıldır Atakurt

Ankara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı
e-mail: elhan@medicine.ankara.edu.tr

Özet

Bu çalışmanın amacı, sağlık değerlendirim ölçeklerinin değerlendirilmesinde kullanılan modern yaklaşımlardan biri olan Rasch analizinin kısaca incelenmesidir. Rasch analizi herhangi bir kişinin herhangi bir maddedeki bir kategoriyi seçme olasılığının kişi yetenek düzeyi ile madde zorluk düzeyi arasındaki farkın bir lojistik fonksiyonu olduğunu varsayar. Genellikle, ölçeklerin birçoğu sıralı ölçeğe sahiptir. Bu nedenle, maddelere verilen doğru cevapların toplanmasıyla elde edilen ham puanları kullanarak anket ya da ölçek değerlendirilmeye çalışıldığı zaman birtakım sorunlarla karşılaşılır. Rasch analizi bu sorunların üstesinden gelmek için kullanılan yöntemlerden biridir. Sonuç olarak, klasik test teorisinin tersine, Rasch modeli ölçeğin hiyerarşik yapısını, tek boyutluluğunu ve toplanabilirliğini incelemeyi sağlayan alternatif bir ölçekleme yöntemi sağlar.

Anahtar Kelimeler: Rasch analizi, Rasch modelleri, Uyum iyiliği istatistikleri.

1. Giriş

Rehabilitasyon, kişide hastalık sonucu ortaya çıkabilecek özrürlülük ve/veya engelliği azaltmaya, kişinin yaşam kalitesini arttırmaya yönelik problemleri çözme ve eğitim sürecidir. Klinik uygulamalarda rehabilitasyon sürecinde, tedavinin belirlenmesi, izlenmesi ve etkinliğinin değerlendirilmesi amacıyla çeşitli sonuç değerlendirme ölçekleri kullanılmaktadır. Sonuç değerlendirim ve izleniminin en doğru şekilde yapılabilmesi için, bu ölçeklerin kullanıldıkları toplumlara göre dil, dil ve sosyo-kültürel adaptasyonlarının yapılması, geçerlilik ve güvenilirlikleri gösterilerek standartlaştırılmaları gerekmektedir. Bu standartlaştırma işleminde klasik psikometrik analiz yöntemlerinin yeterli olmadığı öne sürülmüş ve yeni istatistiksel analiz yöntemlerinden Rasch analizi gündeme gelmiştir. Bu yöntemler, ulusal ve uluslararası düzeyde güvenle kullanılacak ölçeklerin belirlenmesi ve geliştirilmesine olarak sağlamış, böylece tıbbi rehabilitasyon alanında sonuç değerlendirim ve izleminin standartlaştırılması için çalışmalar başlamıştır.

Bu çalışmanın amacı, ölçeklerin psikometrik özelliklerini değerlendirirken Rasch analizinin yeri hakkında kısaca bilgi vermektir. Bu çerçevede kısaca Rasch analizi, modelleri ve uyum iyiliği istatistikleri hakkında bilgi verilecektir.

Sağlık alanındaki sonuç ölçümlerin birçoğu sıralı ölçeğe sahiptir. Bu durumda sağlık bakım sonuçlarını değerlendirmek için doğrusal aritmetik işlemlerin ve parametrik istatistiksel yöntemlerin kullanılması mümkün olmamaktadır (1, 2). Değişkenin aralık ölçeğe sahip olmadığı durumlarda, sadece hastaların fonksiyonel durumlarında değişiklik olup olmadığı test edilebilmekte, eğer değişiklik varsa bu değişikliğin miktarı belirlenememektedir. Genellikle anketlerden elde edilen toplam puanlar maddelere verilen doğru cevapların ya da başarı sayılarının toplanmasıyla elde edilir (3). Bu tür değişkenler de sıralı ölçeklidir. Değişkenlerin ölçüm düzeyinin aralık mı? yoksa sıralı mı? olduğuna karar vermek önemlidir. Aksi takdirde hangi istatistiksel yöntemin kullanılmasının uygun olacağı konusunda bir şey söylemek mümkün değildir.

Sağlık bakımının etkinliğini belirleme oldukça önemli bir sorundur. Rasch analizi özellikle fiziksel tıp ve rehabilitasyon dalında sıralı ölçekli ölçümleri, aralık ölçekli hale dönüştürmek ve sonuçta sağlık bakımının etkinliğini belirlemek amacıyla oldukça yaygın bir şekilde kullanılmaktadır.

Ölçekler için tek boyutluluk önemlidir. Bunun nedeni, birden fazla boyuta ait maddeler bir araya geldiği zaman, boyutlardan birine ait maddelerin puanındaki artışa karşılık diğer boyuta ait maddelerin puanındaki azalış toplam puandaki değişimin önemini ortadan kaldırmaktadır. Bu nedenle, toplam puanlar araştırmacıya konu hakkında gerçek bir fikir vermeyecektir. Maddelere verilen doğru cevapların toplanmasıyla elde edilen ham puanları kullanarak herhangi bir ölçek değerlendirilmeye çalışıldığı zaman bazı sorunlarla karşılaşılmaktadır (4). Rasch analizi bu sorunlara çözüm getirmektedir.

Genel olarak Rasch analizi, verilen bir görevi bağımsız olarak başarmadaki kişinin yeteneğini (β_n) ve o görevin zorluk derecesini (δ_i) kestirerek, kişinin yeteneği ile maddenin zorluk değerlerini ortak bir eksen boyunca yerleştirmeye çalışır. Böylece herhangi bir yetenek düzeyindeki bir kişinin gerçekleştirilmesi istenen göreve karşı ne yapabileceği olasılığı kestirilir. Rasch analizi sonrası dönüşüm sonuçları lojit olarak sunulur. Lojit; odds oranının doğal logaritması olarak tanımlanır (5).

Kişilerin maddelere verdikleri gözlenen cevaplardan yararlanarak, kişinin yetenek düzeyini ve maddenin zorluk düzeyini kestirmek için β_n , δ_i ve x_{ni} 'yi içeren model geliştirmek gereklidir. Geliştirilen bu modelden elde edilen kişi yetenek düzeyi kestiriminin uygulanan ölçekten bağımsız, madde zorluk düzeyinin de örneklemden bağımsız olarak kestirilmesi gereklidir.

Rasch modelini oluşturmaya kişi yetenek düzeyi (β_n) ile madde zorluk düzeyi (δ_i) arasındaki fark alınarak ($\beta_n - \delta_i$) başlanmaktadır. Rasch analizinde, ($\beta_n - \delta_i$) farkıyla, yetenek düzeyi β_n olan bir kişinin zorluk düzeyi δ_i olan bir anket maddesi karşısında doğru cevap verme olasılığı kestirilmek istenmektedir. Fakat ($\beta_n - \delta_i$) farkı $-\infty$ ile $+\infty$ arasında değişirken, doğru cevap verme olasılığı 0 ile 1 arasında değişmektedir. Bu sorun e sabitine göre ($\beta_n - \delta_i$) farkının üssü alınarak çözümlür.

$$e^{(\beta_n - \delta_i)} = \exp(b_n - d_i)$$

Rasch analizinde 3 farklı cevap formatı için 3 model geliştirilmiştir. Bu modeller Rasch'in ikili modeli (6), sıralı ölçekli model (7, 8, 9, 10) ve kısmi kredi modelidir (11). Her üç modelde de örneklemden bağımsız madde zorluk düzeyi ve ölçekteki maddelerden bağımsız kişi yetenek düzeyi kestirimleri elde edilmektedir.

1.1. İkili Model

Herhangi bir soru maddesinde sadece iki kategori (katılıyorum/katılmıyorum, evet/hayır) mevcut ise bu soru maddesi en basit cevap şekline sahiptir. Bu şekilde puanlanan maddeler "tek-basamaklı" maddeler olarak adlandırılır. Eğer bu tek-basamak başarı ile gerçekleştirilirse, kişi bu maddeden 1 puan alır. Eğer bu tek-basamak başarı ile gerçekleştirilmezse, o zaman 0 puan alır. Denklem 1'den anlaşılacağı gibi i 'ninci maddede başarılı olma olasılığı yetenek düzeyi ile madde zorluk derecesine bağlıdır.

Georg Rasch tarafından geliştirilen ikili model aşağıda verilmiştir.

$$\text{İkili (Dikotom) Model: } \phi_{ni1} = \frac{\exp(\beta_n - \delta_{i1})}{1 + \exp(\beta_n - \delta_{i1})} \quad (1)$$

f_{ni1} ifadesi; n 'ninci kişinin i 'ninci soru maddesinden 0 yerine 1 puan alma olasılığı, b_n ; n 'ninci kişinin yetenek düzeyi, d_{i1} ise i 'ninci soru maddesinde tek basamak atlama zorluk derecesi olarak açıklanabilir.

Diğer bir ifadeyle, n 'ninci kişinin i 'ninci maddeden 0 puan alma olasılığı π_{ni0} ile, 1 puan alma olasılığı ise π_{ni1} ile gösterilsin. İlk ve tek basamağı başarıma olasılığı ϕ_{ni1} aşağıdaki gibi gösterilir.

$$\phi_{ni1} = \frac{\pi_{ni1}}{\pi_{ni0} + \pi_{ni1}} = \frac{\exp(\beta_n - \delta_{11})}{1 + \exp(\beta_n - \delta_{11})} \quad (2)$$

Maddelere verilecek cevaplar için sadece iki kategorinin mevcut olduğu durumlar için $\pi_{ni0} + \pi_{ni1} = 1$ ve $\phi_{ni1} = \pi_{ni1}$ eşitlikleri geçerlidir.

1.2. Kısmi Kredi Modeli

Kısmi kredi modelinde maddelere verilen cevapların sonuçlarını sadece doğru/yanlış olarak değil, doğruluk kavramını göreceli olarak geliştirip bir ya da daha fazla orta düzeye yayarak orta derecede bilgiye sahip kişilerin de kredilendirilmesi mümkün olmaktadır. Örnek olarak, Tablo 1’de kısmi kredi biçimine uygun iki madde verilmiştir. Her bir madde için seçenek olarak sıralı 4 kategori verilmiştir. Bu kategoriler 0, 1, 2 ve 3 olarak etiketlenmiştir ve 3 etiketi en yüksek performansı ifade etmektedir.

Dört farklı performans düzeyi aynı maddede tanımlandığında o madde “3 basamaklı” bir madde olarak düşünülür. Bu maddedeki ilk basamak 0 yerine 1 puan almak için sorunun “ $4.0/0.2 = ?$ ” bölümünü çözmektir. İlk basamağın başarıyla tamamlanmasına bağlı olan ikinci basamak 1 yerine 2 puan almak için “ $20-4 = ?$ ” çıkarma işlemini yapmaktır. Son olarak, üçüncü basamak tabii ki ilk iki basamağın başarıyla geçilmesi koşuluyla 2 puan yerine 3 puan almak için “ $\sqrt{16} = ?$ ” işlemini çözmektir.

Tablo 1. Kısmi kredi modeline örnek iki soru maddesi

Matematik maddesi	
$\sqrt{4.0/0.2} - 4 = ?$	
Hiç ilerleme kaydetmemiş.....	0
$4.0/0.2 = 20$	1
$20 - 4 = 16$	2
$\sqrt{16} = 4$	3

Kısmi kredi modeli için genel denklem aşağıdaki gibidir.

$$\phi_{nik} = \frac{\pi_{nik}}{\pi_{nik-1} + \pi_{nik}} = \frac{\exp(\beta_n - \delta_{ik})}{1 + \exp(\beta_n - \delta_{ik})} \quad k= 1, 2, \dots, m_i \quad (3)$$

Denklem 3'ten yararlanarak n 'ninci kişinin madde i den x puanı alma olasılığı aşağıdaki gibi verilmiştir.

$$\text{Kısmi Kredi Modeli: } \pi_{nix} = \frac{\exp \sum_{j=0}^x (\beta_n - \delta_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^k (\beta_n - \delta_{ij})}, \quad x = 0, 1, \dots, m_i \quad (4)$$

$$d_{i0} \equiv 0, \quad \sum_{j=0}^0 (\beta_n - \delta_{ij}) = 0 \quad \text{ve} \quad \exp \sum_{j=0}^0 (\beta_n - \delta_{ij}) = 1$$

Denklem 4'teki x ifadesi başarıyla tamamlanmış olan basamak sayısıdır.

1.3. Sıralama Ölçekli Model

Herhangi bir davranış anketindeki bir madde için " k 'ninci basamağı tamamlamak" maddedeki k 'ninci kategoriye ($k-1$)'inci kategoriye tercih etmek olarak düşünülür. Böylece sıralı kategorileri aşağıdaki gibi olan bir davranış anketinde 3 farklı basamak vardır.

<i>Kesinlikle katılmıyorum</i>	<i>Katılmıyorum</i>	<i>Katılıyorum</i>	<i>Kesinlikle katılıyorum</i>
0	1	2	3

Sıralama ölçekli maddelerde basamakların zorluk derecelerinin maddeden maddeye göre değişmediği varsayılır. Bu beklenti kısmi kredi modelinde madde basamaklarını aşağıdaki gibi iki bölümden oluşacak şekilde yeniden yazarak gerçekleştirilir.

$$\delta_{ik} = d_i + t_k$$

d_i ifadesi i 'ninci maddenin zorluk düzeyidir ve t_k ifadesi ise her maddedeki k 'ninci basamağın o maddenin ölçek değerine göre konumudur. Madde basamakları "eşik" parametreleri t_1, t_2, \dots, t_m tarafından

tanımlanmaktadır ve tüm madde kümesi için geçerli olacağından sadece bir defa kestirilmektedir.

$d_{ik} = d_i + t_k$ eşitliğiyle kısmi kredi model, sıralı ölçekli modele dönüşür (7, 8, 9).

$$\phi_{nik} = \frac{\pi_{nik}}{\pi_{nik-1} + \pi_{nik}} = \frac{\exp[\beta_n - (\delta_i + \tau_k)]}{1 + \exp[\beta_n - (\delta_i + \tau_k)]} \quad k= 1, 2, \dots, m$$

Kısmi kredi modelinde olduğu gibi n 'inci kişinin madde i den kategori x 'i seçme olasılığı aşağıdaki gibidir.

$$\text{Sıralama Ölçekli Model: } \pi_{nix} = \frac{\exp \sum_{j=0}^x [\beta_n - (\delta_i + \tau_j)]}{\sum_{k=0}^m \exp \sum_{j=0}^k [\beta_n - (\delta_i + \tau_j)]}, \quad x= 0, 1, \dots, m \quad (5)$$

$$t_0 \equiv 0, \quad \exp \sum_{j=0}^0 [\beta_n - (\delta_i + \tau_j)] = 1$$

1.4. Rasch Analizinde Kullanılan Kestirim Yöntemleri

Rasch analizinde parametreleri kestirmek amacıyla birkaç yöntem kullanılmaktadır. Bunlardan ikisi PROX ve Koşulsuz Kestirim (UCON) yöntemleridir.

PROX kestirim yöntemi hesap makinesi yardımıyla hesaplanabilecek kadar kolay bir yöntemdir. PROX yönteminde madde zorluk ve hasta yetenek düzeyleri ortak bir doğrusal eksen üzerinde ifade edilir ve örneklerin yerleşiminden, örneklem ve uygulanan testin yaygınlığından bağımsız sonuç verir.

İkinci kestirim yöntemi olan UCON, koşulsuz en çok olabilirlik kestirim yöntemidir ve iteratif işlemlerle hasta parametrelerini madde parametreleriyle birlikte eş zamanlı olarak kestirir. PROX yöntemine göre tercih edilmelidir.

1.5. Rasch Modeline Uyumun İncelenmesi

Yapısal iç geçerliliğin değerlendirilmesi verilerin Rasch modeline uyumunun incelenmesiyle yapılır. Rasch ölçüm modeli ölçüm aracından elde edilen verilerin tek boyutlu olduğunu varsayar ve böylece ölçekte bulunan

maddelerin tek bir yapıya ait olup olmadığını test etmek amacıyla Rasch ölçüm modeli kullanılır. Verilerin Rasch modeline uyumunu test etmek aslında teorik olarak yapısal geçerliliği ve ölçeğin yeterliliğini test etme işlemiyle aynıdır. Madde ve kişi uyumu iki farklı kareler ortalaması uyum istatistiği ile değerlendirilmektedir. Bunlar ağırlıklandırılmamış kareler ortalaması (OUTFIT), ve ağırlıklı kareler ortalamasıdır (INFIT). Bu istatistikler her madde için ayrı ayrı hesaplanır ve birlikte ele alındıkları zaman maddelere verilen cevapların tutarlılıkları hakkında bilgi sağlar. Uzak gözlemlere duyarlı kareler ortalaması uyum istatistiği (OUTFIT) kişinin yetenek düzeyinden uzakta olan anormal cevaplara daha duyarlıdır. Örnek olarak, çok yetenekli bir kişinin çok kolay bir maddeye cevap vermesi gibi. OUTFIT istatistiğindeki artık karelerinin her biri kendi varyansı ile ağırlıklandırıldığında INFIT istatistiği elde edilir. Böylece uzak gözlemlerin etkisi ağırlıklandırma nedeniyle azalmış olur. INFIT istatistiği daha çok merkezi, başka bir ifadeyle, orta zorluk düzeyindeki cevaplar hakkında bilgi sağlar. Yani, kişinin yetenek düzeyi ile benzer zorluk düzeyindeki maddelere vermiş olduğu cevaplara daha duyarlıdır. Kareler ortalaması değerleri 0.6–1.4 arasında olduğu zaman, eldeki örneklem büyüklüğü için modele yeterli uyumun sağlandığı kabul edilir (12, 13). Genel olarak, herhangi bir maddenin kareler ortalamasının 1.4'ün üzerinde (>1.4) olması, ya o maddeye verilen cevapların beklenenin dışında olmasından, ya maddenin iyi anlaşılmasından, ya da o maddenin diğer maddelerin oluşturduğu yapıya ait olmamasından kaynaklanmış olduğu kabul edilir. Herhangi bir maddenin kareler ortalamasının 0.6'nın altında (<0.6) olması ise beklenenden daha az değişimin olduğunu, yani maddenin doğru olamayacak kadar uyumlu olduğunu göstermektedir (12). Sonuçta, bir ölçeğin tek boyutluluğu, maddelerin modele uyum iyiliğiyle test edilir.

Winsteps ve RUMM 2020 paket programları yukarıda verilen 3 model için hem madde zorluk düzeylerini ve kişi yetenek düzeylerini kestirmekte, hem de verilerin Rasch modeline uyumunu test etmektedir. Bu nedenle Rasch analizinin uygulama zorluğu bulunmamaktadır.

2. Sonuç

Klasik test teorisinin tersine, Rasch modeli ölçeğin hiyerarşik yapısını, tek boyutluluğunu ve toplanabilirliğini incelemeyi sağlayan alternatif bir ölçekleme yöntemi sağlar. Sağlık, eğitim ve psikoloji alanlarında çok sık kullanılan ölçeklerin gerek geliştirilmesinde, gerekse Türkiye'ye adaptasyonlarının yapılmasında klasik yöntemlere göre birçok avantajı olan Rasch analizinin kullanılması, bu alanda çalışanlara gerçekten yeni bir bakış açısı getirecektir.

3. Kaynaklar

1. Chang, W.C., Chan, C. (1995). Rasch analysis for outcomes measures: Some methodological considerations. Arch Phys Med Rehabil, 76:334-339.
2. McHorney, C.A., Haley, S.M., Ware, J.E., Jr. (1997). Evaluation of the MOS SF-36 physical functioning scale (PF-10): II. Comparison of relative precision using likert and rasch scoring methods. J Clin Epidemiol, 50 (4): 451-461.
3. Linacre, J.M., Heinemann, A.W., Wright, B.D., Granger, C.V., Hamilton, B.B. (1994). The structure and stability of the Functional Independence Measure. Arch. Phys Med. Rehabil, 75:127-132.
4. Elhan, A.H., Atakurt, Y. (2005) Ölçeklerin değerlendirilmesinde niçin Rasch analizi kullanılmalıdır? Ankara Üniversitesi Tıp Fakültesi Mecmuası, 58:47-50.
5. Haley ve ark., 1994 Haley, S.M., McHorney, C.A., Ware, J.E., JR. (1994). Evaluation of the MOS SF-36 physical functioning scale (PF-10): I. Unidimensionality and reproducibility of the rasch item scale. J Clin Epidemiol, 47 (6): 671-684.
6. Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests (1960). Copenhagen: Danmarks Paedagogiske Institut (Uni. of Chicago Press, Chicago, 1980).
7. Andrich, D. (1978a). Scaling attitude items constructed and scored in the Likert tradition. Educ Psychol Meas, 38: 665-680.
8. Andrich, D. (1978b). Application of a psychometric rating model to ordered categories which are scored with successive integers. Appl Psychol Meas, 2: 581-594.
9. Andrich, D. (1979). A model for contingency tables having an ordered response classification. Biometrics, 35: 403-415.
10. Masters, G.N. (1980). Rasch model for rating scales. Doctoral dissertation, University of Chicago.
11. Masters, G.N., Wright B.D. (1981). A model for partial credit scoring. Rasch Memorandum No. 31, MESA Psychometric Laboratory, Department of Education, University of Chicago.
12. Wright, B.D., Linacre, J.M. (1994). Reasonable mean-square fit values. Transactions of the Rasch Measurement SIG American Educational Research Association, Rasch Meas Trans, 8(3): 370.

13. Smith, R.M., Schumacker, R.E., Bush, M.J. (1998). Using item mean squares to evaluate fit of the Rasch model. *Journal of Outcome Measurement*, 2:66-78.

**MENAPOZ SEMPTOMLARINI DEĞERLENDİRME ÖLÇEĞİ
(MRS), KUPPERMAN MENAPOZ ÖLÇEĞİ (KMÖ) VE
NOTHINGAM SAĞLIK PROFİLİ (NSP) ÖLÇEKLERİNİN
GÜVENİLİRLİK VE YANITLAMA ETKİ
BÜYÜKLÜKLERİNİN YENİ MENAPOZLU HASTALARDA
KARŞILAŞTIRILMASI**

N. Bekiroğlu¹, Ayşe Gürbüz², R. Konyalıoğlu³

Marmara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı¹

Zeynep Kamil Kadın ve Çocuk Hastalıkları Hastanesi²

ARK İstatistiksel Danışmanlık³

e-mail: nural@marmara.edu.tr

Özet

Çalışmamız, 2005 Nisan-Haziran ayları arasında, Zeynep Kamil Kadın ve Çocuk Hastalıkları Hastanesinde gerçekleştirilmiştir. Bu çalışmada, Menapoz Semptomlarını Değerlendirme Ölçeği (MRS), Kupperman Menapoz Ölçeği (KMÖ) ve Nothingam Sağlık Profili (NSP) ölçeklerinin etki büyüklüğü bakımından karşılaştırılması amaçlanmıştır. Çalışmaya 60 yeni menopoz tanısı alan hasta katılmıştır. Her üç anketin güvenilirlik katsayıları ve yanıtlama etki büyüklükleri hesaplanmıştır.

Test-Tekrar test güvenilirlik çözümlemesinde, sıralanmış ve sürekli veriler için sınıf-içi korelasyon katsayısına bakılmıştır. İç tutarlılık katsayısına ise Cronbach α ile bakılmıştır. Yanıtlamadaki etki büyüklüğü için değişim miktarı hesaplanmıştır. Değişime olan duyarlılığın hesaplanması standardize ortalama yanıt (SOY) ile sağlanmıştır. SOY ise, skorlardaki ortalama değişimin skorlardaki değişimin standart sapmasına bölümüyle bulunmuştur. Her 3 ölçeğinde önce ve sonraki ölçümleri arasında oldukça anlamlı fark bulunmuştur ($p < 0.001$).

Sonuç olarak, son yıllarda önemli bir araştırma konusu olan hayat kalitesi ölçeklerini kullanırken, hastanın sağlık durumunu ve gerçek klinik faydayı daha iyi yansıtması bakımından, genel ölçekler yerine hastalıklara özel geliştirilmiş ölçeklerin kullanılmasının daha faydalı olduğunu düşünmekteyiz.

Anahtar Kelimeler: Standardize ortalama yanıt, Yanıtlama etki büyüklüğü, Hayat kalite ölçekleri.

Kaynaklar

1. Heinemann L.A.J., DoMinh T., Strelow F., Gerbsch S., Schnitker J., Schneider H pg. (2004) The Menopause Rating Scale (MRS) as outcome measure for hormone treatment? A validation study Health and Quality of Life Outcomes 2004, 2:67
2. Kazis, L.E., Anderson, J.J., Meenan, R.F., (1989) Effect size for interpreting changes in healthy status. Med. Care, 24:7; 1370-1377.
3. Schneider HP., Heinemann LA., Rosemeier HP., Potthoff P., Behre HM. (2000) The Menopause Rating Scale (MRS): comparison with Kupperman index and quality-of-life scale SF-36. Climacteric. 2000 Mar; 3(1): 50-8.
4. Heinemann L.A.J., Potthoff P., Scheider H PG. (2003) International versions of the Menopause Rating Scale (MRS) Health and Quality of Life Outcomes 2003, 1:28
5. Suvanto-Luukkonen E., Koivunen R., Sundstrom H., Bloigu R., Karjalainen E., Haiva-Mallinen L., Tapanainen JS.(2005). Citalopram and fluoxetine in the treatment of postmenopausal symptoms: a prospective, randomized, 9-month, placebo-controlled, double-blind study. Menopause. January/February 2005 12(1):18-26.

RASGELE VE SİSTEMATİK HATALARIN SINIF İÇİ VE UYUM KORELASYON KATSAYILARI VE BLAND&ALTMAN YÖNTEMİ ÜZERİNE ETKİLERİ: BİR SİMÜLASYON ÇALIŞMASI

Semra Erdoğan, E. Arzu Kanık

*Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı
e-mail: arzukanik@mersin.edu.tr*

Özet

Ölçme yöntemlerinin karşılaştırılması analizlerinde ve özellikle ölçüm yöntemleri arasındaki uyumu incelemek için sınıfıçi (intraclass, ICC) korelasyon katsayısı, uyum (concordance, CCC) korelasyon katsayısı ve bunlara alternatif olarak grafiksel bir yöntem olan Bland & Altman yöntemleri kullanılmaktadır. Ölçüm güvenilirliği ve geçerliliğini etkileyen en önemli hata kaynakları rasgele ve sistematik hatalardır. Rasgele hatalar ölçümlerde güvenilirliği azaltırken sistematik hatalar ölçümün geçerliliğini etkilerler. Bu çalışmada iki farklı ölçme yöntemini veya iki farklı değerlendiricinin uyumunu test eden bu üç yöntemin rasgele ve sistematik hataların varlığından nasıl etkilendikleri bir simülasyon çalışması ile deneysel olarak araştırılmıştır.

Bu çalışmada, ortalamaları sıfır, varyansları 1 ve farklı korelasyon yapısına sahip (0.3, 0.6 ve 0.9) $N=1000$ olan popülasyonlar üretilmiştir. Bu popülasyonlar için ICC, CCC ve Bland & Altman uyum sınırları ve grafikleri elde edilmiştir. Bu verilere $N(0,1)$ olarak üretilen ve ölçümlerden bağımsız olan rasgele hatalar ile ölçümlerle korelasyonlu olan sistematik hatalar eklenmiş ve ICC, CCC ve Bland & Altman için hesaplamalar yapılmıştır.

Bu çalışmanın sonuçlarına göre; Bland & Altman yöntemine göre güven sınırlarının dışında gözlemlerin oranı hiçbir şekilde ne uyum iyiliğinden ne de rasgele ve sistematik hatadan etkilenmemektedir. Ancak Bland&Altman grafikleri sistematik hatayı çok iyi ortaya koyduğundan bu çalışmanın sonucuna göre grup içi veya uyum korelasyonlarıyla birlikte sistematik bir hatanın olup olmadığını gözlemlmek için Bland & Altman grafiklerinden yararlanılması araştırmacılara önerilmektedir.

Anahtar Kelimeler: Sınıfıçi korelasyon katsayısı, uyum, Bland& Altman yöntemi, metot karşılaştırması, sistematik, rasgele hata.

Giriş

Ölçme yöntemlerinin karşılaştırılması analizlerinde ve özellikle ölçüm yöntemleri arasındaki uyumu karşılaştırmak için intraclass korelasyon katsayısı, concordance korelasyon katsayısı ve bunlara alternatif olarak grafiksel bir yöntem olan Bland & Altman yöntemleri kullanılmaktadır. Bu yöntemler sadece ölçme yöntemlerinin karşılaştırılmasında değil aynı zamanda tek bir ölçme yönteminin tekrarlanması durumunda veya iki farklı değerlendirici tarafından elde edilen ölçümlerin uyumlarını test etmek için de kullanılmaktadırlar. Özellikle biyokimya ve radyoloji alanlarında sıklıkla bu yöntemlere ihtiyaç duyulmaktadır. Ölçüm güvenilirliği ve geçerliliğini etkileyen en önemli hata kaynakları rasgele ve sistematik hatalardır. Rasgele hatalar ölçümlerde duyarlılığı azaltırken sistematik hatalar ölçümün geçerliliğini etkilerler. Bu çalışmada iki farklı ölçme yöntemini veya iki farklı değerlendiricinin uyumunu test eden bu üç yöntemin özelliklerine değinilmiştir. Ayrıca bu yöntemlerin rasgele ve sistematik hataların varlığından nasıl etkilendikleri bir simülasyon çalışması ile deneysel olarak araştırılmıştır.

Gereç ve Yöntem

Intraclass korelasyon katsayısı (ICC)

Sınıf içi korelasyon katsayısı olarak adlandırılır ve iki veya daha fazla ölçme yönteminin karşılaştırılmasında kullanılan popüler yöntemlerden birisidir. Sınıf içi (intraclass) sözcüğü, aynı değişkenin tekrarlı ölçümlerini analiz etmek gibi bir kavram içerirken, sınıflar arası (interclass) sözcüğü iki farklı değişkenin analizi için uygun bir kavramdır (korelasyon katsayısı gibi). Korelasyon katsayısı, daha çok farklı ölçüm birimiyle ölçülmüş iki veri setinin ilişki miktarının ölçümünde kullanılır. Bundan dolayı, sınıflar arası analize uygundur ve veri iki farklı ölçüm birimiyle ölçüldüğü için ortalamalardaki değişikliklerle ilgilenmez.

Sürekli değişkenler için ölçüm uyumu analizlerinde araştırmacılar tarafından en çok yapılan hatalardan birisi bu uyumu test etmek için eşleştirilmiş (paired) t testinin veya Pearson korelasyon katsayısının kullanılmasıdır. Genellikle ölçüm uyumu yüksek olduğunda bu iki test ile de uyumun iyi olduğuna ilişkin bir bilgi edinilebilmektedir. Ancak bu iki testin ölçüm uyumu analizlerinde kullanılması hatalıdır.

Sınıf içi güvenilirlik katsayısı küçük örneklerde ($n < 15$) veya ikiden daha çok test yapıldığında bu testler arasında bir ilişki olduğunda tercih edilir. Sınıf içi korelasyon katsayısı kompleks tasarımlar için geliştirilmiştir. Normal dağılım gösteren iki değişken gibi basit durumlarda bile, sınıf içi

korelasyon katsayısı Pearson korelasyon katsayısından daha avantajlıdır (1, 2).

Sınıf içi korelasyon katsayısının kabul edilebilir seviyeleri Çizelge 1'deki gibi verilebilir (3).

Çizelge 1. Sınıf İçi Korelasyon Katsayısının Kabul Edilebilir Seviyeleri

Sınıf içi korelasyon değeri	Kabul edilebilir seviye
< 0.40	Zayıf
0.40 – 0.59	Orta
0.60 – 0.74	İyi
> 0.74	Çok İyi

Bu korelasyon katsayısının kabul edilebilir seviyeleri verilerin özelliğine ve kullanıldığı alana göre farklılık göstermektedir. Örneğin, fizyolojik veriler üzerinde bir çalışma yapılıyor ise bu durumda sınıf içi korelasyon katsayısının 0.90'nın üzerinde olması kabul edilebilir seviyesinin "yüksek", 0.80 ile 0.89 arasında olması "orta", ve 0.80'nin altında olması "şüpheli" olarak nitelendirilir. Birden fazla değerlendirici arasında bir uyum olduğunu söyleyebilmek için sınıf içi korelasyon katsayısının en az 0.70 olması gerekmektedir. Sınıf içi korelasyon katsayısı 0.95 ile 1.00 arasında bir değer alıyorsa değerlendiriciler arasındaki uyumun "mükemmel" olduğunu, 0.85 ile 0.94 arasında bir değer alıyorsa değerlendiriciler arasındaki uyumun "yüksek" olduğunu, korelasyon değerinin 0.70'in altında olması durumunda değerlendiricilerin birbirleriyle hiçbir şekilde uyumlu olmadığını söyleyebiliriz (4).

ICC (sınıf içi korelasyon katsayısı);

$$\rho = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_w^2}$$

ile tanımlanır (5).

Concordance Korelasyon Katsayısı (CCC)

Concordance korelasyon katsayısı Lin tarafından 1989 yılında tanımlanmıştır. Concordance korelasyon katsayısı (CCC)

$$\rho_c = \frac{2\sigma_{12}}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2}$$

şeklinde elde edilir (6,7) .

CCC (ρ_c)'nın Özellikleri

1. -1 ile +1 arasında bir değer alır ve Pearson korelasyon katsayısının mutlak değerinden daha büyük bir değer alamaz.
2. Pearson korelasyon katsayısının sadece sifıra eşit olduğu bir durumda CCC'da sifıra eşit olacaktır.
3. Ortalamalar ve varyanslar birbirine eşit ise CCC Pearson korelasyon katsayısına eşit olur. Yani; $\sigma_1 = \sigma_2$ ve $\mu_1 = \mu_2$ olduğu durumda $\rho_c = \rho$ olur.
4. Aşağıdaki koşullardan en az biri sağlandığında
 - a. $(\mu_1 - \mu_2)^2 + (\sigma_1 - \sigma_2)^2 + 2(1 \pm \rho)\sigma_1\sigma_2 = E[(Y_1 - Y_2)^2] = 0$ veya
 - b. $\sigma_1 = \sigma_2$ ve $\mu_1 = \mu_2$ veya
 - c. Değerlendiricilerin her birinde mükemmel bir ilişki (uyum) var ise (örneğin, 1,1; 2,2; 3,3; 4,4; 5,5) veya mükemmel bir ters ilişki (uyum) var ise (örneğin, 5,1; 4,2; 3,3; 2,4; 1,5) $\rho_c = \pm 1$ olur (6,7,8).

Bland&Altman Yöntemi (Uyum Sınırları, Limits of Agreement)

Bu yöntem iki ölçme yönteminin karşılaştırılmasında intraclass ve concordance korelasyon katsayılarına alternatif olarak Bland & Altman tarafından (1986) geliştirilmiş olan bir grafik yöntemidir. Bu yöntem ilk olarak yöntemler arasındaki farklılığa karşı ortalamaların saçılım grafiği çizilerek başlanır. Ortalamaya karşı iki yöntem arasındaki farklılığın grafiği aynı zamanda ölçüm hataları ile gerçek değerler arasındaki herhangi bir olası ilişkiyi de araştırmamıza müsaade eder (9). Ölçümlerin ortalamasına karşılık bu farklılıkların grafiği ölçüm hatasındaki sistematik hatadaki olası eğilimleri de ortaya çıkarır. Bu eğilim ortalama ve farklar arasındaki spearman rank korelasyonu kullanılarak test edilebilir. Benzer durumlarda, bu eğilimi çoğu orijinal değerlerin yerine ölçümlerin logaritması alınarak ortadan kaldırılabilir. Bu özel dönüşümler kullanıldığında ölçümlerin farklılıkların yerine oranlar için uyum sınırları elde edebiliriz. Logaritma dönüşümü aynı zamanda normallik varsayımını da başarmaktadır (10).

İki ölçme yöntemi arasındaki farka ait ortalama (\bar{d}) ve standart sapma (S) hesaplanarak ölçme yöntemleri arasındaki uyum sınırları incelenebilir. Eğer farklar normal dağılım gösteriyorsa farkların sıfırın etrafında rasgele dağılması ve % 95'inin $\bar{d} - 2s$ ile $\bar{d} + 2s$ arasında yer alması beklenir. Bu sınırlar içerisinde sağlanan farklılıklar klinik olarak önemli olmayabilir ve böyle bir durumda bu iki ölçme yöntemini birbirlerinin yerine kullanılabilir. Bu yöntemde $\bar{d} \pm 2s$ Bland&Altman'ın uyum sınırları olarak tanımlanmaktadır. Farklılıklar normal dağılım gösteriyorken tahmin edilen uyum sınırlarının doğruluğunu test etmek için güven aralıklarına başvurulmaktadır. Uyum sınırlarına ait güven aralıkları,

$$(\bar{d} \pm 2S) \pm t * S_d$$

şeklinde hesaplanır. Yani, uyumun alt sınırı $\bar{d} - 2s$ için güven aralığı; $(\bar{d} - 2S) \pm t * S_d$ şeklinde hesaplanırken, üst sınırı $\bar{d} + 2s$ için güven aralığı $(\bar{d} + 2S) \pm t * S_d$ şeklinde hesaplanmaktadır. Burada S_d farklara ait standart hatayı vermektedir ve uyum sınırlarına ait standart hata,

$$S_d = \sqrt{\frac{3S^2}{n}}$$

şeklinde hesaplanır. Güven aralıkları formülünde küçük örnekler için (n-1) serbestlik dereceli t tablo değeri, büyük örnekler için de tablo değeri olarak 1.96 verilmektedir (9,11).

Eğer farklar ve ortalamalar arasında ilişki varsa böyle bir durumda kullanılacak ilk yöntem logaritma dönüşümü uygulamaktır. Eğer bu yöntemle ilişki ortadan kaldırılabiliyse uyum sınırları anti-logaritma kullanılarak tekrar hesaplanır: Aksi durumda, fark değerleri ancak saçılım grafikleri ile gösterilebilir ve uyum sınırları istatistiksel olarak bir değer değil de tanımlayıcı bir değer gibi verilebilir (12).

Rasgele Hata

Değişkenin ölçümlerini rasgele olarak etkileyen ve geçici etmenlere bağlı olan pek çok faktörden etkilenmektedir. Rasgele hatanın kaynağı ve ölçmedeki miktarları kesin olarak bilinmemektedir. Aynı zamanda rasgele

hatalar ölçmelere hem pozitif hem de negatif olmak üzere iki farklı şekilde karışabilmektedirler. Bu özelliğinden dolayı dağılımdaki bütün rasgele hata toplamının sıfıra eşit olması gerekmektedir. Daha az rasgele hata içeren ölçümler daha güvenilir ölçümlerdir (13). Rasgele hatanın artmasıyla varyasyon katsayısı da artar.

Sistemik Hata

Sistemik hata, ölçümleri ya pozitif ya da negatif olarak etkiler ve genellikle ölçümün değeri arttıkça artar yani sistemik hata genellikle ölçümlerle pozitif korelasyonludur. Hem ölçümlerle korelasyonlu olması hem de bu sebeple ortalamasının sıfırdan farklı olmasından dolayı Bu iki özelliği bakımından rasgele hatadan farklılık gösterir. Bu özelliğinden dolayı sistemik hata bazen ölçümlerde “**bias**” olarak adlandırılır (13).

Simülasyon Çalışması

Bu çalışma ile iki ölçme yöntemi arasındaki uyum analizinde kullanılan sınıf içi korelasyon katsayısı (ICC), concordance korelasyon katsayısı (CCC) ve Bland&Altman yöntemlerinin rasgele hata, sistemik hata ve hatasızlık durumunda nasıl değiştiklerini incelemek amacıyla bir simülasyon çalışması yapılmıştır. Bu veriler, bilgisayar ortamında MNRDGV v.2 (14) programı ile üretildikten sonra ICC'ı SPSS 11.5 paket programı (15) yardımı ile CCC Excel'de yapılmış olan Linconcordance (16) adlı bir program ile ve Bland&Altman grafikleri ise MEDCALC (17) paket programı ile elde edilmiştir.

x_1 ve x_2 ölçümlerinin üretilmesi:

Ortalamaları ve varyansları eşit ve pearson korelasyon katsayıları aşağıda verildiği gibi olan normal popülasyonlar simülasyonla üretilmiştir.

$$N = 1000$$

$$\mu_1 = \mu_2 = 0$$

$$\sigma^2 = 1$$

$$\rho_{x_1x_2} = 0.3, 0.6, 0.9$$

Yukarıda verilen varyans ve pearson korelasyon katsayıları için ayrı ayrı veriler üretilip her biri için ICC, CCC ve Bland&Altman uyum sınırları ve

grafikleri elde edilmiştir. 3 farklı korelasyon için ICC, CCC ve Bland&Altman hesaplanmıştır.

Rasgele hata eklenmiş x_1 ve x_2 ölçümlerinin üretilmesi:

Rasgele hatanın bu üç yöntem üzerine etkisini araştırmak için yukarıda verilen varyans ve korelasyon kombinasyonlarından faydalanılmıştır. Rasgele üretilen birbirinden bağımsız iki farklı $N(0,1)$ e_1 ve e_2 değişkenleri x_1 ve x_2 ölçümlerinin üzerine eklenerek rasgele hata oluşturulmuştur.

Sistemik hata oluşturulmuş x_1 ve x_2 ölçümlerinin üretilmesi:

Rasgele üretilen ve x_1 ölçümüyle sırasıyla 0.3, 0.6 ve 0.9 seviyesinde korelasyonlu $N(0,1)$ verileri x_1 üzerine eklenerek sistemik hata oluşturulmuştur. Ölçümlerin birinde oluşturulan sistemik hatanın üç yöntem üzerine etkisi araştırılmıştır.

Bulgular

Ortalamaları 0 ve varyansları 1, pearson korelasyon katsayıları 0.3, 0.6, 0.9 olduğu durumlarda ICC, CCC ve Bland&Altman uyum sınırları ve bu sınırlara ait güven aralıkları çizelge 2'de, rasgele bir hata oluşturduktan sonra 0.3, 0.6 ve 0.9 pearson korelasyonuna sahip olduğu durumlarda ICC, CCC ve Bland&Altman uyum sınırları ve bu sınırlara ait güven aralıkları çizelge 3'de ve son olarak da sistemik bir hata oluşturduktan sonra 0.3, 0.6 ve 0.9 pearson korelasyonuna sahip olduğu durumlarda ICC, CCC ve Bland&Altman uyum sınırları ve bu sınırlara ait güven aralıkları çizelge 4'de verilmiştir. Bu üç farklı durum için Bland&Altman saçılım grafikleri ise şekil 1, şekil 2 ve şekil 3'de yer almaktadır.

Çizelge 2. Hata oluşturulmamış durumda ICC, CCC ve Bland&Altman'a ait uyum sınırları(Limits of Agreement, LOA) ve Güven aralıkları

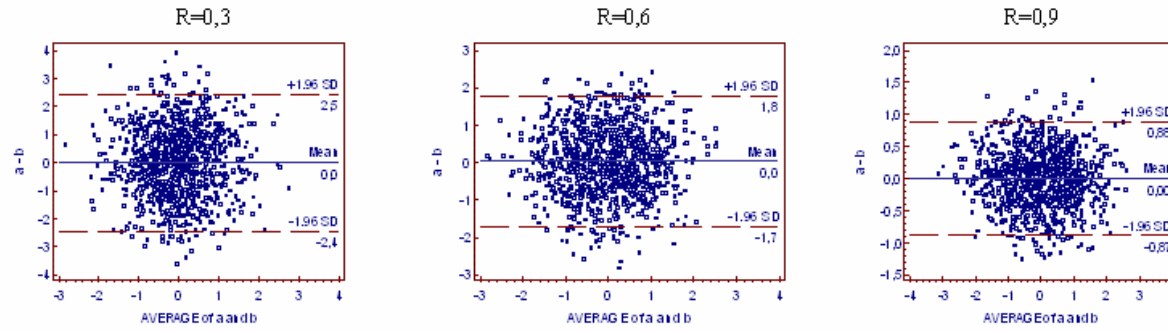
r	ICC	CCC	\bar{d}	S	Limits of agreement			%CI_LOA'ın Alt Sınırı için		%CI_LOA'ın Üst sınırı	
					$\bar{d} + 2s$	$\bar{d} + 2s$	S_d	$(\bar{d}-2S)-1.96*S_d$	$(\bar{d}-2S)-1.96*S_d$	$(\bar{d}-2S)-1.96*S_d$	$(\bar{d}-2S)-1.96*S_d$
0,3	0,241	0,241	0,006	1,2478	-2,49	2,501	0,068	-2,624	-2,356	2,367	2,635
0,6	0,6	0,6	0,03	0,89	-1,75	1,81	0,049	-1,846	-1,654	1,714	1,906
0,9	0,903	0,903	0,005	0,45	-0,895	0,905	0,025	-0,943	-0,847	0,857	0,953

Çizelge 3. Rasgele Hata oluşturulduğunda ICC, CCC ve Bland&Altman'a ait uyum sınırları (Limits of Agreement, LOA) ve Güven aralıkları

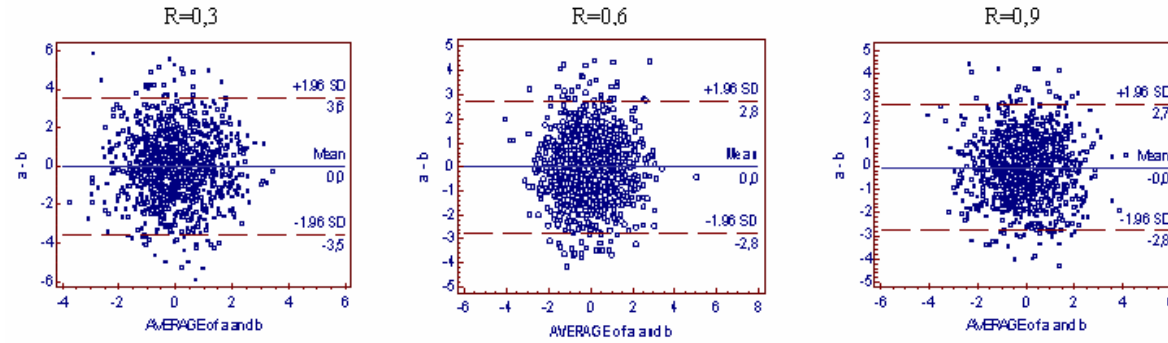
r	r	ICC	CCC	\bar{d}	S	Limits of agreement			%CI_LOA'ın Alt Sınırı için		%CI_LOA'ın Üst sınırı	
						$\bar{d} + 2s$	$\bar{d} + 2s$	S_d	$(\bar{d}-2S)-1.96*S_d$	$(\bar{d}-2S)-1.96*S_d$	$(\bar{d}-2S)-1.96*S_d$	$(\bar{d}-2S)-1.96*S_d$
0,3	0,204	0,204	0,204	0,012	1,817	-3,623	3,647	0,100	-3,818	-3,428	3,452	3,842
0,6	0,488	0,489	0,489	0,001	1,405	-2,808	2,811	0,077	-2,959	-2,657	2,660	2,962
0,9	0,51	0,51	0,51	-0,045	1,394	-2,833	2,744	0,076	-2,982	-2,683	2,594	2,893

Çizelge 4. Sistematik Hata oluşturulduğunda ICC, CCC ve Bland &Altman'a ait uyum sınırları (Limits of Agreement, LOA) ve Güven aralıkları

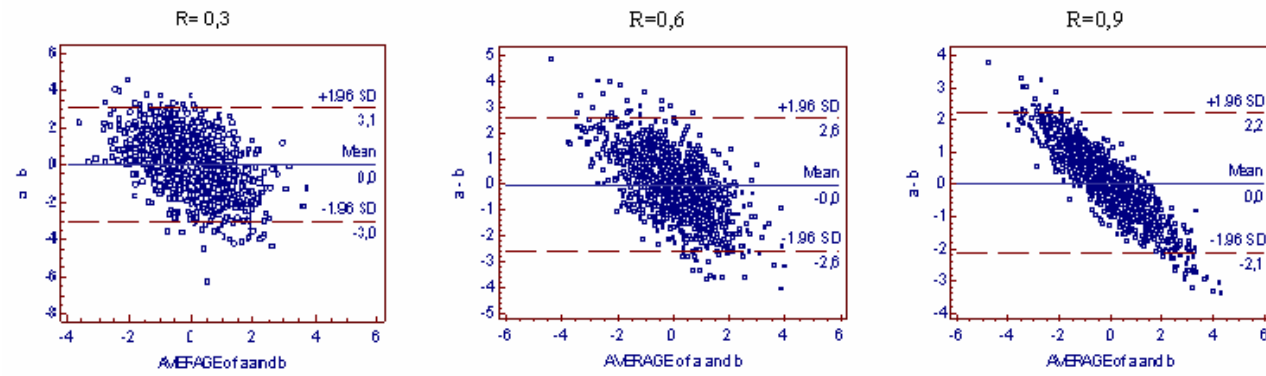
ρ	r	ICC	CCC	\bar{d}	S	Limits of agreement			%CI_LOA'ın Alt Sınırı için		%CI_LOA'ın Üst sınır	
						$\bar{d} - 2s$	$\bar{d} + 2s$	S_d	$(\bar{d} - 2S) - 1.96 * S_d$	$(\bar{d} - 2S) - 1.96 * S_d$	$(\bar{d} - 2S) - 1.96 * S_d$	$(\bar{d} - 2S) - 1.96 * S_d$
0,3	0,378	0,332	0,332	0,037	1,554	-3,07	3,144	0,085	-3,237	-2,903	2,978	3,311
0,6	0,676	0,586	0,586	-0,017	1,319	-2,655	2,621	0,072	-2,797	-2,513	2,480	2,763
0,9	0,928	0,748	0,748	0,046	1,112	-2,179	2,271	0,061	-2,299	-2,060	2,151	2,390



Şekil 1. Hata oluşturulmadığı durumda Bland & Altman'a ait saçılım grafikleri



Şekil 2. Rasgele hata oluşturulduğunda Bland & Altman'a ait saçılım grafikleri



Şekil 3. Sistemik hata oluşturulduğunda Bland & Altman'a ait saçılım grafikleri

Tartışma

Hata oluşturulmayan veri setinde (Çizelge 2) gerçek uyum popülasyon parametrelerine olduğu gibi yansımakta ICC ve CCC değerleri bu parametreye eşit olarak hesaplanmaktadır. Bu durum, iki ölçme yönteminin de gerçek ölçümü gösterdiği yani güvenilir ve geçerli oldukları anlamına gelmektedir. Bland&Altman güven sınırları ise uyum artıkça daralmaktadır. Grafikler ise daire şeklini korumaktadır (Şekil 1). Bland&Altman yöntemine göre $\rho=0.3$ olduğunda verilerin % 4'ü, $\rho=0.6$ olduğunda % 5.2'si, $\rho=0.9$ olduğunda ise % 4.9'u güven sınırları dışında kalmaktadır. Ortalamaları eşit iki yöntem üzerinde durulduğundan bu değerler yaklaşık % 5 ile ifade edilebilir.

Rasgele hata oluşturulan veri seti analiz edildiğinde (Çizelge 3) ise gerçek uyum popülasyonda ölçüm sırasında rasgele hata tarafından gölgelenmektedir ve parametre gerçek değerinden daha düşük olarak ölçülmektedir. ICC ve CCC değerleri ise bu korelasyonu aynen yansıdığı gibi tahmin edebilmektedir. Yani ölçüm yöntemi geçerliliğini korumakta ancak güvenilirliği azalmaktadır. Bland&Altman yönteminde şekiller daireselliğini korumakta ve hata oluşturulmayan gruba göre güven aralığının genişlemesine rağmen hesaplamalara göre aralık dışında kalan gözlemlerin oranının tüm korelasyonlarda yaklaşık %5'i koruduğu görülmektedir (Şekil 2). Ancak rasgele hata ne grafiklerde ne de istatistiklerin değerlerinden açıkça fark edilememektedir.

Sistemik hata oluşturulan veri setinde (Çizelge 4) ölçüm tekniği sonucunda gerçek uyum korelasyonlara yansımaya rağmen ICC ve CCC değerleri sistemik hata yüzünden bu korelasyonları olduğundan düşük hesaplanmaktadır. Bu durumda sistemik hata parametrenin yanlış tahmin edilmesine neden olmaktadır. Bland & Altman yönteminde güven sınırları hata oluşturulmayan gruba göre daha geniş olmasına rağmen, sınırların dışında kalan gözlemlerin oranı yine % 5'i geçmemektedir. Burada önemli sorun bu istatistiklerden ölçümde sistemik bir hata olduğunun anlaşılmasıdır. Ancak ortalamalar arasında fark olmamasına rağmen ICC ve CCC değerlerinin veri setindeki korelasyon katsayısından küçük çıkması, sistemik hatanın varlığı için bir ipucu olabilir. Bunun yanında Bland & Altman grafikleri çok çarpıcı bir şekilde sistemik hatayı ortaya koyabilmektedir. Bu korelasyon $\rho=0.3$ olduğunda bile grafiklerde belirgin bir şekilde fark edilmektedir. Sistemik hatanın grafiklerden görülebileceği Hollis tarafından da ifade edilmektedir (18). Sistemik hatanın ölçümlerle olan korelasyonun artması kötü etkisinin daha çok ortaya çıkmasına sebep olmaktadır. Çizelge 4'de $\rho=0.9$ olduğunda ICC ve CCC değerlerinin veri setindeki korelasyondan daha çok uzaklaştığı görülmektedir.

Bland &Altman yöntemine göre güven sınırlarının dışında gözlemlerin oranı hiçbir şekilde ne uyum iyiliğinden ne de rasgele ve sistematik hatadan etkilenmemekte ve bu sınırlar araştırmacıya yöntem uyumu veya hatalar hakkında herhangi bir bilgi vermemektedir. Yöntem sayısı ikiden çok olduğunda kullanılamamasına rağmen Bland &Altman grafikleri sistematik hatayı çok iyi ortaya koyduğundan bu çalışmanın sonucuna göre grup içi veya uyum korelasyonlarıyla birlikte sistematik bir hatanın olup olmadığını gözlemlemek için Bland &Altman grafiklerinden yararlanılması araştırmacılara önerilmektedir.

Kaynaklar

1. Lewis, R.J. Reliability and Validity: Meaning and measurement. Erişim: [http:// www. Ambpeds.org/reliabilityandvalidity.pdf](http://www.Ambpeds.org/reliabilityandvalidity.pdf). Erişim tarihi: 17.05.1999
2. PA765: Reliability. Erişim: <http://www2.chass.ncsu.edu/garson/pa765/reliab.htm>. Erişim Tarihi: 25.08.2003
3. Barrett P. Assessing the reliability of Rating Data. Erişim: <http://www.pbarett.net/rater.pdf>. Erişim Tarihi: 2001
4. Alpar, R. Spor Bilimlerinde Uygulamalı İstatistik. Ankara: Nobel Yayın Dağıtım.2001
5. Fleiss J.L, Shrout PE. Intraclass Correlation: uses in Assessing Rater Reliability. Psychological Bulletin. 1979; 86(2): 420-428.
6. Carosso LJ, Jover L. Estimating the Generalized Concordance Correlation Coefficient through Variance Component. Biometrics, 2003; 59: 849-858.
7. Lin LI-Kuei. Assay Validation Using the Concordance Correlation Coefficient. Biometrics, 1992; 48: 599-604.
8. Erdoğan S, Kanık EA. Sürekli Değişkenler için Güvenilirlik Analizinde Kullanılan İlişki Katsayılarının Değişken Varyansından ve Etki Büyüklüğünden Etkilenme Durumları. 7.Ulusal Biyoistatistik Kongresi Özet Kitapçığı, sayfa:43, 29 Eylül-2 Ekim Mersin, 2004.
9. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; i: 307-310.
10. Rousson V, Gasser T., Seifert B. Assessing intrarater, interrater and test-retest reliability of continuous measurements. Statistics in Medicine, 2002; 21; 3431-3446.

11. Bland J.M., Altman DG. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol* 2003; 22: 85-93.
12. Bland JM, Altman DG. Measurement in medicine: The analysis of comparison studies. *Statistician* 1983;32: 307-317.
13. Trochim, WMK. Measurement Error. Erişim: [http://www.tech.farmingdale.edu/Nibrahiaz/MET 409/True Score Theory. Mht.](http://www.tech.farmingdale.edu/Nibrahiaz/MET_409/True_Score_Theory.Mht) Erişim Tarihi: 01.08.2003.
14. Kanık EA, Mengi T. Simülasyon Yardımıyla Çok Değişkenli Veri Üretimi: MNRDG.v.2, 7.Ulusal Biyoistatistik Kongresi Özet Kitapçığı, sayfa:43, 29 Eylül-2 Ekim Mersin, 2004.
15. SPSS Inc. SPSS for windows. Version 11.5
16. Lin's Concordance vs Pearson Corelation. Erişim: <http://www.ucsf.edu/cando/resources/linconcordance.xls>. Erişim tarihi: 10.08.2003.
17. MEDCALC. Medcalc for Windows. Version 8.0.1.
18. Hollis S. Analysis of method comparison studies *JIFCC* 1997; 9.

HEPATİT A VİRÜS ENFEKSİYONUNUN İLERİYE YÖNELİK TAHMİN EDİLMESİNDE DÖRT FARKLI ZAMAN SERİSİ YÖNTEMİNİN KARŞILAŞTIRILMASI

Mevlüt Türe¹, İmran Kurt²

*Trakya Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı¹
Osmangazi Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı²
e-mail: [ture@trakya.edu.tr](mailto:tire@trakya.edu.tr)*

Özet

Hepatit A virüs (HAV) enfeksiyonu, sadece gelişmekte olan ülkelerin değil gelişmiş ülkelerin de problemi. Bu çalışmada, HAV olgularını tahmin etmek için üç Yapay Sinir Ağları (YSA) algoritmaları (çok tabakalı algılayıcı (Multi Layer Perceptron, MLP), radyal tabanlı fonksiyon (Radial Basis Function, RBF) ve zaman gecikmeli sinir ağları (Time Delay Neural Networks, TDNN)) ve otoregresif tamamlanmış hareketli ortalama süreci (Auto Regressive Integrated Moving Average, ARIMA) yöntemlerinin zaman serilerinde tahmin performanslarının karşılaştırılması amaçlandı. Bu yöntemlerin etkinliklerini değerlendirmek için Türkiye’de aylara göre 13 yıllık (Ocak 1992 – Haziran 2004) HAV olguları kullanıldı. Uygulama sonucunda MLP’nin RBF, TDNN ve ARIMA yöntemlerinden daha az hata ile tahminleme yaptığı bulundu.

Anahtar Kelimeler: Sinir Ağları, MLP, RBF, TDNN, ARIMA, Hepatit A, İleriye yönelik tahminleme.

1. Giriş

Viral hepatit pek çok insanı etkileyen ve dünyadaki en önemli sağlık problemlerinden birisidir [1]. Akut enfeksiyon ve kronik bozukluklar nedeniyle yüksek mortalite ve morbiditeye neden olmaktadır [2]. Hepatit A virüsü (HAV) enfeksiyonu genellikle kendi kendini sınırlayan, tam şifa ile sonuçlanan ve hayat boyu bağışıklık bırakan bir enfeksiyon hastalığıdır. HAV enfeksiyonu, hem gelişmekte olan ülkelerin hem de gelişmiş ülkelerin sağlık sorunudur. Hijyenik olmayan bölgelerde çocuk hastalığı olarak da görülmektedir [3-5]. Fakat yetişkinlerde HAV enfeksiyonu önemli mortalite nedenidir [6]. Aşı

maliyetlerinin ve stoklarının belirlenmesi gibi nedenlerle ülkeler için HAV enfeksiyonunun tahmini oldukça önemlidir. Bu nedenle zaman serisi yöntemleri, toplumdaki hastalıkların zaman içindeki hareketliliklerinin belirlenmesinde ve gelecekteki hareketliliğin tahmini için zaman kalıplarının ortaya çıkarılmasında kullanılan yöntemlerdir [7].

Zaman serisi olaylarını tahminlemede kullanılan otoregresif tamamlanmış hareketli ortalama süreci (Auto Regressive Integrated Moving Average, ARIMA) yöntemi, Box ve Jenkins [8] tarafından geliştirilmiştir. ARIMA modelleme yaklaşımı, değişkenler arasında doğrusallık olması varsayımıyla kısıtlıdır. Bununla birlikte araştırmacılar, doğrusallık varsayımını sağlamayan zaman serisi olaylarının tahminlenmesi için alternatif modelleme yaklaşımları geliştirmişlerdir.

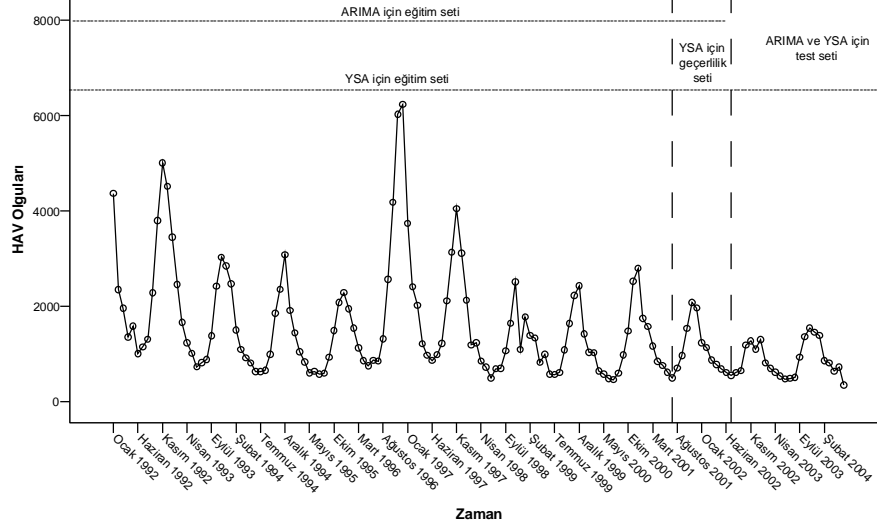
Yapay Sinir Ağları (YSA), insan beynindeki sinir ağları gibi çalışarak bilgisayar yardımıyla en zor problemlerin çözümünü sağlayan ve değişken yapısı konusunda herhangi bir kısıtlama getirmeksizin değişkenler arasındaki ilişkiyi ortaya koyan çok esnek bir yöntemdir [9]. Bir çok araştırmacı, doğrusal olmayan zaman serisi olaylarını tahmin etmede YSA'yı kullanmışlardır [10-13]. YSA modellerinin performanslarının klasik tahminleme yöntemleriyle karşılaştırılması pek çok çalışmaya konu olmuştur [10,14-16].

Bu çalışmanın amacı: (i) aylara göre toplanmış 13 yıllık (Ocak 1992 – Haziran 2004) HAV olgularını kullanarak ARIMA, zaman gecikmeli sinir ağları (Time Delay Neural Networks, TDNN), çok tabakalı algılayıcı (Multi Layer Perceptron, MLP) ve radyal tabanlı fonksiyon (Radial Basis Function, RBF) yöntemleri ile modelleme yapılması, (ii) son iki yıllık (Temmuz 2002 – Haziran 2004) test verisiyle modellerden elde edilen tahmin değerleri karşılaştırılarak en uygun modelin belirlenmesidir.

2. Gereç ve Yöntemler

2.1. Veri Seti

Çalışmamızda kullanılan HAV olgu sayıları, Sağlık Bakanlığı kayıtlarından Ocak 1992 – Haziran 2004 tarihleri arasında aylık olarak elde edildi. Analiz öncesinde YSA için veri seti; eğitim, geçerlilik ve test seti olmak üzere üç alt sete, ARIMA için eğitim ve test seti olmak üzere iki alt sete ayrıldı. YSA için Ocak 1992 – Haziran 2001 ayları arası eğitim seti, Temmuz 2001 – Haziran 2002 ayları arası geçerlilik seti, Temmuz 2002 – Haziran 2004 ayları arası test seti; ARIMA için Ocak 1992 – Haziran 2002 ayları arası eğitim seti, Temmuz 2002 – Haziran 2004 ayları arası test seti olarak seçildi (Şekil 1).



Şekil 1. Aylara göre HAV olguları ve YSA ile ARIMA yöntemleri için ayrılan eğitim, geçerlilik ve test setlerinin gösterimi

2.2. ARIMA

Box-Jenkins yöntemi olarak da bilinen ARIMA yöntemi, bir değişkenin mevcut değerlerinden yararlanarak gelecekteki değerlerinin tahminlenmesinde kullanılan bir yöntemdir. ARIMA modelleme yaklaşımı üç aşamadan oluşur: (1) otokorelasyon ve kısmi otokorelasyon yöntemlerini kullanarak p , d ve q parametrelerinin belirlenmesi; (2) modele anlamlı katkıların olup olmadığını ya da biri ya da diğerinin çıkarılıp çıkarılmadığını görmek için p (otoregresif) ve q (hareketli ortalama) bileşenlerinin tahmini ve (3) iyi bir modelin belirtisi olan artıkların rasgele ve normal dağılılıp dağılmadıklarının belirlenmesi [8,17].

Eğer ARIMA modeli mevsimsel olmayan (p,d,q) ve mevsimsel olan (P,D,Q) bileşenlerin her ikisini de içeriyorsa, model $ARIMA(p,d,q) \times (P,D,Q)$ olarak tanımlanır. Burada p ve P sırasıyla mevsimsel olmayan ve mevsimsel otoregresif operatör; d ve D sırasıyla mevsimsel olmayan ve mevsimsel fark alma operatörü; q ve Q sırasıyla mevsimsel olmayan ve mevsimsel hareketli ortalama operatörüdür. Aylık HAV verisindeki yapıyı tanımlamak için, otokorelasyon fonksiyonu (ACF) ve kısmi otokorelasyon fonksiyonu (PACF) grafikleri kullanıldı. Bu grafiklere dayanarak, veri seti için olası bütün ARIMA modelleri belirlendi. HAV verisi için olası modeller arasından en iyi modelin seçilmesinde Akaike Bilgi Kriteri (AIC) [18] kullanıldı.

2.3. Zaman Gecikmeli Sinir Ağları (TDNN)

TDNN, çok tabakalı ileri beslemeli bir sinir ağıdır. Waibel tarafından 1988 yılında ses tanıma için geliştirilmiş olmasıyla birlikte aynı zamanda zaman serisi tahmini için de kullanılan bir yöntemdir. TDNN, çok tabakalı ileri beslemeli bir sinir ağının girdi tabakada bellek hücreleri yerleştirilerek oluşturulan dinamik bir sinir ağıdır [19]. Girdi ve gizli tabakalar arasındaki bellek hücrelerinde, sonlu dürtü cevaplı (finite impulse response, FIR) filtreleme kullanılır. TDNN girdi verisini, bu filtreleme sayesinde zaman gecikmelerini dikkate alarak ardışık biçimde ele alır. FIR filtresi, girdi tabakasında her bir girdi nöronu için,

$$w_{ji} = [w_{ji}(0), w_{ji}(1), \dots, w_{ji}(m)]^T$$

ağırlık vektörünü elde eder. Burada w_{ji} , i nöronundan j nöronuna taşınan sinyal vektörüdür.

i girdi nöronu için x_i girdi vektörü,

$$x_i(t) = [x_i(t), x_i(t-1), \dots, x_i(t-m)]^T$$

biçiminde gösterilir. Burada m ağırlık vektörüyle ilgili olan gecikme katsayısıdır. TDNN, değişkenin artan zaman noktalarındaki $\{x(t), x(t-1), x(t-2), \dots, x(t-m)\}$ değerlerine dayanarak y çıktısını elde etmeyi amaçlar. x_i girdi vektörüne FIR filtresi uygulanmasıyla i girdisi için FIR ağırlıklı çıktısı, $s_{ji}(t) = w_{ji}^T x_i(t)$ eşitliğiyle elde edilir. Ağın filtre uygulanarak elde edilen yeni girdileri $s_{ji}(t)$ 'lerin toplamına doğrusal olmayan bir aktivasyon fonksiyonu uygulanarak j gizli nöronu için elde edilen çıktı,

$$y_j(t) = f\left(\sum_{i=0}^p s_{ji}(t)\right)$$

eşitliğiyle gösterilir. Burada p girdi sayısı, y_j j 'inci gizli nöronun çıktısı ve f doğrusal olmayan aktivasyon fonksiyonudur [20,21]. k 'inci çıktı nöronu için ağ çıktısı,

$$y_k(t) = \varphi \left(\sum_{j=0}^u w_{kj} f \left(\sum_{i=0}^p s_{ji}(t) \right) \right)$$

eşitliğiyle elde edilir.

TDNN, ağın ağırlıkların ayarlanması sürecinde temporal geriye yayılma algoritmasını kullanır. n'inci adımdan n+1'inci adıma geçişte gizli tabakadaki j gizli nöronundaki ağırlıkların güncellenmesi için temporal geriye yayılma algoritması, $w_{ji}(n+1) = w_{ji}(n) + \eta \delta_j(n-m)x_i(n-m)$ eşitliğiyle gösterilir. Burada m gecikme katsayısı, her bir FIR filtresinin sırasını gösterir. Çıktı tabakasındaki j çıktı nöronu için algoritma, $w_{kj}(n+1) = w_{kj}(n) + \eta \delta_k(n)y_j(n)$ eşitliğiyle gösterilir. Burada $y_j(n)$, i gizli nöronunun çıktısıdır.

Sinir ağının öğrenme süreci, gözlenen çıktı ve beklenen çıktı arasındaki karesel hataların toplamını minimum yapmak amacıyla çalıştırılır. Ağırlıkların güncellenmesi süreci, ağ hatasını minimum yapan ağırlıkları elde edene kadar temporal geriye yayılma algoritmasıyla adım adım sürer [21,22].

2.4. Çok Tabakalı Algılayıcı (MLP)

MLP, geriye yayılma algoritmasıyla çalışan ileri beslemeli bir YSA'dır. MLP, sinir ağı tarafından elde edilen çıktıyı gözlenen değerlerle karşılaştırarak en iyi ağı belirlemeye çalıştığı için danışmanlı öğrenme stratejisiyle çalışır. MLP ağı, girdi ve çıktı tabakasının yanında bir ya da daha fazla gizli tabakadan oluşabilir. Gizli tabakadaki nöronlar, girdi vektörü ile çıktı vektörleri arasındaki doğrusal ve doğrusal olmayan ilişkilerin açıklanması için gerekli bilginin kayba uğramadan taşınmasına izin verir. MLP, özellikle sınıflandırma problemlerinde yaygın olarak kullanılan en popüler YSA'dır [21, 23, 24].

MLP, gözlenen ve beklenen çıktı arasındaki farktan kaynaklanan hatayı en küçük yapan en iyi ağırlıkları belirlemek için geriye yayılma algoritmasını kullanır. Bu algoritma, ağ hatasını en küçük yapmak için ağın ağırlıklarının adım adım hesaplanması sürecidir.

Geriye yayılma algoritmasının t'inci adımdan (t+1)'inci adıma geçişteki en genel ifadesi,

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_i(t)x_j(t)$$

eşitliğiyle gösterilir. Burada η , öğrenme oranı parametresi ve $\delta_i(t)$, t'inci adımda i nöronunun yerel hatasıdır [21,23].

2.5. Radyal Tabanlı Fonksiyon Ağları (RBF)

RBF, MLP'ye alternatif olarak geliştirilen ileri beslemeli bir YSA algoritmasıdır. RBF, ileri beslemeli YSA yapılarına benzer şekilde girdi, gizli ve çıktı tabakasından oluşur. Ancak girdi tabakasından gizli tabakaya geçişteki dönüşüm, radyal tabanlı aktivasyon fonksiyonlarını kullanan doğrusal olmayan bir dönüşümdür.

RBF, öğrenme sürecinde danışmansız ve danışmanlı öğrenme olmak üzere karma strateji kullanır ve böylece öğrenme süreci iki aşamaya ayrılır. Danışmansız öğrenme stratejisi ile gerçekleştirilen ilk aşama, girdi verilerine gizli tabakada radyal tabanlı fonksiyon uygulanması sürecidir. Danışmanlı öğrenme stratejisi ile gerçekleştirilen ikinci aşama ise, gizli tabakadan elde edilen çıktılara doğrusal dönüşüm uygulanarak ağ çıktıları ile gözlenen çıktılar arasındaki hatayı en küçük yapan ağırlıkların belirlenmesi sürecidir [21,23,24].

RBF'de gizli tabakadaki her bir nöron, ortalaması μ_i ve varyansı σ_i olan ve

$$G_i(x) = \exp\left[-\frac{1}{2\sigma_i^2} \sum_{k=1}^p (x_k - \mu_{ik})^2\right]$$

eşitliğiyle gösterilen çok değişkenli Gauss fonksiyonunu kullanır. Burada p bağımsız değişken sayısı ve i gizli tabakada yer alan herhangi bir nöronu gösterir. Bu fonksiyonlar, radyal tabanlı fonksiyonlar olarak isimlendirilir. Sonuç olarak, çıktı tabakasında doğrusal dönüşüm uygulanarak elde edilen ağ çıktıları

$$f(x) = \sum_{i=1}^n w_{ji} (G_i(x))$$

eşitliğiyle elde edilir [21,23].

2.6. Performans Göstergeleri

ARIMA ve YSA yöntemlerinin performansları; hata kareler ortalaması $(HKO = \sum_{i=1}^n (d_i - y_i)^2 / n)$, normleştirilmiş hata kareler

ortalaması $(NHKO = HKO / \text{Var}(d))$ ve ortalama mutlak hata $(OMH = \sum_{i=1}^n |d_i - y_i| / n)$ değerlerine göre karşılaştırıldı [23].

3. Bulgular

3.1. ARIMA Modelinin Katsayı Tahminleri

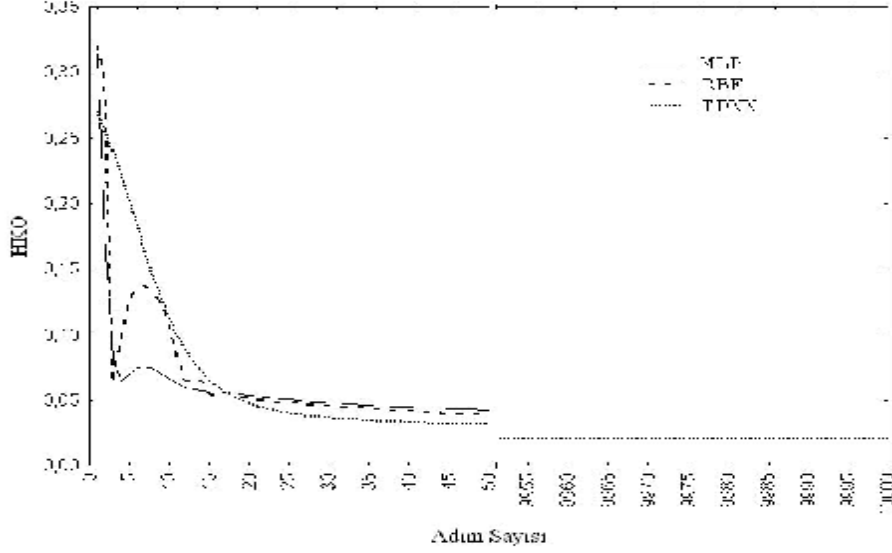
ARIMA yöntemiyle zaman serisi analizi sonucunda AIC, ACF ve PACF kriterlerine dayanarak en iyi model, ARIMA(1,0,1)x(0,1,1) modeli olarak bulundu. Bu model, hem mevsimsel olan hem de mevsimsel olmayan bileşenleri içermektedir. AR (1), MA(1) ve mevsimsel MA(1)'in katsayıları sırasıyla 0.958, 0.339 ve 0.891 değerinde bulundu (Tablo 1). Modelin bütün katsayıları istatistiksel olarak anlamlı bulundu ($p < 0.001$).

Tablo 1. ARIMA modelinin katsayıları

	Katsayı	Standart Hata	t	p
AR(1)	0.958	0.037	26.225	<0.001
MA(1)	0.339	0.093	3.654	<0.001
Mevsimsel MA(1)	0.891	0.168	5.301	<0.001

3.2. Yapay Sinir Ağları Algoritmalarının Sonuçları

Denemeler sonucunda; MLP ve TDNN algoritmasında sinir ağı yapısı, 2 girdi nöronlu (ay ve yıl) girdi tabakası, 4 nöronlu bir gizli tabaka ve 1 çıktı nöronlu (HAV enfeksiyonu) çıktı tabakasından oluşturuldu. RBF algoritmasında ise 2 girdi nöronlu girdi tabakası, 10 nöronlu bir gizli tabaka ve 1 çıktı nöronlu çıktı tabakasından oluşturuldu. Şekil 2, YSA'ların öğrenme sürecinde HKO'larının adım adım değişimini göstermektedir. Öğrenme sürecinde HKO bakımından YSA'ların eğitim setinin performansı, geçerlilik setine karşı test edildi ve geçerlilik setinin HKO'su artmaya başladığı noktada öğrenme süreci sona erdi. Böylece MLP, RBF ve TDNN'nin öğrenme süreci sırasıyla 3401, 1082 ve 10000'inci adımda sonlandı.



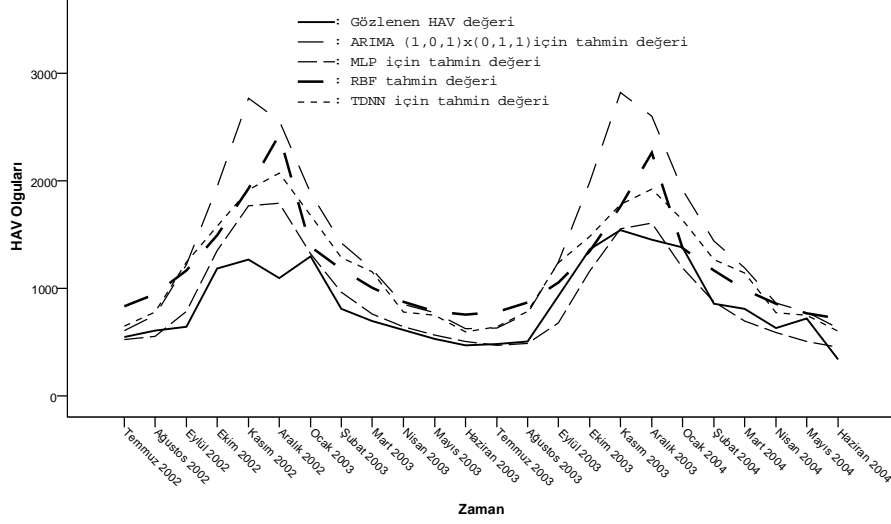
Şekil 2. Eğitim seti için YSA'ların adım sayısına göre HKO'sunun değişimi

3.3. ARIMA Modeli, MLP, RBF ve TDNN Yöntemlerinin Karşılaştırılması

Çalışmamızda yöntemlerin tahmin performanslarının karşılaştırılması için HKO, NHKO ve OMH ölçüleri kullanıldı. Tablo 2'de ARIMA ve YSA yöntemlerinin performans ölçüleri ve Şekil 3'de yöntemlerin test seti verisi için tahmin değerleri verilmiştir. Test seti için HKO, NHKO ve OMH değerleri sırasıyla 43820.407 – 453028.369, 0.346 – 3.432 ve 135.724 – 529.102 aralığında elde edildi. Tablo 2'de görüleceği gibi, en küçük HKO, NHKO ve OMH değeri MLP'de bulundu. MLP'yi ikinci sırada TDNN, üçüncü sırada RBF ve dördüncü sırada ARIMA yöntemi izledi. Dört yöntem arasından minimum hatayla en iyi performansı MLP gösterdi. Böylece MLP'nin, HAV frekansının tahminlenmesinde en uygun yöntem olduğu bulundu.

Tablo 2. Test için YSA ve ARIMA yöntemlerinin performans ölçüleri

	ARIMA	MLP	RBF	TDNN
HKO	453028.369	43820.407	187076.380	146644.297
NHKO	3.432	0.346	1.479	1.159
OMH	529.102	135.724	333.274	321.708



Şekil 3. Test seti için aylara göre gözlenen HAV frekansı ve ARIMA ile YSA yöntemlerinin tahmin edilen değerleri

4. Tartışma

Türkiye, Dünya Sağlık Örgütü'nün verilerine göre HAV enfeksiyonu yönünden orta endemik bölge sınırında yer almaktadır [25]. HAV olgularının tahmini, hastalığın önlenmesi ve kontrolü bakımından çok önemlidir. Bu tahmin probleminin çözümü, hastalıkların bilinmeyen zaman kalıplarının ortaya konmasında yatmaktadır. Çalışmamızda, veri yapılarının genel eğilimi hakkında bilgi edinmek ve geleceğe yönelik tahmin problemlerinin çözümünde araştırmacılara yardım etmek amacıyla gerçek bir veri setini kullanarak dört tahminleme yöntemi (ARIMA, MLP, RBF ve TDNN) karşılaştırıldı. Sonuç olarak, MLP'nin RBF, TDNN ve ARIMA yöntemlerinden daha iyi performans gösterdiği bulundu.

Standart zaman serileri yöntemleri, zaman serisinin durağanlığı ya da uzunluğu gibi model varsayımlarıyla kısıtlanır. Buna karşın, zaman serisi modelleme yaklaşımı olarak YSA'nın kullanımında, hem modelin varsayımları ve doğrusallık hem de zaman serisinin düzensiz örneklenmesi ve kısa zaman aralığı içermesi gibi kısıtlamalar yoktur. YSA algoritmalarının, karmaşık doğrusal olmayan ve dinamik sistemleri modellemede yararlı olduğu bir çok araştırmacı tarafından kabul edilmektedir [26, 27]. YSA, ilgilenilen sistemler hakkında ayrıntılı ön bilgilere ihtiyaç duymaz [28]. YSA ayrıntılı ön bilgiye muhtaç kalan modellerin dışında kalarak, en iyi modelin

araştırılmasında ilişkilerin tam fonksiyonel biçimlerini doğru olarak belirlemek için modelleme uzmanlarına ihtiyacı tamamen ortadan kaldırır [29].

Hükümetler, toplumda görülen bulaşıcı ve bulaşıcı olmayan hastalıkların zaman kalıplarının doğru bir şekilde tahminlenmesiyle aşı maliyetleri ve stokları gibi konular hakkında stratejik sağlık planlamalarını en iyi biçimde yapabilir.

Araştırmacılar, HAV ya da diğer enfeksiyon hastalıkları verilerinin gelecekte sadece tek yönetime bağlı kalarak değerlendirilmesinden kaçınmalı ve veriyi daha iyi tahminleyecek en iyi yöntemle çalışmalarını yürütmeye özen göstermelidirler. Bu düşüncelere dayanarak, yüksek performanslı zaman serileri modelleme yöntemleriyle zaman kalıplarının belirlenmesi ve geleceğe yönelik tahminlerin daha az hatayla yapılabilmesi için eldeki veriye uygun en iyi yöntemi araştırmanın daha iyi olacağını önermekteyiz.

Kaynaklar

1. Lavanchy D. Public health measures in the control of viral hepatitis: a world health organization perspective for the next millennium. *J Gastroenterol Hepatol* 2002; 17: 452-529.
2. Wisnom C, Siegel MA. Advances in the diagnosis and management of human viral hepatitis. *Dent Clin North Am* 2003; 47: 431-47.
3. Committee on Infectious Diseases. Prevention of hepatitis A infections: guidelines for use of hepatitis A vaccine and immune globulin. *Pediatrics* 1996; 98: 1207-1215.
4. Erdogan MS, Oktun M, Tatman-Otkun M, Akata F, Ture M. The Epidemiology of hepatitis a virus infection in children. *European Journal of Epidemiology* 2004; 19: 267-273.
5. Immunization Practices Advisory Committee. Centers for Disease Control and Prevention: Hepatitis B virus: a comprehensive strategy for eliminating transmission in the United States through universal childhood vaccination: Recommendations of the Immunization Practices Advisory Committee (ACIP). *MMWR Morb Mortal Wkly Rep* 1991; 40(RR-13): 1-19.
6. Cuthbert, J. A. (2001). Hepatitis A: old and new. *Clin Microbiol Rev*, 14, 38-58.
7. Bustamante-Sa C, Nobre FF. Forecasting epidemiological time series with backpropagation neural networks. *Midwest Symposium on Circuits and Systems* 1995; 2: 1365-1368.

8. Box GEP, Jenkins G. Time Series Analysis, Forecasting and Control. San Francisco: Holden-Day; 1976.
9. Kurt İ. Yapay Sinir Ağları ile Lojistik Regresyon Analizi'nin Karşılaştırılması (tez). Edirne: Trakya Üniversitesi; 2004.
10. Guan P, Huang D, Zhou B. Forecasting model for the incidence of hepatitis A based on artificial neural network. World J Gastroenterol 2004; 10(24): 3579-3582.
11. Hill T, O'Connor M, Remus W. Neural network models for time series forecasts. Management Science 1996; 42(7): 1082-1092.
12. Tang Z, Fishwick PA. Feedforward neural nets as models for time series forecasting. ORSA Journal of Computing 1993; 5: 374-385.
13. Zhang G. Time Series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 2003; 50: 159-175.
14. De Groot C, Wurtz D. Analysis of univariate time series with connectionist nets: a case study of two classical examples. Neurocomputing 1991; 3: 177-192.
15. Kurunc A, Yurekli K, Cevik O. Performance of two stochastic approaches for forecasting water quality and streamflow data from Yesilirmak River, Turkey. Environmental Modelling & Software 2005; 20: 1195-1200.
16. Zhang G, Patuwo EB, Hu MY. Forecasting with artificial neural network: the state of the art. International Journal of Forecasting 1998; 14: 35-62.
17. Berthouex PM, Brown LC. Statistics for Environmental Engineers. New York: Lewis Publishers; 2002.
18. Akaike H. A new look at statistical model identification. IEEE Transactions on Automatic Control 1974; 19: 716-723.
19. Zhang H, Balaban MÖ, Principe JC. Improving pattern recognition of electronics nose data with time-delay neural networks. Sensors and Actuators B 2003; 96: 385-389.
20. Hassoun MH. Fundamentals of Artificial Neural Networks. Cambridge: Mass:MIT Press; 1995.
21. Haykin S. Neural Network: A Comprehensive Foundation. Upper Saddle River, NJ: Prentice Hall; 1999.
22. Wintoft P, Cander LjR. Short-term prediction of foF2 using time-delay neural network. Phys Chem Earth (C) 1999; 24(4): 343-347.
23. Principe J, Euliano NR, Lefebvre WC. Neural and adaptive systems: fundamentals through simulations. New York: John Wiley & Sons Inc.; 1999.

24. Ture M, Kurt I, Kurum AT, Ozdamar K. Comparing classification techniques for predicting essential hypertension. *Expert Systems with Applications* (in press).
25. Previsani N, Lavanchy D. Hepatitis A. Available at: URL: <http://www.who.int/emc-documents/hepatitis/docs/whocdscsredc2007.html>. 2003.
26. Jerez-Aragones JM, Gomez-Ruiz JA, Ramos-Jimenez G, Munoz-Perez J, Alba-Conejo E. A combined neural network and decision trees model for prognosis of breast cancer relapse. *Artificial Intelligence in Medicine* 2003; 27: 45-63.
27. Traeger M, Eberhart A, Geldner G, Morin AM, Putzke C, Wulf H, Eberhart LH. Prediction of postoperative nausea and vomiting using an artificial neural network. *Anaesthesist* 2003; 52: 1132-1138.
28. Ma L, Khorasani K. New training strategies for constructive neural networks with application to regression problems. *Neural Networks* 2004; 17: 589-609.
29. Callan DE, Kent RD, Guenther FH, Vorperian HK. An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. *J Speech Lang Hear Res* 2000; 43: 721-736.

TEK YÖNLÜ VARYANS ÇÖZÜMLEMESİNDE FARKLI VARYANSLILIK PROBLEMİ

A.Fırat Özdemir, Serdar Kurt

Dokuz Eylül Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü

e-mail: serdar.kurt@deu.edu.tr

Özet

Varyans çözümü yöntemi tıp, mühendislik, tarım, eğitim, psikoloji, sosyoloji, biyoloji, gibi birçok alanda değişim kaynaklarının araştırılması sürecinde başvurulan en güçlü araçlardan biridir. Yöntemin dayandığı olasılık dağılımının (F) geçerli olması ve verdiği sonuçların istenilen hata sınırları içinde kalması için her bir deneme ile elde edilen gözlemlerin eşit varyanslı kitlelerden çekilmesi gerekir. Bu varsayımın sağlanmaması literatürde farklı varyanslılık (heteroscedasticity) olarak adlandırılır. Farklı varyanslılık durumunda F dağılımı çoğunlukla deneme ortalamaları arasında gerçekte olmayan anlamlı farklar bulma yönünde eğilim gösterir. Bu çalışmada tek yönlü varyans çözümlemesinde farklı varyanslılığın sonuçları üzerinde durularak literatürde alternatif olarak yer alan üç test (Welch testi, Kruskal-Wallis testi, Alexander-Govern testi) ile F testi farklı tekrar sayısı, deneme sayısı ve etki büyüklüğü koşullarında gözlenen I.tip hata olasılığı ve testin gücü ölçütleri bakımından karşılaştırılmıştır. Farklı kombinasyonların oluşturduğu durumlar için hangi testin daha uygun olduğuna dair önerilerde bulunulmuştur.

Anahtar Kelimeler: Varyans çözümü, farklı varyanslılık, F testi, Welch testi, Kruskal-Wallis testi, Alexander-Govern testi.

1. Giriş

Varyans analizi genellikle iki den fazla kitle ortalamasının birbirine eşitliğini test ederken yararlanılan bir yöntemdir. Burada test edilen hipotez

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 : \text{En az bir } \mu_i \text{ farklıdır}$$

biçiminde kurulur. Bu hipotez, cevap değişkeninde gözlenen toplam değişim varyans analizi yardımıyla açıklanan ve açıklanamayan olmak üzere iki kısma ayrıldıktan sonra

$$\frac{\sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y}_{..})^2 / k-1}{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 / N-k} = \frac{MSTr}{MSE} \sim F_{k-1, N-k}$$

biçiminde tanımlanan test istatistiğine sahip F testi ile test edilir. F testi'nin uygulanabilmesi için

- Gözlemlerin birbirinden bağımsız elde edilmesi
- Gözlemlerin elde edildiği kitle dağılımlarının normal olması
- Her bir kitle varyansının eşit olması

varsayımlarının sağlanmış olması gereklidir. Her bir deneme ile elde edilen gözlemlerin çekildiği kitle varyanslarının eşit olmaması durumu farklı varyanslılık (heteroscedasticity) olarak adlandırılır. Farklı varyanslılığın gözleendiği durumlarda F testine ait test istatistiğinin dağılımı F dağılımı olmaz (Box, G.E.P., 1954a).

1.1 Farklı Varyanslılık Durumunda F Oranında Gözlenen Değişim

Deneme kareler ortalamasının beklenen değeri

$$\begin{aligned} & E \left[\frac{1}{k-1} \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y}_{..})^2 \right] \\ &= \frac{1}{k-1} \left[\sum_{i=1}^k n_i E(\bar{Y}_i - \mu)^2 - N E(\bar{Y}_{..} - \mu)^2 \right] \end{aligned}$$

$$= \frac{1}{k-1} \left[\sum_{i=1}^k n_i \frac{\sigma_i^2}{n_i} - N \frac{\sum_{i=1}^k n_i \sigma_i^2}{N^2} \right]$$

$$= \frac{1}{N(k-1)} \sum_{i=1}^k (N - n_i) \sigma_i^2$$

biçimindedir. Hata kareler ortalamasının beklenen değeri

$$E \left[\frac{1}{N-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \right] = \frac{1}{N-k} \sum_{i=1}^k (n_i - 1) \sigma_i^2$$

biçimindedir. Bu değerlerin ikisi de kitle varyanslarının ağırlıklı ortalamasıdır. *Deney dengeli olduğunda* bu iki değer birbirine eşit olur ve F oranının değeri 1' e yaklaşır. Ancak deneme kareler ortalamasının varyansı

$$\frac{2\bar{\sigma}^4}{k-1} \left[1 + \frac{(k-2)}{k(k-1)} \frac{\sum_{i=1}^k (\sigma_i^2 - \bar{\sigma}^2)^2}{\bar{\sigma}^4} \right] \quad \bar{S} : \text{varyansların ortalaması}$$

biçimindedir. Kitle varyansları eşit olduğunda parantez içinde yer alan değer 1 olurken kitle varyansları farklı olduğunda 1'den büyük bir değer alır. Farklı varyanslılık durumunda deneme kareler ortalamasının varyansı büyümüş olur. Deneme kareler ortalamasının beklenen değeri de daha büyük değerler alma eğilimi gösterir. Bunun sonucu olarak F oranı büyür ve test, H_0 'ı reddetme yönünde bir eğilim gösterir (liberal eğilim). Gözlenen I. tip hata yapma olasılığı nominal değerinden daha büyük bir değer alır.

Dengesiz deney düzenlerinde ise farklı varyanslılığın etkisi daha ciddi boyutlardadır. Burada iki uç durum incelenmelidir.

- Varyansı büyük olan kitleden büyük örneklemin çekilmesi, varyansı küçük olan kitleden küçük örneklemin çekilmesi (pozitif eşleşme)

- Varyansı büyük olan kitleden küçük örneklemin çekilmesi, varyansı küçük olan kitleden büyük örneklemin çekilmesi (negatif eşleşme)

Pozitif eşleşme durumunda deneme kareler ortalamasının beklenen değeri küçülürken

$$E[DKT] = \frac{1}{N(k-1)} \sum_{i=1}^k (N - n_i) s_i^2$$

hata kareler ortalamasının beklenen değeri büyüme eğilimi gösterir.

$$E[HKT] = \frac{1}{N-k} \sum_{i=1}^k (n_i - 1) s_i^2$$

Bu durumda F oranı küçülür ve test H_0 'ı reddetmeme yönünde bir eğilim gösterir (tutucu eğilim).

Negatif eşleşme durumunda deneme kareler ortalamasının beklenen değeri büyürken, hata kareler ortalamasının beklenen değeri küçülür. F oranı büyür ve test H_0 'ı reddetme yönünde eğilim gösterir (liberal eğilim). Farklı varyanslılığın derecesinin artması, nominal I. tip hata düzeyinden daha çok uzaklaşma ve testin gücünde azalma biçiminde ortaya çıkar.

2. Benzetim Çalışması

F, Welch (W), Kruskal-Wallis (KW) ve Alexander-Govern (AG) testleri farklı tekrar sayısı, deneme sayısı ve etki büyüklüğü koşullarında gözlenen I. tip hata olasılığı ve testin gücü ölçütleri bakımından karşılaştırılmıştır. MINITAB paket programının 13. versiyonu kullanılarak normal dağılmış veriler türetilmiş, $k = 3, 6, 9$ grup için $\alpha = 0.05$ anlam düzeyinde her bir düzen için 10000 tekrar olacak şekilde gözlenen I. tip hata olasılıkları ve güç değerleri hesaplanmıştır. Çalışmada kullanılan düzenler tablo 1 ve tablo 2'de verilmiştir.

Tablo 1. k=3 ve k=6 grup için örneklem genişlikleri ve varyanslar

Düzen	i	K = 3			K = 6						
		1	2	3	1	2	3	4	5	6	
A ₁	n _i	5	5	5	5	5	5	5	5	5	5
	σ_i^2	4	4	4	4	4	4	4	4	4	4
A ₂	n _i	5	5	5	5	5	5	5	5	5	5
	σ_i^2	2	6	10	2	6	10	2	6	10	10
B ₁	n _i	10	10	10	10	10	10	10	10	10	10
	σ_i^2	4	4	4	4	4	4	4	4	4	4
B ₂	n _i	10	10	10	10	10	10	10	10	10	10
	σ_i^2	2	6	10	2	6	10	2	6	10	10
C ₁	n _i	5	10	15	5	10	15	5	10	15	15
	σ_i^2	4	4	4	4	4	4	4	4	4	4
C ₂	n _i	5	10	15	5	10	15	5	10	15	15
	σ_i^2	2	6	10	2	6	10	2	6	10	10
C ₃	n _i	5	10	15	5	10	15	5	10	15	15
	σ_i^2	10	6	2	10	6	2	10	6	2	2
D ₁	n _i	10	20	30	10	20	30	10	20	30	30
	σ_i^2	4	4	4	4	4	4	4	4	4	4
D ₂	n _i	10	20	30	10	20	30	10	20	30	30
	σ_i^2	2	6	10	2	6	10	2	6	10	10
D ₃	n _i	10	20	30	10	20	30	10	20	30	30
	σ_i^2	10	6	2	10	6	2	10	6	2	2

Tablo 2. k=9 için grup için örneklem genişlikleri ve varyanslar

Pattern	i	K = 9								
		1	2	3	4	5	6	7	8	9
A ₁	n _i	5	5	5	5	5	5	5	5	5
	σ_i^2	4	4	4	4	4	4	4	4	4
A ₂	n _i	5	5	5	5	5	5	5	5	5
	σ_i^2	2	6	10	2	6	10	2	6	10
B ₁	n _i	10	10	10	10	10	10	10	10	10
	σ_i^2	4	4	4	4	4	4	4	4	4
B ₂	n _i	10	10	10	10	10	10	10	10	10
	σ_i^2	2	6	10	2	6	10	2	6	10
C ₁	n _i	5	10	15	5	10	15	5	10	15
	σ_i^2	4	4	4	4	4	4	4	4	4
C ₂	n _i	5	10	15	5	10	15	5	10	15
	σ_i^2	2	6	10	2	6	10	2	6	10
C ₃	n _i	5	10	15	5	10	15	5	10	15
	σ_i^2	10	6	2	10	6	2	10	6	2
D ₁	n _i	10	20	30	10	20	30	10	20	30
	σ_i^2	4	4	4	4	4	4	4	4	4
D ₂	n _i	10	20	30	10	20	30	10	20	30
	σ_i^2	2	6	10	2	6	10	2	6	10
D ₃	n _i	10	20	30	10	20	30	10	20	30
	σ_i^2	10	6	2	10	6	2	10	6	2

Tablo 3. k= 3 grup için gözlenen anlam düzeyleri

Düzen	n	σ^2	F	W	KW	AG
A ₁	5,5,5	4,4,4	0,051	0,052	0,044	0,039
A ₂	5,5,5	2,6,10	0,059	0,059	0,054	0,044
B ₁	10,10,10	4,4,4	0,047	0,050	0,048	0,049
B ₂	10,10,10	2,6,10	0,054	0,053	0,053	0,047
C ₁	5,10,15	4,4,4	0,068	0,012	0,046	0,055
C ₂	5,10,15	2,6,10	0,029	0,003	0,029	0,048
C ₃	5,10,15	10,6,2	0,150	0,025	0,082	0,057
D ₁	10,20,30	4,4,4	0,069	0,011	0,045	0,051
D ₂	10,20,30	2,4,6	0,026	0,003	0,029	0,050
D ₃	10,20,30	10,6,2	0,160	0,020	0,088	0,045

Tablo 4. k= 6 grup için gözlenen anlam düzeyleri

Düzen	n	σ^2	F	W	KW	AG
A ₁	5,5,5,5,5,5	4,4,4,4,4,4	0,051	0,080	0,036	0,055
A ₂	5,5,5,5,5,5	2,6,10,2,6,10	0,064	0,087	0,045	0,057
B ₁	10,10,10,10,10,10	4,4,4,4,4,4	0,049	0,058	0,041	0,05
B ₂	10,10,10,10,10,10	2,6,10,2,6,10	0,059	0,060	0,053	0,047
C ₁	5,10,15,5,10,15	4,4,4,4,4,4	0,061	0,012	0,042	0,059
C ₂	5,10,15,5,10,15	2,6,10,2,6,10	0,025	0,007	0,027	0,054
C ₃	5,10,15,5,10,15	10,6,2,10,6,2	0,180	0,025	0,091	0,058
D ₁	10,20,30,10,20,30	4,4,4,4,4,4	0,061	0,007	0,050	0,049
D ₂	10,20,30,10,20,30	2,6,10,2,6,10	0,026	0,005	0,028	0,047
D ₃	10,20,30,10,20,30	10,6,2,10,6,2	0,160	0,012	0,100	0,051

Tablo 5. k= 9 grup için gözlenen anlam düzeyleri

Düzen	n	σ^2	F	W	KW	AG
A ₁	5,5,5,5,5,5,5,5,5	4,4,4,4,4,4,4,4,4	0,049	0,098	0,029	0,060
A ₂	5,5,5,5,5,5,5,5,5	2,6,10,2,6,10,2,6,10	0,060	0,100	0,036	0,057
B ₁	10,10,10,10,10,10,10,10,10	4,4,4,4,4,4,4,4,4	0,049	0,064	0,041	0,049
B ₂	10,10,10,10,10,10,10,10,10	2,6,10,2,6,10,2,6,10	0,067	0,059	0,051	0,049
C ₁	5,10,15,5,10,15,5,10,15	4,4,4,4,4,4,4,4,4	0,057	0,014	0,039	0,060
C ₂	5,10,15,5,10,15,5,10,15	2,6,10,2,6,10,2,6,10	0,023	0,010	0,024	0,058
C ₃	5,10,15,5,10,15,5,10,15	10,6,2,10,6,2,10,6,2	0,200	0,024	0,098	0,056
D ₁	10,20,30,10,20,30,10,20,30	4,4,4,4,4,4,4,4,4	0,069	0,007	0,043	0,053
D ₂	10,20,30,10,20,30,10,20,30	2,6,10,2,6,10,2,6,10	0,027	0,004	0,025	0,049
D ₃	10,20,30,10,20,30,10,20,30	10,6,2,10,6,2,10,6,2	0,200	0,010	0,100	0,047

Tablo 6. k=3 grup için tahminlenen güç değerleri

Düzen	n	σ^2	μ_i	F	W	KW	AG
A ₁₁	5,5,5	4,4,4	2,0,0	0,27	0,27	0,24	0,24
A ₁₂	5,5,5	4,4,4	-1,0,1	0,21	0,21	0,19	0,18
A ₂₁	5,5,5	2,6,10	2,0,0	0,20	0,29	0,20	0,25
A ₂₂	5,5,5	2,6,10	-1,0,1	0,17	0,18	0,15	0,15
B ₁₁	10,10,10	4,4,4	2,0,0	0,58	0,56	0,53	0,54
B ₁₂	10,10,10	4,4,4	-1,0,1	0,45	0,44	0,42	0,42
B ₂₁	10,10,10	2,6,10	2,0,0	0,42	0,58	0,48	0,57
B ₂₂	10,10,10	2,6,10	-1,0,1	0,31	0,35	0,30	0,35
C ₁₁	5,10,15	4,4,4	2,0,0	0,46	0,19	0,33	0,32
C ₁₂	5,10,15	4,4,4	-1,0,1	0,46	0,12	0,35	0,36
C ₂₁	5,10,15	2,6,10	2,0,0	0,21	0,14	0,21	0,44
C ₂₂	5,10,15	2,6,10	-1,0,1	0,21	0,05	0,19	0,33
C ₃₁	5,10,15	10,6,2	2,0,0	0,44	0,13	0,27	0,16
C ₃₂	5,10,15	10,6,2	-1,0,1	0,46	0,11	0,32	0,26
D ₁₁	10,20,30	4,4,4	2,0,0	0,77	0,50	0,66	0,63
D ₁₂	10,20,30	4,4,4	-1,0,1	0,77	0,36	0,68	0,68
D ₂₁	10,20,30	2,6,10	2,0,0	0,51	0,52	0,57	0,81
D ₂₂	10,20,30	2,6,10	-1,0,1	0,48	0,22	0,45	0,64
D ₃₁	10,20,30	10,6,2	2,0,0	0,66	0,27	0,45	0,32
D ₃₂	10,20,30	10,6,2	-1,0,1	0,71	0,29	0,56	0,52

Tablo 7. k=6 grup için tahminlenen güç değerleri

Düzen	n	σ^2	μ_i	F	W	KW	AG
A ₁₁	5,5,5,5,5,5	4,4,4,4,4,4	2,0,0,2,0,0	0,39	0,39	0,31	0,32
A ₁₂	5,5,5,5,5,5	4,4,4,4,4,4	-1,0,1,-1,0,1	0,29	0,31	0,22	0,29
A ₂₁	5,5,5,5,5,5	2,6,10,2,6,10	2,0,0,2,0,0	0,26	0,40	0,72	0,42
A ₂₂	5,5,5,5,5,5	2,6,10,2,6,10	-1,0,1,-1,0,1	0,22	0,25	0,57	0,24
B ₁₁	10,10,10,10,10,10	4,4,4,4,4,4	2,0,0,2,0,0	0,77	0,74	0,25	0,84
B ₁₂	10,10,10,10,10,10	4,4,4,4,4,4	-1,0,1,-1,0,1	0,62	0,60	0,17	0,70
B ₂₁	10,10,10,10,10,10	2,6,10,2,6,10	2,0,0,2,0,0	0,58	0,76	0,64	0,86
B ₂₂	10,10,10,10,10,10	2,6,10,2,6,10	-1,0,1,-1,0,1	0,42	0,48	0,41	0,57
C ₁₁	5,10,15,5,10,15	4,4,4,4,4,4	2,0,0,2,0,0	0,62	0,24	0,44	0,42
C ₁₂	5,10,15,5,10,15	4,4,4,4,4,4	-1,0,1,-1,0,1	0,62	0,13	0,47	0,47
C ₂₁	5,10,15,5,10,15	2,6,10,2,6,10	2,0,0,2,0,0	0,27	0,18	0,28	0,61
C ₂₂	5,10,15,5,10,15	2,6,10,2,6,10	-1,0,1,-1,0,1	0,27	0,06	0,24	0,44
C ₃₁	5,10,15,5,10,15	10,6,2,10,6,2	2,0,0,2,0,0	0,59	0,14	0,39	0,20
C ₃₂	5,10,15,5,10,15	10,6,2,10,6,2	-1,0,1,-1,0,1	0,62	0,13	0,43	0,34
D ₁₁	10,20,30,10,20,30	4,4,4,4,4,4	2,0,0,2,0,0	0,98	0,65	0,84	0,82
D ₁₂	10,20,30,10,20,30	4,4,4,4,4,4	-1,0,1,-1,0,1	0,92	0,86	0,86	0,86
D ₂₁	10,20,30,10,20,30	2,6,10,2,6,10	2,0,0,2,0,0	0,69	0,69	0,76	0,94
D ₂₂	10,20,30,10,20,30	2,6,10,2,6,10	-1,0,1,-1,0,1	0,65	0,25	0,63	0,83
D ₃₁	10,20,30,10,20,30	10,6,2,10,6,2	2,0,0,2,0,0	0,83	0,29	0,62	0,42
D ₃₂	10,20,30,10,20,30	10,6,2,10,6,2	-1,0,1,-1,0,1	0,87	0,34	0,75	0,64

Tablo 8. k=9 grup için tahminlenen güç değerleri

Düzen	n	σ^2	μ_i	F	W	KW	AG
A ₁₁	5,5,5,5,5,5,5,5,5	4,4,4,4,4,4,4,4,4	2,0,0,2,0,0,2,0,0	0,48	0,50	0,40	0,39
A ₁₂	5,5,5,5,5,5,5,5,5	4,4,4,4,4,4,4,4,4	-1,0,1,-1,0,1,-1,0,1	0,36	0,41	0,28	0,29
A ₂₁	5,5,5,5,5,5,5,5,5	2,6,10,2,6,10,2,6,10	2,0,0,2,0,0,2,0,0	0,32	0,52	0,32	0,42
A ₂₂	5,5,5,5,5,5,5,5,5	2,6,10,2,6,10,2,6,10	-1,0,1,-1,0,1,-1,0,1	0,25	0,33	0,20	0,24
B ₁₁	10,10,10,10,10,10,10,10,10	4,4,4,4,4,4,4,4,4	2,0,0,2,0,0,2,0,0	0,88	0,86	0,86	0,84
B ₁₂	10,10,10,10,10,10,10,10,10	4,4,4,4,4,4,4,4,4	-1,0,1,-1,0,1,-1,0,1	0,75	0,72	0,71	0,70
B ₂₁	10,10,10,10,10,10,10,10,10	2,6,10,2,6,10,2,6,10	2,0,0,2,0,0,2,0,0	0,71	0,88	0,78	0,87
B ₂₂	10,10,10,10,10,10,10,10,10	2,6,10,2,6,10,2,6,10	-1,0,1,-1,0,1,-1,0,1	0,53	0,58	0,51	0,57
C ₁₁	5,10,15,5,10,15,5,10,15	4,4,4,4,4,4,4,4,4	2,0,0,2,0,0,2,0,0	0,72	0,30	0,56	0,53
C ₁₂	5,10,15,5,10,15,5,10,15	4,4,4,4,4,4,4,4,4	-1,0,1,-1,0,1,-1,0,1	0,73	0,16	0,60	0,59
C ₂₁	5,10,15,5,10,15,5,10,15	2,6,10,2,6,10,2,6,10	2,0,0,2,0,0,2,0,0	0,33	0,24	0,37	0,73
C ₂₂	5,10,15,5,10,15,5,10,15	2,6,10,2,6,10,2,6,10	-1,0,1,-1,0,1,-1,0,1	0,33	0,07	0,31	0,55
C ₃₁	5,10,15,5,10,15,5,10,15	10,6,2,10,6,2,10,6,2	2,0,0,2,0,0,2,0,0	0,70	0,17	0,41	0,24
C ₃₂	5,10,15,5,10,15,5,10,15	10,6,2,10,6,2,10,6,2	-1,0,1,-1,0,1,-1,0,1	0,72	0,14	0,51	0,41
D ₁₁	10,20,30,10,20,30,10,20,30	4,4,4,4,4,4,4,4,4	2,0,0,2,0,0,2,0,0	0,98	0,77	0,95	0,92
D ₁₂	10,20,30,10,20,30,10,20,30	4,4,4,4,4,4,4,4,4	-1,0,1,-1,0,1,-1,0,1	0,98	0,58	0,95	0,95
D ₂₁	10,20,30,10,20,30,10,20,30	2,6,10,2,6,10,2,6,10	2,0,0,2,0,0,2,0,0	0,84	0,82	0,89	0,98
D ₂₂	10,20,30,10,20,30,10,20,30	2,6,10,2,6,10,2,6,10	-1,0,1,-1,0,1,-1,0,1	0,80	0,37	0,76	0,92
D ₃₁	10,20,30,10,20,30,10,20,30	10,6,2,10,6,2,10,6,2	2,0,0,2,0,0,2,0,0	0,93	0,36	0,74	0,53
D ₃₂	10,20,30,10,20,30,10,20,30	10,6,2,10,6,2,10,6,2	-1,0,1,-1,0,1,-1,0,1	0,95	0,42	0,86	0,82

3. Sonuç ve Yorumlar

Elde edilen sonuçlar, deney düzeni, örneklem büyüklüğü ve kitle varyanslarının eşit olup olmamasına göre farklı düzenler için yorumlanmıştır. Yorumlamada kriter, gözlenen I. tip hata yapma olasılığının, araştırmacı tarafından belirlenen nominal değerine ($\alpha = 0,05$) olan yakınlığıdır. Bu konuda Bradley (1978) gözlenen I. tip hata yapma olasılığının $0,5\alpha < \alpha < 1,5\alpha$ aralığında olması durumunda testin farklı varyanslılığa karşı güçlü kabul edilebileceğini belirtmiştir. Ancak 0,05 değerine mutlak olarak aynı uzaklıkta olan iki değerden küçük olanı tercih edilmiştir. Güç değeri, I.hata yapma olasılığından doğrudan etkilenen bir büyüklük olduğu için gözlenen I. tip hata yapma olasılıkları yaklaşık aynı olan testler arasında tercih yapılırken kullanılmıştır. Benzetim çalışmasının bileşenlerine göre kullanılan düzenler aşağıdaki biçimde sıralanmış ve her bir durumda kullanılması önerilen testler tablo 9'da gösterilmiştir.

- Dengeli-Küçük Örneklem-Eşit Varyans: A_1
Dengeli-Küçük Örneklem-Farklı Varyans: A_2
Dengeli-Büyük Örneklem-Eşit Varyans: B_1
Dengeli-Büyük Örneklem-Farklı Varyans: B_2
Dengesiz-Küçük Örneklem-Eşit Varyans: C_1
Dengesiz-Küçük Örneklem-Farklı Varyans-Pozitif Eşleşme: C_2
Dengesiz-Küçük Örneklem-Farklı Varyans-Negatif Eşleşme: C_3
Dengesiz-Büyük Örneklem-Eşit Varyans: D_1
Dengesiz-Büyük Örneklem-Farklı Varyans-Pozitif Eşleşme: D_2
Dengesiz-Büyük Örneklem-Farklı Varyans-Negatif Eşleşme: D_3

Tablo 9. Farklı Benzetim Çalışması düzenlerine göre kullanılması önerilen testler

		Eşit Varyans	Farklı Varyans	
Dengeli	Küçük Örneklem	F	KW - AG	
	Büyük Örneklem	F	AG	
Dengesiz	Küçük Örneklem	F-AG	PE: AG	NE: AG
	Büyük Örneklem	F-AG	PE: AG	NE: AG

PE: Pozitif Eşleşme

NE: Negatif Eşleşme

4. Kaynaklar

1. Alexander RA, Govern DM. A new and Simpler Approximation for ANOVA Under Variance Heterogeneity. Journal of Educational Statistics 1994; 19: 91-101.
2. Bishop TA. Heteroscedastic ANOVA, MANOVA and Multiple Comparisons (Doktora Tezi) The Ohio State University Department of Statistics; 1976.
3. Box GEP. Some theorems on quadratic forms applied in the study of analysis of variance problems. Annals of Mathematical Statistics 1954a; 25: 290-302.
4. Bradley JV. Robustness? British Journal of Mathematical and Statistical Psychology 1978; 31:144-152.

5. Hartung J, Argaç D, Makambi KH. Small Sample Properties of Tests on Homogeneity in One-Way Anova and Meta-Analysis. *Statistical Papers* 2002; 43: 197-235.
6. Miller. *Beyond ANOVA, basics of applied statistics*. Newyork: John Wiley; 1986.
7. Welch BL. On the comparison of several mean values. *Biometrika* 1951; 38: 330-336.

DENGELENMEMİŞ VERİLERDE VARYANS ANALİZİ TEKNİĞİ VE BİR UYGULAMA

Hayal Boyacıoğlu¹, Saim Kendir¹, Soner Duman²

Ege Üniversitesi Fen Fakültesi İstatistik Bölümü, İzmir¹

Ege Üniversitesi Tıp Fakültesi, İzmir²

e-mail: hayal.boyacioglu@ege.edu.tr

Özet

Bu çalışmada Dengelenmemiş Verilerde Varyans Analizi Tekniği konusu ele alınmış, buna ilişkin, Tip I (Weighted Means), Tip II (Method of Fitting Constants), Tip III (Weighted Squares of Means) yöntemleri incelenmiştir. Çalışmanın amacı, en dengesiz koşullarda bile, anlamlı değerlendirmelerin yapılabileceğini göstermek suretiyle bu alandaki çalışmalara önemli bir katkı sağlamaktır.

Anahtar Kelimeler: Fitting Constants, nonortogonal, dengelenmemiş veri.

Tartışma

A ve B gibi iki faktörlü dengelenmiş bir deneme deseninde, iki-yönlü varyans analizi çözümlenmesi yapıldığında, satır, sütun ve genel toplamlar üzerinden toplamsallık varsayımı sağlanmaktadır. Diğer bir ifade ile $KT_{Genel} = KT_A + KT_B + KT_{AxB} + KT_{Hata}$ 'dır. Bundan dolayı yapılacak olan varyans analizi çözümlenmesi üç yöntem ile aynı sonucu verecektir. (Tip I=Tip II=Tip III). Dengelenmemiş deneme desenlerinde iki-yönlü varyans analizi çözümlenmesi için, en uygun yöntemin belirlenmesinde, dengesizliğin durumu çok önemlidir. Dengesizlik, kitlenin bir özelliği ise, Tip I yöntemi, kitlenin bir özelliği olmayıp, örnekte yapay olarak ortaya çıkmış ise ve etkileşim'in önemsiz olduğu varsayılıyor ise bununla birlikte deneme deseninde boş hücre varsa Tip II, aksi durumda Tip III yöntemi ile çözüm en iyi sonucu verecektir.

VARYANS HETEROJENLİĞİNDE GRUP ORTALAMALARININ KARŞILAŞTIRILMASINDA KULLANILAN ALTERNATİF YAKLAŞIMLAR

Semra Erdoğan, Handan Ankaralı, Sema Erden

Mersin Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı- MERSİN
e-mail: hcamdeviren@yahoo.com

Özet

Yaygın olarak kullanılan varyans analizi yönteminin uygulanabilmesi için bazı ön şartların yerine gelmiş olması gerekmektedir. Bu ön şartlar yerine geldiği durumda varyans analizi yöntemi güçlü sonuçlar vermektedir. Ancak ön şartlardan biri olan varyansların homojenliği ön şartı yerine gelmediği durumda parametrik ve parametrik olmayan bazı testler kullanılmaktadır. Bu çalışmada, alternatif testler tanıtılmış ve normal dağılım gösteren ancak varyansların homojen olmadığı durum için bir simülasyon çalışması planlanarak Anova, Welch, Brown-Forsythe ve Kruskal-Wallis testleri gerçekleşen I. tip hata olasılıkları bakımından karşılaştırmalı olarak incelenmiştir. Aynı zamanda bu verilere logaritmik transformasyon yapılarak aynı testler uygulanmış ve bunlardan elde edilen I. Tip hata olasılıkları belirlenmiştir.

Herhangi bir grubun varyansı diğer grupların varyanslarının en fazla 4 katı olduğunda, genel olarak Welch testinin I. Tip hatası en uygun bulunurken, 3 grubun varyansı da birbirinden farklı olduğu durumlarda ve küçük örneklerde Brown-Forsythe testi daha iyi sonuçlar vermiştir. Bu koşulda örnek genişlikleri orta ve büyük olduğu zaman ise yine Welch testinin I. Tip hatası başlangıçta belirlenen düzeye en yakın bulunmuştur.

Anahtar Kelimeler: Heterojen varyans, Anova, Kruskal-Wallis, Welch testi, Brown-Forsythe testi, James Second Order testi.

Giriş

Aynı popülasyondan rasgele seçilen ikiden fazla örneklemin ortalamaları arasında fark yoktur şeklinde kurulan hipotezi test etmek için bilindiği gibi varyans analizi yöntemi kullanılmaktadır. Varyans analizi yöntemini uygulayabilmek için belli başlı ön şartların yerine gelmiş olması

gerekmektedir. Bu ön şartlar sırasıyla verilerin normal dağılım göstermesi ve grup varyanslarının homojen olması şeklinde özetlenebilir. Bu ön şartların gerçekleştiği durumlarda verilere varyans analizi yöntemi uygulandığı zaman güçlü sonuçlar elde edilmektedir (1, 2). Ancak pratikte bu tip ideal koşullar bir çok çalışmada sağlanmamaktadır. Ortalamalar karşılaştırılırken grup varyanslarının heterojen olması durumunda, hata varyansı büyüyecek ve bu test istatistiğinin (F) olması gerekenden daha küçük çıkmasına neden olacaktır. Bu durumda ise reddedilmesi gereken bazı hipotezler kabul edilecek ve II. Tip hata yapma olasılığı artacaktır. Bu ön şartın yerine gelmediği durumlarda, Kruskal-Wallis testi ve son yıllarda da Welch testi diğer bazı alternatiflerine göre daha yaygın kullanılmaktadır (2). En uygun analiz yöntemi belirlenirken dikkat edilecek olan bir nokta ise başlangıçta belirlenmiş olan I. Tip hata yapma olasılığını değiştirmeyecek bir yöntem kullanmaktır. Bu yöntemlerden bazıları şöyle sıralanabilir. Welch testi, Brown-Forsythe testi, James Second Order testi, Rank transform yöntemi, % 15 trim edilmiş değerlere veya monoton transformasyon uygulanmış verilere varyans analizi veya Kruskal-Wallis testidir (2, 3, 4, 5, 6).

Bu çalışmada, varyans analizinin ön şartlarından biri olan varyansların homojenliği ön şartı yerine gelmediğinde kullanılan parametrik ve parametrik olmayan alternatif testlerin özelliklerine değinilmiş ve bu testlerin uygulama adımları bir örnek üzerinde gösterilmiştir. Ayrıca pratikte bu amaçla sık kullanılan 4 testin varyansların heterojenliği durumunda, I. Tip hata yapma olasılıklarının nasıl bir değişim gösterdiği bir simülasyon çalışması ile incelenmiştir.

Gereç ve Yöntem

Çalışmada kullanılan testlerin özellikleri ve işlem adımları sırasıyla aşağıdaki gibi tanımlanmıştır.

Parametrik Testler

Welch Testi

Welch testi varyansların eşit olmadığı durumlarda ortalamalar arasında fark yoktur şeklinde kurulan hipotezi test eder. Varyansların eşit olmadığı durumlarda I. Tip hatayı kontrol altında tutmak için düzeltilmiş serbestlik derecesi kullanır. Bu teste ait ön şartlar; her bir gruptaki bağımlı değişkene ait artık değerlerin normal dağılım göstermesi ve bu artıkların her birinin birbirinden bağımsız olması şeklinde özetlenebilir ve aşağıdaki eşitlik ile hesaplanmaktadır.

$$F' = \frac{\sum_{j=1}^J d_j (\bar{x}_j - \bar{x}_w)^2}{(J-1) \left[1 + \frac{2A(J-2)}{J^2-1} \right]}$$

Burada ;

$$d_j = \frac{n_j}{S_j^2} \quad \bar{x}_w = \frac{\sum_{j=1}^J c_j \bar{x}_j}{D} \quad D = \sum_{j=1}^J d_j \quad \text{ve } A = \sum_{j=1}^J \frac{\left(1 - \frac{d_j}{D} \right)^2}{v_j}$$

Şeklinde elde edilir. v_j Değeri serbestlik derecesini göstermektedir. Hesaplanan F değeri aşağıda eşitlikleri verilen v_1 ve v_2 serbestlik dereceli F tablo değeri ile karşılaştırılır (4,7).

$$v_j = n_j - 1 \quad v_1 = J - 1 \quad v_2 = \frac{J^2 - 1}{3A}$$

James Second Order Testi

İşlem karmaşıklığından dolayı bu test genellikle tercih edilmemektedir. Aşağıda bu teste ait hesaplama adımları verilmiştir.

$$U = \sum_{j=1}^J d_j (\bar{Y}_{.j} - \bar{Y}_{..}^*)^2$$

$$d_j = \frac{n_j}{S_j^2} \quad \bar{Y}_{..}^* = \frac{\sum_{j=1}^J d_j \bar{Y}_{.j}}{D} \quad D = \sum_{j=1}^J d_j$$

Hesaplanan U test istatistiği, aşağıda verilen h (α) değerinden büyük olursa H_0 hipotezi reddedilir.

$$\begin{aligned}
h(\alpha) = & c + (1/2)(3\chi_4 + \chi_2) \sum_{j=1}^J (1 - d_j/D)^2 / v_j \\
& + \left\{ (1/16)(3\chi_4 + \chi_2)^2 (1 - (J-3)/c) \left[\sum_{j=1}^J (1 - d_j/D)^2 / v_j \right]^2 \right. \\
& + (1/2)(3\chi_4 + \chi_2) \{ (8R_{23} - 10R_{22} + 4R_{21} - 6R_{12}^2 + 8R_{12}R_{11} - 4R_{11}^2) \\
& + (2R_{23} - 4R_{22} + 2R_{21} - 2R_{12}^2 + 4R_{12}R_{11} - 2R_{11}^2)(\chi_2 - 1) \\
& + (1/4)(-R_{12}^2 + 4R_{12}R_{11} - 2R_{12}R_{10} - 4R_{11}^2 + 4R_{11}R_{10} - R_{10}^2) \\
& \times (3\chi_4 - 2\chi_2 - 1) \} \\
& + (R_{23} - 3R_{22} + 3R_{21} - R_{20})(5\chi_6 + 2\chi_4 + \chi_2) \\
& + (3/16)(R_{12}^2 - 4R_{23} + 6R_{22} - 4R_{21} + R_{20})(35\chi_8 + 15\chi_6 + 9\chi_4 + 5\chi_2) \\
& + (1/16)(-2R_{22} + 4R_{21} - R_{20} + 2R_{12}R_{10} - 4R_{11}R_{10} + R_{10}^2) \\
& \times (9\chi_8 - 3\chi_6 - 5\chi_4 - \chi_2) \\
& + (1/4)(-R_{22} + R_{11}^2)(27\chi_8 + 3\chi_6 + \chi_4 + \chi_2) \\
& + (1/4)(R_{23} - R_{12}R_{11})(45\chi_8 + 9\chi_6 + 7\chi_4 + 3\chi_2) \}.
\end{aligned}$$

Buradaki c, (k-1) serbestlik dereceli χ^2 değeri ve R_{st} ise

$$R_{st} = \sum_{j=1}^J \frac{1}{v_j^S} \left(\frac{d_j}{D} \right)^t$$

eşitliği ile hesaplanmaktadır (7,8,9).

Brown-Forsythe Testi

Aşağıda bu teste ait hesaplama adımları verilmiştir.

$$F'' = \frac{GAKT}{B}$$

$$B = \sum_{j=1}^J \left(1 - \frac{n_j}{N}\right) S_j^2 \quad v_1 = J-1 \text{ ve } v_2 = \frac{B^2}{\sum_{j=1}^J \left[\left(1 - \frac{n_j}{N}\right) S_j^2 \right]^2}$$

serbestlik dereceli F tablo değeri ile karşılaştırılır (7,10).

Parametrik Olmayan Testler

Varyansların heterojenliği durumunda pratikte en çok tercih edilen testler arasında Kruskal-Wallis ve Friedman testleri gelmektedir. Bu testlerden Kruskal-Wallis tek yönlü varyans analizine alternatif bir test iken, Friedman testi iki yönlü ve etkisiz varyans analizi testine alternatiftir. Bunların dışında Rank transform metodu hem tek yönlü, hem iki yönlü hem de daha kompleks varyans analizi yöntemlerine alternatif olarak kullanılmaktadır (7).

Rank Transform Metot

Bu metot, gerçek gözlemleri ranklara dönüştürerek, bu değerler için parametrik varyans analizi modeli kullanır. Kısaca ranklara klasik varyans analizi uygulanmış olur ve elde edilen test istatistiği F olup grup ve hata serbestlik dereceli tablo değeri ile karşılaştırılarak ilgili hipotez kontrol edilir (7,11).

Simülasyon

Bu çalışma ile varyans analizinin ön şartı olan varyansların homojen olmaması durumunda; Anova, Welch, Brown-Forsythe ve Kruskal-Wallis test istatistiklerini ve de verilere logaritmik transformasyon yapıldıktan sonra bu testleri tekrarlayarak ortalamalar arasında fark yoktur hipotezini test ederken I. Tip hata yapma olasılığı araştırılmıştır. Bu amaçla bir simülasyon çalışması yapılmıştır. Bu veriler bilgisayar ortamında MNRDGv.2 (12) programı ile üretildikten sonra SPSS 11.5 paket programı yardımıyla hesaplamalar yapılmıştır. Varyansların homojenliği ön şartı yerine gelmemesi durumunda hangi yöntem kullanılacağına karar verebilmek amacıyla ortalamaların eşit, varyansların heterojen olduğu kombinasyonlar oluşturulmuştur. Ayrıca 3 grup ele alınarak gruplardaki gözlem sayıları eşit kabul edilmiştir. Bu grup büyüklükleri ise sırasıyla küçük, orta ve büyük örnek genişlikleri olarak belirlenmiştir. Oluşturulan kombinasyonlara sırasıyla Anova, Welch, Brown - Forsythe ve Kruskal-Wallis test istatistikleri uygulanmıştır. Her kombinasyon

için 1000 kez deneme yapılmıştır. Her bir deneme sonucunda orijinal veriler için ve logaritması alınmış veriler için bu dört test istatistiği uygulanarak verilerde reddedilen H_0 hipotezi sayıları belirlenmiş ve bu sayılar tablolar halinde verilmiştir.

Yapılan simülasyon çalışmasında, ortalamaları eşit, grup sayısı 3 ve gruplardaki gözlem sayıları eşit iken varyans oranları 1:4:1 ve örnek genişlikleri 5, 20 ve 100 olmak üzere 3 farklı durum için 3000 farklı kombinasyon, varyans oranları 1:25:100 ve örnek genişlikleri 5, 20 ve 100 için de 3000 farklı kombinasyon olmak üzere toplam 6000 kombinasyon incelenmiştir.

Bulgular

Alternatif Yaklaşımlara Ait Uygulamalar

Varyansların heterojenliği söz konusu olduğunda kullanılan alternatif testlerin sonuçlarını göstermek amacıyla bir örnek veri seti ele alınmıştır. Bu örnekte hemoglobin düzeyleri bakımından, Jüvenil romatoid artrit (JRA) hastalığına yakalanan çocuklar, hematolojik açıdan demir eksikliği anemisi (DEA) olan ve transferin saturasyonu % 16'nın altında olan çocuklar ve sağlıklı çocuklar arasında anlamlı bir farklılığın olup olmadığını araştırmak amacıyla yapılan bir çalışmada benzer yaşlardan kızlar, her bir gruba 10'ar çocuk seçilmiş ve hemoglobin ölçümleri aşağıdaki gibi tespit edilmiştir. Hemoglobin düzeyleri bakımından gruplar arasında anlamlı bir farklılık var mıdır?

Tablo 1. 10 çocuğa ait hemoglobin düzeyleri

Hemoglobin Düzeyleri (gr/dl)		
JRA Grubu	DEA Grubu	Kontrol Grubu
6.1	7.5	10.2
5.9	3.2	11.5
6.9	9.2	8.7
5.6	3.4	9.9
5.5	1.1	9.9
5	3.3	9.4
5.1	4.6	9.1
4.7	4	9.6
6.3	3.9	8
7.1	5.2	10.2

Bu veri seti için varyansların homojenlik testi sonuçları aşağıdaki tabloda görülmektedir. Bu sonuç incelendiğinde varyansların homojen olmadığı görülmektedir ($p=0.038$).

Test of Homogeneity of Variances

V1			
Levene Statistic	df1	df2	Sig.
3,694	2	27	,038

Parametrik Testler

Basit Varyans Analizi

ANOVA

V1					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	140,554	2	70,277	31,135	,000
Within Groups	60,944	27	2,257		
Total	201,498	29			

Bu sonuca göre en az bir grup diğerlerinden farklılık göstermektedir.

Welch Testi

Descriptive Statistics

GRUP	N	Mean	Variance
1,000 V1	10	5,822	,609
Valid N (listwise)	10		
2,000 V1	10	4,546	5,270
Valid N (listwise)	10		
3,000 V1	10	9,641	,894
Valid N (listwise)	10		

$$d_1 = \frac{10}{0.609} = 16.4 \quad d_2 = \frac{10}{5.27} = 1.9 \quad d_3 = \frac{10}{0.894} = 11.19$$

$$D = 16.4 + 1.9 + 11.19 = 29.49$$

$$\bar{X}_W = \frac{16.4 * 5.822 + 1.9 * 4.546 + 11.19 * 9.641}{29.49} = 7.19$$

$$A = \frac{\left(1 - \frac{16.4}{29.49}\right)^2}{9} + \frac{\left(1 - \frac{1.9}{29.49}\right)^2}{9} + \frac{\left(1 - \frac{11.19}{29.49}\right)^2}{9} = 0,162$$

$$F' = \frac{16.4 * (5.822 - 7.19)^2 + 1.9 * (4.546 - 7.19)^2 + 11.19 * (9.641 - 7.19)^2}{2 \left[1 + \frac{2 * 0.162 * 1}{8}\right]}$$

$$F' = 53.4$$

$$v_1 = 2 \quad v_2 = 16.4 \quad F_{2, 16, 0.05} = 3.63 \text{ Ho red.}$$

Welch testine göre en az bir grup diğerlerinden farklılık göstermektedir.

James Second Order Testi

$$d_1 = \frac{10}{0.609} = 16.4 \quad d_2 = \frac{10}{5.27} = 1.9 \quad d_3 = \frac{10}{0.894} = 11.19$$

$$D = 16.4 + 1.9 + 11.19 = 29.49$$

$$\bar{Y}_{..} = \frac{16.4 * 5.822 + 1.9 * 4.546 + 11.19 * 9.641}{16.4 + 1.9 + 11.19} = 7.19$$

$$U = 16.4 * (5.822 - 7.19)^2 + 1.9 * (4.546 - 7.19)^2 + 11.19 * (9.641 - 7.19)^2$$

$$U = 111.19$$

$$R_{10} = 0.375 \quad R_{11} = 0.125 \quad R_{12} = 0.33 \quad R_{20} = 0.047$$

$$R_{21} = 0.016 \quad R_{22} = 0.007 \quad R_{23} = 0.0035$$

$$h(\alpha) = 5.99 + 0.846 + \left\{ 3.65 + \frac{1}{2} * 34.46 * [(-0.4215) + (-0.6345) + (-0.802)] + (-1.4515) + (28.196) \right. \\ \left. + (0.615) + (1.018) + (-8.4536) \right\}$$

$$h(\alpha) = -57.669$$

$U > h(\alpha)$ olduğu için H_0 reddedilir. Bu sonuca göre ortalamalardan en az biri farklılık göstermektedir.

Brown-Forsythe Testi

$$B = \left(1 - \frac{10}{30}\right) * 0.609 + \left(1 - \frac{10}{30}\right) * 5.27 + \left(1 - \frac{10}{30}\right) * 0.894 = 4.54$$

$$F' = \frac{\text{GAKT}}{B} = \frac{140.554}{4.54} = 30.96 \quad \nu_1 = 2 \text{ ve } \nu_2 = 14.27$$

$$F_{2, 14, 0.05} = 3.74 \quad H_0 \text{ red.}$$

Brown-Forsythe testine göre en az bir grup diğerlerinden farklılık göstermektedir.

Parametrik Olmayan Testler

Kruskal-Wallis Testi

Test Statistics^{a,b}

	V1
Chi-Square	19,889
df	2
Asymp. Sig.	,000

a. Kruskal Wallis Test

b. Grouping Variable: GRUP

Bu sonuca göre en az bir grup diğerlerinden farklılık göstermektedir.

Rank Transform Metot

Hemoglobin Düzeyleri (gr/dl)		
JRA Grubu	DEA Grubu	Kontrol Grubu
6.1 (15)	7.5 (19)	10.2 (28.5)
5.9 (14)	3.2 (2)	11.5 (30)
6.9 (17)	9.2 (23)	8.7 (21)
5.6 (13)	3.4 (4)	9.9 (26.5)
5.5 (12)	1.1 (1)	9.9 (26.5)
5 (9)	3.3 (3)	9.4 (24)
5.1 (10)	4.6 (7)	9.1 (22)
4.7 (8)	4 (6)	9.6 (25)
6.3 (16)	3.9 (5)	8 (20)
7.1 (18)	5.2 (11)	10.2 (28,5)

Descriptives

RANK

	N	Mean	Std. Deviation	Std. Error
1,000	10	13,2000	3,42540	1,08321
2,000	10	8,1000	7,41545	2,34497
3,000	10	25,2000	3,40914	1,07806
Total	30	15,5000	8,80145	1,60692

ANOVA

RANK

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1541,400	2	770,700	29,512	,000
Within Groups	705,100	27	26,115		
Total	2246,500	29			

Bu metot'a göre en az bir grup diğerlerinden farklılık göstermektedir.

ÖZET SONUÇLAR TABLOSU

Tablo 2. Parametrik ve Parametrik Olmayan Testlerin Sonuçları

PARAMETRİK TESTLER				PARAMETRİK OLMAYAN TESTLER	
ANOVA	WELCH	B-F	JSO	K-W	R-T
F = 31.132	F' = 53.439	F'' = 30.96	U = 111.19	$\chi^2 = 19.889$	F = 29.512
Sd ₁ : 2 Sd ₂ = 27	Sd ₁ : 2 Sd ₂ = 16.476	Sd ₁ : 2 Sd ₂ = 14.278	H(α) = - 57.66	Sd : 2	Sd ₁ : 2 Sd ₂ = 27
Ho RED	Ho RED	Ho RED	Ho RED	Ho RED	Ho RED
En az bir grup diğerlerinden farklılık göstermektedir.	En az bir grup diğerlerinden farklılık göstermektedir.	En az bir grup diğerlerinden farklılık göstermektedir.	En az bir grup diğerlerinden farklılık göstermektedir.	En az bir grup diğerlerinden farklılık göstermektedir.	En az bir grup diğerlerinden farklılık göstermektedir.

Simülasyon Sonuçları

N =5, 20 ve 100 lük setler için 1000 tekrarlı olarak üretilen ve 1:4:1 ve 1:25:100 varyans kombinasyonlarında ortalamalar arasında fark yoktur hipotezine ait testin sonuçları incelendiği zaman, orijinal verilere ve bu verilere logaritmik transformasyon yapıldıktan sonra Anova, Welch, Brown-Forsythe ve Kruskal-Wallis testleri uygulandığında I. Tip hata yapma olasılığı sayıları Tablo 3 ve Tablo 4'de verilmiştir.

Tablo 3. H₀: $\mu_1 = \mu_2 = \mu_3$ Hipotezini Reddederken Yanılma Olasılığı

Kombinasyon	TESTLER								
	Orijinal Ölçümler için Anova	Log. için Anova	Orijinal Ölçümler için Welch	Log. için Welch	Orijinal Ölçümler için Brown-Forsythe	Log. için Brown-Forsythe	Orijinal Ölçümler için K-W	Log. için K-W	
$\sigma_1^2 : \sigma_2^2 : \sigma_3^2 = 1 : 4 : 1$	n _i = 5	5.8	5.7	4.9	4.8	4.6	4.8	4.9	4.9
	n _i = 20	6.5	7.1	4.9	5.5	6.2	6.9	6.4	6.4
	n _i = 100	5.5	5.4	4.5	4.6	5.3	5.5	5.7	5.7

Grup sayısı 3, varyans oranları 1:4:1 olarak belirleyip her bir gruptaki örnek genişliğini 5 aldığımız zaman ortalamalar arasında fark yoktur şeklinde kurulan hipotezi test etmek için kullanılan testlere ait I. Tip hata yapma olasılıkları Tablo 3' te verilmiştir. Tablo incelendiğinde Welch veya Kruskal-Wallis testlerinin I. Tip hata yapma olasılıkları aynı bulunmuş ve diğer iki teste göre başlangıçta belirlenen sınıra daha yakın bulunmuştur. Bu koşul için en büyük hata Anova testinden elde edilmiştir.

Örnek genişliğini artırıp orta büyüklükte bir örnek seçtiğimizde (n=20) ortalamalar arasında fark yoktur hipotezinin test edilmesi amacıyla kullanılan testler karşılaştırıldığında, orijinal ve logaritması alınmış değerlere Welch testi uygulandığı zaman elde edilen sonuçların en iyi düzeyde olduğu görülmüştür. Orijinal değerlere Welch testi yapıldığında H_0 hipotezini reddetme olasılığı % 4.9, logaritmik transformasyon yapılmış değerlere Welch testi uygulandığında ise bu hata oranı % 5.5'e çıkmıştır. Fakat bu iki hata oranı da kabul edilebilir sınırlar içerisindedir. Buna karşılık hem orijinal hem de logaritması alınmış değerlere Anova, Brown-Forsythe ve Kruskal-Wallis testleri uygulandığında, I. Tip hata olasılığı % 5'ten büyük çıkmış yani hata oranı kabul edilebilir sınırların dışına çıkmıştır. Bu sonuçlar bize söz konusu kombinasyonda denenen testlerin hepsinin H_0 hipotezini reddetme olasılıklarının beklenene (0.05) göre biraz yüksek olduğunu yani I. Tip hatayı artırdıklarını göstermektedir. Buna karşılık en küçük I. Tip hata değeri Welch testinden elde edilmiştir.

Büyük örneklerle çalışıldığında (n=100) ortalamalar arasında fark yoktur şeklindeki hipotezi test etmek için kullanılan testlerin, I Tip hata yapma olasılıkları birbirine biraz daha benzer bulunmuştur. En küçük hata oranı Brown-Forsythe testinden elde edilmiştir (% 5.3). Bunun dışında diğer 3 testin hata oranları birbirine daha yakındır. Ancak bütün örnek genişliği kombinasyonlarında Welch testinin hata oranı genellikle % 5'in altında çıkarken, diğer testlerin hata oranı % 5'in üzerinde çıkmıştır. Ayrıca ölçümlerin logaritmasını aldıktan sonra bu testlerin uygulanması, I. Tip hatayı koruyucu yönde etki etmemektedir. Bir başka ifadeyle grup varyanslarının homojen olmadığı durumlarda grup ortalamalarını karşılaştırmak için logaritmik değerlerin kullanılması, I. Tip hatanın korunması yönünde iyi bir yaklaşım değildir.

Özet olarak, örnek genişlikleri eşit olduğunda ve bir grubun varyansı diğer grubun varyansının 4 katı ise ortalamalar arasında fark yoktur hipotezini incelerken I. Tip hata yapma olasılığını en iyi düzeyde koruyan test Welch testidir.

Grup sayısı 3 ve bu üç gruptaki varyansların birbirinden farklı olduğu durumlarda (1:25:100) farklı örnek büyüklükleri için ortalamalar arasında fark yoktur şeklinde kurulan hipotezin test edilmesi sonucunda elde edilen I. Tip hata olasılıkları Tablo 4'te topluca verilmiştir.

Tablo 4. $H_0: \mu_1 = \mu_2 = \mu_3$ Hipotezini Reddederken Yanılma Olasılığı

	Kombinasyon	TESTLER							
		Orijinal Ölçümler için Anova	Log. İçin Anova	Orijinal Ölçümler için Welch	Log. İçin Welch	Orijinal Ölçümler için Brown-Forsythe	Log. İçin Brown-Forsythe	Orijinal Ölçümler için K-W	Log. İçin K-W
$\sigma_1^2 : \sigma_2^2 : \sigma_3^2 = 1 : 25 : 100$	$n_i = 5$	7.5	7.9	6.4	6.0	5.2	4.8	7.2	7.2
	$n_i = 20$	6.2	9.3	4.4	6.0	5.6	8.1	7.6	7.6
	$n_i = 100$	6.2	42.1	4.3	43.8	6.3	42.3	5.8	5.8

Her bir gruptaki örnek genişliği 5 iken, orijinal ve logaritması alınmış ölçümlere Brown-Forsythe testi uygulandığı zaman, I. Tip hata yapma olasılığı başlangıçtaki düzeyini korumuştur. Ancak diğer test sonuçlarında bu olasılık başlangıçta belirlenene göre yüksek bulunmuştur (en küçük I. Tip hata % 6). Bu sonuçlar dikkate alındığında, küçük örnek genişlikleri için I. Tip hata yapma olasılığı bakımından Brown-Forsythe testinin diğer testlere göre daha avantajlı olduğunu söyleyebiliriz.

Örnek genişliği 20'ye çıkarıldığında, Welch ve Brown-Forsythe testleri uygulandığı zaman, I. Tip hata yapma olasılıkları diğer testlere nazaran başlangıçta belirlenen hata olasılığına daha yakın bulunmuştur (sırasıyla % 4.4 ve % 5.6). Anova ve Kruskal-Wallis testlerinin hata olasılıkları ise % 5'ten önemli düzeyde yüksek bulunmuştur.

Örnek genişliği 100'e çıkarıldığında ise, sadece orijinal ölçümlere Welch testi uygulandığında gerçekleşen I. Tip hata % 5'ten küçük çıkmış diğer koşullarda % 5'ten büyük çıkmıştır. Bu koşul için en iyi sonuç ise Welch ve Kruskal-Wallis testlerinden elde edilmiştir. Testlerin I. Tip hata yapma olasılıkları oldukça yüksek çıkmıştır. Diğer sonuçlar incelendiğinde ise hata olasılıklarının % 6'nın üzerinde olduğu görülmektedir. Ayrıca grupların varyansları birbirinden tamamen farklı olduğu durumda, grup ortalamalarını karşılaştırmak için logaritmik değerlerin kullanılması I. Tip hatanın korunması yönünde iyi bir yaklaşım değildir.

Özet olarak, örnek genişlikleri eşit olduğunda ve grupların varyansları farklı olduğunda, ortalamalar arasında fark yoktur hipotezini incelerken küçük örneklerde Brown-Forsythe testinin, orta ve büyük örneklerde ise Welch testinin I. Tip hata yapma olasılığı bakımından en uygun testler olduğu gözlenmektedir.

Tartışma ve Öneriler

Konuyla ilgili yapılan araştırmalarda verilere monoton bir transformasyon kullanmanın varyansların homojenliğini sağlamada etkili olmadığı görülmüştür (2, 6, 7). Bu çalışmada, logaritmik transformasyon uygulanan verilere 4 testte uygulanmış ancak testlerin I. Tip hata yapma olasılıklarında istenilen düzeye ulaşamamıştır.

Eğer varyanslar eşit değil ise dağılım normal olsun ya da olmasın Kruskal-Wallis testi karşılaştıracağı dağılımların benzer şekilli olduğunu varsayar (7).

Welch testi uygulandığında, bağımlı değişkene ait artık değerler normal dağılım göstermese bile posterior I. Tip hata (p) yapma olasılığı kontrol altında kalmakta ve aynı zamanda testin gücü de korunmaktadır. Ancak veriler karışık-normal dağılıma sahip ise Welch testinin gücü düşer. Ayrıca artıkların pozitif korelasyonlu olması Welch testi sonucunda, kabul edilemeyecek düzeyde büyük bir p değerini verirken negatif korelasyonlu olması kabul edilemeyecek düzeyde düşük bir p değerini verecektir. Böyle durumlarda farklı bir istatistik yöntem tercih edilmelidir (4, 5).

Welch testi varyanslar heterojen olduğu tek faktörlü denemelerde tercih edilir. Örnek genişliklerinin dengeli veya dengesiz olmasından etkilenmez. Verilerin dağılımı eğri olduğu durumda bu test kullanılmaz. Dengeli denemelerde örnek genişlikleri küçük ise bu testi kullanmaktan kaçınmak gerekir. Dengesiz denemelerde de bir grubun örnek genişliği diğerlerine göre oldukça küçük ise yine bu testi kullanmak uygun olmaz. (7)

Veriler simetrik olmayan bir dağılımdan çekiliyorsa ve küçük örneklerde James Second Order testinin I. Tip hatası genellikle kontrol edilememektedir. Bu testin, gruplar dengeli olsa bile veriler eğri bir dağılıma sahip olduğu durumda da kullanılmaması önerilmektedir. Grupların örnek genişlikleri 10'dan küçük olduğu zaman bu test Welch testine göre I. Tip hatayı daha iyi kontrol etmektedir. Ama oldukça karmaşık bir yöntem olmasından ve henüz istatistik paket programında yer almamasından dolayı pek tercih edilmemektedir (8, 13).

Eğer sadece bir grubun gözlemlerine ait varyans ekstrem bir değer alıyorsa Brown-Forsythe testini kullanmaktan kaçınmak gerekmektedir. James Second Order ve Welch testlerine göre varyansların heterojenliği konusunda onlar kadar dayanıklı bir test olmamasına karşılık James Second Order ve Welch testlerine göre eğri dağılım gösteren verilerden daha az etkilenir (8, 10).

Özet olarak, normal dağılım gösteren verilerde varyanslar heterojen ise I. Tip hata yapma olasılığı bakımından James Second Order testi hem Brown-Forsythe, hem de Welch testine göre daha iyi sonuç vermektedir.

Genel olarak parametrik testler verilerin normal dağılım göstermediği durumlar dışında parametrik olmayanlara göre daha iyi sonuçlar vermektedirler. Verilerin dağılımı normal olmadığı zaman Kruskal-Wallis testi istatistiksel gücü iyi olan bir testtir. Pratikte kullanımı zor değildir. Gruplar dengesiz olduğu zaman varyansların heterojenliğine duyarlı bir testtir (7).

Rank transform yöntemi varyanslar heterojen olduğu durumlarda iyi sonuçlar vermediği belirlenmiştir (11).

Sonuç olarak, varyansların homojenlik varsayımı sağlandığı zaman Anova testi bütün alternatif parametrik testlere göre en iyi sonucu veren testtir. Buna karşılık bu ön şartın yerine gelmediği durumlarda ise bütün alternatifler Anova testinden daha iyidir (14).

Bu çalışmada da, veriler normal dağılım gösteriyorsa ve varyans oranları çok farklı değil ise küçük örneklerde Welch ve Brown-Forsythe testlerinin I. Tip hata olasılıkları benzer bulunur iken varyans oranları farklılaştıkça Brown-Forsythe testinin daha iyi sonuç verdiği görülmüştür. Ancak orta ve büyük örneklerde ise Welch testinin diğer testlere oranla I. Tip hatası başlangıçta belirlenen seviyeye daha yakın bulunmuştur.

Kaynaklar

1. Shun-Yi Chen, Hubert J. Chen, Cherng G. Ding. An ANOVA test for the equivalency of means under unequal variances. Computational Statistics & Data Analysis 33, 2000; 151-175.
2. Gutti Jogesh Babu, A. R. Padmanabhan and Madan L. Puri. Robust One-Way ANOVA under Possibly Non-Regular Conditions Biometrical Journal 41 1999; 3, 321±339.
3. Brunner, E., Puri, M. L., and Sun, S. Nonparametric methods for stratified two-sample designs with applications to multi-clinical trials. J. Am. Statist. Assoc. 90. 1995; 1004±1014.
4. Welch, B. L. On the comparison of several mean values: an alternative approach. Biometrika 1951; 38: 330-336.
5. "Welch's One-Way Analysis of variance for Independent Samples (Welch's Anova)"; Erişim Adresi: <http://www.mtsu.edu/~dkfuller/notes302/Welchanova.pdf>. Erişim Adresi: 3.06.2005
6. Tekindal, Benian,1998. Varyans Analizinin Önşartları ve Transformasyonlar (Doktora Tezi). Ankara. Ankara Üniversitesi 1998.
7. "Alternatives to the Anova". Erişim Adresi : [biology.queensu.ca/~biol843/ANOVA-ranked%20 data.pdf](http://biology.queensu.ca/~biol843/ANOVA-ranked%20data.pdf). Erişim Tarihi: 06.06.2005

8. James'2 nd Order Approximation and Alexander's Normalized t-Transformation", Erişim Adresi : members.aol.com/IMSAP/James/James.html. Erişim Tarihi: 20.06.2005
9. Aguinis H., Petersan A. S., Pierce A. C. Appraisal of the Homogeneity of Error Variance Assumption and Alternatives to Multiple Regression for estimating Moderating Effects of Categorical Variables. Organization Research Methods, Vol.2, No.4, October 1999, 315-339.
10. Brown M.B, Forsythe A.B. The Anova and Multiple Comparisons for Data with Heterogeneous Variances. Biometrics 1974 ; 30, 719-724.
11. Erişim Adresi: www.engr.mun.ca/~llye/PP08-comparison2.ppt. Erişim Tarihi: 16.06.05 "Comparison Among Groups (more than 2 groups, 1 factor)".
12. Kanık EA, Mengi T. Simülasyon Yardımıyla Çok Değişkenli Veri Üretimi: MNRDG.v.2, 7.Ulusal Biyoistatistik Kongresi Özet Kitapçığı, sayfa:43, 29 Eylül-2 Ekim Mersin, 2004.
13. Wilcox, R.R. A new alternative to the ANOVA F and new results on James's Second-Order Method. British Journal of Mathematical and Statistical Psychology 1988; 41,109-117.
14. Coombs W.T, Algina J and Oltman D. O. Univariate and Multivariate Omnibus Hypothesis tests Selected to control Type I Error rates When population Variances are not Necessarily Equal. Review of educational Research 1996; 66 (2), 137-79.

NORMAL DAĞILIM GÖSTEREN POPULASYONLARDAN ALINAN ÖRNEKLERDE VARYANSLARIN HOMOJENLİĞİNİ TEST ETMEK İÇİN KULLANILAN BAZI YÖNTEMLERİN KARŞILAŞTIRILMASI

Sıddık Keskin¹, Ensar Başpınar¹, Handan Ankaralı², Fikret Gürbüz³

¹: Yüzüncü Yıl Üniversitesi, Tıp Fakültesi, Temel Tıp Bilimleri Bölümü,
Biyoistatistik Anabilim Dalı

²: Mersin Üniversitesi, Tıp Fakültesi, Temel Tıp Bilimleri, Bölümü
Biyoistatistik Anabilim Dalı

³: Ankara Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Biyometri ve
Genetik Anabilim Dalı

e-mail: skeskin973@hotmail.com

Özet

Bu çalışmada; normal dağılım gösteren popülasyonlardan alınmış örneklerde, varyansların homojenliğini test etmek için kullanılan yöntemlerden olan; Bartlett, Cochran, Hartley (Fmax), Levene, Brown-Forsythe ve Satterthwaite testleri I. Tip hata olasılığı (α) ve testin gücü ($1-\beta$) bakımından karşılaştırılmıştır. Bu amaçla; örnek genişliği her grupta eşit olacak şekilde, 3 ve 5 gruplu örnekler alınmış ve bu örneklerde 100.000 (yüz bin) simülasyon denemesi sonucunda, I. Tip hata olasılığı ve güç değerleri hesaplanmıştır. Güç değerlerinin hesaplanması için gruplardan birinin varyansı diğerlerinin 4 katı olarak alınmıştır.

Sonuçta; Bartlett, Cochran ve Hartley (Fmax), testlerinin I. Tip hata olasılığı ve güç bakımından diğer testlere göre daha tatminkar sonuçlar verdiği gözlenmiştir.

Anahtar Kelimeler: Bartlett testi, I. Tip hata, simülasyon, testin gücü, varyansların homojenliği.

Giriş

Varyans analizi tekniği; Normal dağıldığı varsayılan bir popülasyondan rasgele alınmış iki ve/veya daha fazla sayıdaki örnek ortalamalarını birbiriyle karşılaştırmada kullanılan yöntemlerden birisidir. Bu yöntem aslında;

üzerinde durulan özellik bakımından genel varyasyonu (Kareler Toplamını) bu özelliğe etkili olabileceği düşünülen faktör veya faktörlerin yer aldığı modele göre unsurlarına (bileşenlerine) ayırır (Keskin, 2002). Ayrıca, çeşitli biyolojik modellerdeki faktörlere ait varyans unsurlarının tahmininde de bu teknikten yararlanılmaktadır.

Ancak, varyans analizi tekniği; her ne kadar bir yöntem olarak kendinden bekleneni vermekte ise de, yöntemin dayandığı ön şartları (varsayımları) göz ardı etmek, pratikte yanlış veya yanıltıcı kararlara yol açabilmektedir (Keskin, 2002).

Varyans Analizi tekniğinin ön şartları (varsayımları) kısaca aşağıdaki gibi özetlenir.

- Üzerinde durulan özelliğin Normal dağılım göstermesi (normality),
- Gözlemlerin bağımsızlığı (independency),
- Etkilerin eklenebilir oluşu (additivity),
- Varyansların homojenliği (homogeneity).

Bu ön şartlardan birinin veya bir kaçının sağlanamaması durumunda; yapılacak olan hipotez testlerinde hem testin önemlilik seviyesi, hem de F istatistiğinin duyarlılığı etkilenebilmektedir (Steel vd. 1997). Gruplarda eşit sayıda gözlem olması durumunda hata terimlerinin Normal dağılım göstermeyişi, test sonucunu pek fazla etkilememektedir. Ancak aşırı çarpıklılık olduğunda, I. tip hata olasılığı değişebilmektedir (Sokal ve Rohlf 1995). Varyans analizinde test istatistiği için elde edilen sonuçların önemlilik seviyesi ve duyarlılığı, en fazla varyansların heterojen olması durumundan etkilenmektedir (Cochran ve Cox 1957). Bu nedenle, varyans analizinin ön şartlarından en önemli olanının “varyansların homojenliği” ön şartı olduğu söylenebilir (Bisop ve Dudewicz 1978). Varyansların homojenliği ön şartı; varyans analizi tekniğinin en önemli ön şartlarından birisi olunca, bu analiz tekniğine başvurmadan önce adı geçen ön şartın sağlanıp sağlanmadığının doğru olarak belirlenebilmesi büyük önem taşımaktadır. Bu nedenle; varyansların homojenliğini test etmeye yönelik, farklı koşullarda uygulanabilen ve bazıları birbirinin değişik hali olan 60’a yakın sayıda test yöntemi bulunmaktadır. Bu test yöntemlerinin uygulanması sonucunda elde edilecek sonuçlar; grup sayısından, gruptaki gözlem sayısından ve özelliğin göstermiş olduğu dağılımdan etkilenebilmektedir. Varyansların homojenliğini test etmede kullanılan yöntemlerin sonuçları her zaman birbirini desteklemekte, çoğunlukla çelişkili olabilmektedir.

Bu çalışmada; normal dağılım gösteren popülasyonlardan alınan 3 ve 5 grup içeren örneklerde, varyansların homojenliğini test etmede kullanılan yöntemlerden; Bartlett, Cochran, Hartley (Fmax), Levene, Brown-Forsythe ve Satterthwaite testleri I. tip hata olasılığı ve testin gücü bakımından simulasyon tekniği yardımıyla karşılaştırılmıştır.

Gereç ve Yöntem

Bu çalışmanın materyalini; FORTRAN Power Station derleyicisinin IMSL Library alt programlarından yararlanarak yazılan programlarla üretilen tesadüf sayıları oluşturmaktadır (Anonymous 1995). Bu tesadüf sayıları, $N(0,1)$ parametrelili dağılım gösteren popülasyondan üretilmiştir. Grup sayısı uygulamada sık karşılaşılan 3 ve 5 grup ile sınırlı tutulmuştur. Gruplardaki gözlem sayıları her grupta eşit olmak üzere; $n=4, 6, 8, \dots, 30$ olarak belirlenmiştir.

Ele alınan testler için ampirik olarak gerçekleşen I. tip hata olasılıkları, 100000 (yüz bin) simülasyon denemesi sonunda ret edilen H_0 hipotezi adetleri sayılarak, bu değer yüzdeye dönüştürülmesiyle elde edilmiştir.

Testler için ampirik güç değerlerinin elde edilmesinde; gruplardan birinin varyansı diğer grupların varyansının 4 katı olacak şekilde uygun katsayı ile çarpılarak varyanslar heterojen hale getirilmiş ve yine 100000 (yüz bin) simülasyon denemesi sonucunda ret edilen H_0 hipotezi adetleri sayılarak yüzdeye dönüştürülmüştür.

Test İstatistikleri

Her gruptaki gözlemler $i = 1, \dots, n_k$, grup sayısı $j = 1, \dots, k$ ve X_{ij} , j. muamele grubundaki i. gözlemin değerini göstermek üzere; her hangi bir j. grubun ortalaması ($\bar{X}_{.j}$) ve varyansı (S_j^2);

$$X_{.j} = \frac{\sum_{i=1}^{n_j} X_{ij}}{n_j} \quad S_j^2 = \frac{1}{n_j - 1} \cdot \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{.j})^2 \quad (1)$$

eşitlikleri ile ifade edilir. k adet grup için hesaplanan toplanmış varyans (pooled variance) ise;

$$S_p^2 = \frac{\sum_{j=1}^k (X_{ij} - \bar{X}_{.j})^2}{\sum_{j=1}^k v_j}, \quad (v_j = n_j - 1) \quad (2)$$

olarak yazılır. Burada v_j , j. grubun serbestlik derecesidir.

Bartlett (BA) Testi

Bartlett testi varyansların homojenliğini test etmede en yaygın olarak kullanılan testlerden birisidir. Gruplardaki gözlem sayısının eşit olması zorunluluğu yoktur. Bartlett testi için M istatistiği;

$$M = \log S_p^2 \left(\sum_{j=1}^k v_j \right) - \sum_{j=1}^k v_j \log S_j^2 \quad (4)$$

eşitliği ile hesaplanır.

M istatistiği yaklaşık olarak k-1 serbestlik dereceli χ^2 dağılımı göstermektedir. Ancak χ^2 dağılımına uyumu artırmak amacıyla, M istatistiğine bir düzeltme yapılır. C ile gösterilen düzeltme katsayısı;

$$C = 1 + \frac{1}{3(k-1)} \left[\sum_{j=1}^k \frac{1}{v_j} - \frac{1}{\sum_{j=1}^k v_j} \right] \quad (5)$$

eşitliği ile hesaplanır. Buradan düzeltilmiş Bartlett istatistiği; B = M / C olarak bulunur. B istatistiği, α yanılma düzeyindeki k-1 serbestlik dereceli χ^2 tablo değerinden büyükse ($B > \chi^2_{\alpha; k-1}$) ilgili H_0 hipotezi ret edilir. Bartlett testi Normal dağılımdan sapmaya karşı duyarlı bir testtir (Zar, 1999).

Cochran (CO) Testi

Cochran testi, k grulu bir denemede sapan varyansı bulmaya yönelik bir test olarak düşünülebilir. I. tip hatayı deneme sonunda koruyabilmektedir. Ancak, Normal dağılımdan sapmaya karşı duyarlı bir testtir. Cochran testinin kullanılabilmesi için gruplarda eşit sayıda gözlem bulunması gerekmektedir. Cochran testi için gerekli olan C istatistiği;

$$C = \frac{S_{\max}^2}{\sum_{j=1}^k S_j^2} \quad (6)$$

olarak hesaplanır. Bulunan bu değer k ve n-1'e göre düzenlenmiş Cochran tablo değerinden büyük olduğu durumda H_0 hipotezi ret edilir.

Hartley veya F_{\max} (HF) Testi

Hartley (1940, 1950) tarafından önerilen bu test yöntemi Hartley veya F_{\max} testi olarak bilinir (Zar, 1999). Hartley testinin kullanılabilmesi için, Cochran testinde olduğu gibi gruptaki gözlem sayılarının eşit olma zorunluluğu vardır. Hartley F_{\max} istatistiği;

$$F_{\max} = \frac{S_{\max}^2}{S_{\min}^2} \quad (7)$$

olarak hesaplanır. Buradan hesaplanan test istatistiği; α yanılma düzeyinde, k ve $n-1$ 'e göre düzenlenmiş Hartley tablo değeri ile karşılaştırılır ve hesaplanan değer, tablo değerinden büyük olduğu durumda H_0 hipotezi ret edilir.

Hartley testi, k gruplu bir denemede sapan (outlier) varyansı belirlemeye yönelik bir test olarak düşünülebilir. Test istatistiğinin hesaplanmasında, tüm grupların varyansı yerine sadece en büyük ve en küçük varyansların kullanılması Hartley testinde bir sakınca olarak görülebilir (Keskin, 2002).

Levene (LE) Testi

Levene testinde, her gruptaki gözlemlerin kendi grup ortalamasından olan mutlak değerce sapmaları alınarak, elde edilen bu değerlere varyans analizi yapılır. Y_{ij} , j . gruptaki i . gözlemin değerini göstermek üzere;

$$X_{ij} = |Y_{ij} - \bar{Y}_j| \quad (8)$$

olarak elde edilen X_{ij} değerlerine

$$L = \frac{\sum_{j=1}^k n_j (\bar{X}_j - \bar{X}_{..})^2 / (k-1)}{\sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2 / (N-k)}, \quad \left(N = \sum_{j=1}^k n_j \right) \quad (9)$$

eşitliğine göre varyans analizi yapılır ve elde edilen L istatistiği, $F_{\alpha; (k-1), (N-k)}$ tablo değerinden büyük olduğu durumda H_0 hipotezi ret edilir.

Brown-Forsythe (BF) Testi

Brown ve Forsythe (1974) tarafından önerilen bu test; her gruptaki gözlem değerlerinin kendi grup medyanından olan mutlak değerce

sapmasını alarak, bu değerlere varyans analizi yapılması esasına dayanır (Loh, 1987).

Y_{ij} ; j. gruptaki i. gözlem değerini ve \bar{Y}_j 'da j. grubun medyanını göstermek üzere, her gruptaki gözlem değerlerinin medyandan olan mutlak değerce sapması;

$$X_{ij} = |Y_{ij} - \bar{Y}_j| \quad (10)$$

olarak ifade edilir. Daha sonra X_{ij} değerlerine varyans analizi yapılması sonucu BF istatistiği;

$$BF = \frac{\sum_{j=1}^k n_j (\bar{X}_j - \bar{X}_{..})^2 / (k-1)}{\sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2 / (N-k)}, \quad \left(N = \sum_{j=1}^k n_j \right) \quad (11)$$

olarak hesaplanır. Eğer hesaplanan BF istatistiği, $F_{\alpha; (k-1), (N-k)}$ tablo değerinden büyükse ilgili H_0 hipotezi ret edilir.

Satterhwaite (SA) Testi

Satterwaite testinde hesaplanan test istatistiği, Brown-Forsythe (BF) istatistiğidir. Ancak, bu istatistik F tablo değeri ile karşılaştırılırken serbestlik derecesinde bir düzeltme yapılır. Test istatistiği, k-1 ve N-k serbestlik dereceli F tablo değeri ile değil de, k-1 ve v serbestlik dereceli F tablo değeri ile karşılaştırılır. v değeri;

$$v = \frac{\left(\sum_{j=1}^k X_j \right)^2}{\sum X_j^2 / v_j}, \quad (X_j) = \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2 \quad \text{ve} \quad v_j = n_j - 1 \text{ dir} \quad (12)$$

olarak hesaplanır (Loh, 1987).

Bulgular ve Tartışma

Normal dağılım gösteren popülasyondan alınan örneklerde; testlere ait ampirik olarak gerçekleşen I. Tip hata ve güç değerleri, grup sayısı 3 ve 5 olduğu durumlar için sırasıyla; Çizelge 1. ve Çizelge 2 de verilmiştir.

Çizelge 1. Üç grup için ampirik olarak gerçekleşen I. Tip Hata (α) ve Güç ($1-\beta$) değerleri

Gözlem Sayısı		1-b için varyans oranı: 1:1: 4					
		BA	CO	HF	LE	BF	SA
4: 4: 4	α	0.04878	0.05090	0.05164	0.09550	0.06794	0.02154
	$1-\beta$	0.17445	0.22755	0.14330	0.19684	0.14688	0.14584
6: 6: 6	α	0.04904	0.04994	0.04937	0.07472	0.03854	0.04057
	$1-\beta$	0.31014	0.37372	0.25836	0.29771	0.17539	0.21880
8: 8: 8	α	0.05028	0.05090	0.05022	0.06675	0.03360	0.04094
	$1-\beta$	0.44174	0.51171	0.38493	0.39401	0.26285	0.31257
10: 10: 10	α	0.04976	0.05065	0.04997	0.06392	0.03306	0.04080
	$1-\beta$	0.55087	0.61875	0.49565	0.48561	0.36279	0.40964
12: 12: 12	α	0.04889	0.04952	0.04901	0.05951	0.03264	0.04023
	$1-\beta$	0.64776	0.70757	0.59665	0.57342	0.46429	0.50722
14: 14: 14	α	0.04949	0.04922	0.04960	0.05860	0.03468	0.04153
	$1-\beta$	0.72670	0.77837	0.68154	0.64957	0.55795	0.59338
16: 16: 16	α	0.05051	0.05116	0.04957	0.05773	0.03511	0.04146
	$1-\beta$	0.79110	0.83505	0.75097	0.71804	0.64270	0.67118
18: 18: 18	α	0.04881	0.04979	0.04903	0.05684	0.03613	0.04183
	$1-\beta$	0.84143	0.87876	0.81033	0.77527	0.71548	0.73901
20: 20: 20	α	0.04950	0.04938	0.04922	0.05609	0.03678	0.04225
	$1-\beta$	0.88151	0.91153	0.85426	0.82349	0.77483	0.79373
22: 22: 22	α	0.05010	0.04893	0.05064	0.05578	0.03858	0.04396
	$1-\beta$	0.90869	0.93442	0.88820	0.85918	0.82087	0.83526
24: 24: 24	α	0.05063	0.05006	0.05125	0.05525	0.03970	0.04451
	$1-\beta$	0.93333	0.95244	0.91595	0.88896	0.85952	0.87044
26: 26: 26	α	0.05240	0.05187	0.05114	0.05596	0.04116	0.04556
	$1-\beta$	0.95026	0.96608	0.93605	0.91306	0.88927	0.89817
28: 28: 28	α	0.05004	0.04969	0.05025	0.05465	0.04016	0.04438
	$1-\beta$	0.96463	0.97562	0.95372	0.93455	0.91695	0.92334
30: 30: 30	α	0.05114	0.05093	0.04964	0.05446	0.04154	0.04553
	$1-\beta$	0.97251	0.98203	0.96386	0.94830	0.93529	0.93994

Çizelge 1 ve Çizelge 2, gerçekleşen I. tip hatalar bakımından incelendiğinde; Bartlett, Cochran ve Hartley testlerinin, başlangıçta karşılaştırılan I. tip hatayı, büyük ölçüde deneme sonunda da koruyabildikleri görülmektedir. Levene testi için gerçekleşen I. tip hata değerinin, küçük örneklerde % 5'ten belirgin bir şekilde büyük olduğu, örnek genişliğinin 30 olması durumunda bile ancak % 5.446'ya düştüğü görülmektedir. Brown-

Forsythe ve Satterhwaite testlerinin grup sayısından bağımsız olarak, başlangıçta karşılaştırılan % 5'lik I. Tip hatayı genelde % 5'in altında tuttukları dikkat çekmektedir. Dolayısıyla bu iki testin diğer testlere göre daha tutucu (conservative) oldukları söylenebilir.

Çizelge 2. Beş grup için ampirik olarak gerçekleşen I. Tip Hata (α) ve Güç ($1-\beta$) değerleri

Gözlem Sayısı		1-b için varyans oranı: 1: 1: 1: 1: 4					
		BA	CO	HF	LE	BF	SA
4: 4: 4: 4: 4	α	0.04835	0.05071	0.05018	0.11225	0.06318	0.03505
	$1-\beta$	0.18872	0.27824	0.12543	0.25520	0.15960	0.16182
6: 6: 6: 6: 6	α	0.04814	0.04938	0.04959	0.08497	0.03468	0.03901
	$1-\beta$	0.32860	0.43522	0.24001	0.34467	0.19505	0.23033
8: 8: 8: 8: 8	α	0.04990	0.05004	0.05026	0.07667	0.03095	0.03793
	$1-\beta$	0.45527	0.566659	0.35787	0.43894	0.28658	0.32708
10: 10: 10: 10: 10	α	0.04950	0.05047	0.04996	0.06831	0.02860	0.03576
	$1-\beta$	0.56596	0.67349	0.58511	0.53274	0.50022	0.53353
12: 12: 12: 12: 12	α	0.05025	0.04996	0.05010	0.06555	0.03039	0.03682
	$1-\beta$	0.66605	0.76120	0.58511	0.62021	0.50022	0.53353
14: 14: 14: 14: 14	α	0.04910	0.05057	0.04994	0.06240	0.03102	0.03679
	$1-\beta$	0.74320	0.82533	0.67538	0.69491	0.59712	0.62333
16: 16: 16: 16: 16	α	0.04934	0.04990	0.04950	0.05951	0.03201	0.03929
	$1-\beta$	0.80283	0.87337	0.74476	0.75365	0.67342	0.69664
18: 18: 18: 18: 18	α	0.04843	0.04924	0.04828	0.05951	0.03366	0.03929
	$1-\beta$	0.85206	0.90866	0.80631	0.80549	0.74331	0.76088
20: 20: 20: 20: 20	α	0.05070	0.05013	0.04993	0.05834	0.03437	0.04013
	$1-\beta$	0.88989	0.93639	0.85205	0.84844	0.79849	0.81287
22: 22: 22: 22: 22	α	0.04979	0.04947	0.04979	0.05702	0.03524	0.04008
	$1-\beta$	0.91929	0.95489	0.89114	0.88416	0.84644	0.85683
24: 24: 24: 24: 24	α	0.04952	0.04984	0.04944	0.05711	0.03605	0.04040
	$1-\beta$	0.94087	0.96862	0.91810	0.91055	0.88115	0.88936
26: 26: 26: 26: 26	α	0.05022	0.04946	0.05012	0.05687	0.03796	0.04210
	$1-\beta$	0.95736	0.97823	0.94111	0.93123	0.90913	0.91536
28: 28: 28: 28: 28	α	0.05035	0.04940	0.05107	0.05689	0.03796	0.04210
	$1-\beta$	0.96931	0.98544	0.95646	0.94832	0.93191	0.93643
30: 30: 30: 30: 30	α	0.04816	0.04976	0.04837	0.05520	0.03784	0.04174
	$1-\beta$	0.97796	0.98976	0.96786	0.96113	0.94916	0.95230

Gartside (1972)'ın yapmış olduğu çalışmada, 3 grup için grupta eşit sayıda (n = 16) gözlem olduğu durumda; Bartlett, Cochran ve Hartley testleri için gerçekleşen I. tip hatalar sırasıyla; % 4.7, % 5.5 ve % 5.1 olarak bulunmuşken, çalışmada bu değerler sırasıyla; % 5.05, % 5.12 ve % 4.96 olarak gerçekleşmiştir. Overall ve Woodward (1974), Bartlett testi için gerçekleşen I. tip hatayı (1000 deneme), n = 12 olduğu durumda, % 5.1 olarak belirtirken, Geng vd (1979) n = 6 olduğunda (1000 deneme) % 5.1, n = 10 olduğunda ise % 4.6 olarak belirtmişlerdir.

Çalışmada bu değer, n = 12 olduğunda; % 4.89, n = 6 olduğunda; % 4.90 ve n =10 olduğunda % 4.98 olarak gerçekleşmiştir. Levene testi, başlangıçta karşılaştırılan I. tip hatayı deneme sonunda koruyamamaktadır. Bu sonuç, Keskin ve Gürbüz (2002) ve Keskin (2002) tarafından bildirilen sonuçlara benzerlik göstermektedir.

Keskin (2002) 3 grulu denemelerde; Bartlett, Cochran, Hartley, Levene, Brown-Forsythe ve Satterhwaite testlerine ait I. Tip hata değerlerini, gruptaki gözlem sayısı 4 olduğu durumda sırası ile; % 4.92, % 5.13, % 5.14, % 9.35, % 6.61 ve % 2.08 olarak bildirirken, gruptaki gözlem sayısının 30 olması durumunda sırası ile; % 4.94, % 4.91, % 4.90, % 5.32, % 4.03 ve % 4.43 olarak bildirmiştir. Çalışmada ise bu değerler; gruptaki gözlem sayısı 4 olduğunda sırası ile; % 4.88, % 5.09, % 5.16, % 9.55, % 6.79 ve % 2.15 olarak gerçekleşirken, gruptaki gözlem sayısının 30 olması durumunda sırası ile; % 5.11, % 5.09, % 4.96, %5.45, % 4.15 ve % 4.55 olarak gerçekleşmiştir. Çalışmadan, elde edilen I. tip hata değerlerinde olduğu gibi güç değerleri de Keskin (2002)'in bildirdiği değerlere büyük ölçüde benzerlik göstermektedir. Bartlett, Cochran ve Hartley testlerinin tutarlı sonuçlar verdiği söylenebilir. Conover vd (1981), yapmış oldukları çalışmada, ele aldıkları testlerde gerçekleşen I. tip hata % 10'a kadar olduğunda bu testleri "robust" (sağlam) testler olarak kabul etmişler; ancak Sharma (1991) yapmış olduğu çalışmada bu sınırı % 7 olarak almıştır. Çizelge 1 ve Çizelge 2, ampirik olarak gerçekleşen güç değerleri bakımından incelendiğinde; grupta 4'er adet gözlem olduğu durumda, en yüksek güç değerinin % 27.824 değeri ile Cochran testinde gerçekleştiği görülür. Cochran testi, bu yüksek güç değerini bütün örnek genişliklerinde koruyabilmekte ve örnek genişliği 30 olduğu durumda, bu testin gücü % 98.976' ya çıkmaktadır (Çizelge 2). Grupta 4'er adet gözlem olduğu ve grup sayısının da 3 olduğu durumda Levene testinin gücü yaklaşık % 19.68 olurken, bu değer örnek genişliği 30 olduğu durumda yaklaşık % 95 olmaktadır.

Grup sayısı 3 iken, Bartlett testi % 80 güç değerine (I. Tip hata yaklaşık % 5 iken) örnek genişliği 18 olduğu durumda ulaşırken, Cochran testi 16 olduğu durumda ulaşmakta, grup sayısının 5 olduğu durumda ise Bartlett testi n =16 iken, Cochran testi de n = 14 iken % 80 güç değerine ulaşmaktadır.

Sonuç

Bartlett ve Cochran testleri, Normal dağılımından alınan örneklerde; grup sayısından ve gruptaki gözlem sayısından bağımsız olarak, başlangıçta karşılaştırılan I. tip hatayı koruyabilmekte olup, Cochran testine ait güç değeri, genel olarak diğer testlerden daha yüksektir.

Bartlett ve Cochran testleri için güç değeri ve I. tip hata olasılıkları birlikte dikkate alındığında; I. tip hata olasılığının % 5 ve güç değerinin % 80 olması için örnek genişliğinin yaklaşık olarak 15 olması gerekmektedir.

Gerçekleşen I. tip hata bakımından, Hartley testinden elde edilen sonuçların, Cochran testinden elde edilen sonuçlara genelde benzer olmasına rağmen, güç değerleri bakımından elde edilen sonuçların, Cochran testine ait güç değerlerinden daha düşük olduğu görülmektedir. Ancak, yine de, Cochran testinin kullanılmasının uygun olduğu koşullarda, Hartley testinin de kullanılabilirliği söylenebilir.

Levene testi, başlangıçta karşılaştırılan I. tip hatayı koruyamamaktadır. Grup sayısının artmasıyla birlikte, teste ait gerçekleşen I. tip hata değeri de artmakta, yani bu test, grup sayısının artışından etkilenmektedir.

Brown- Forsythe ve Satterhwaite testleri her ne kadar I. Tip hatayı başlangıçta karşılaştırılan değer (% 5'in) altında tutabilmişler de genel olarak Bartlett, Cochran ve Hartley testleri kadar iyi sonuçlar vermemektedir.

Sonuç olarak, Normal dağılım gösteren popülasyonlardan eşit hacimli olmak şartıyla alınan örneklerde, Bartlett, Cochran ve Hartley testlerinin diğer testlere göre daha iyi sonuçlar verdiği ve kullanılmasının uygun olabileceği söylenebilir.

Kaynaklar

1. Keskin, S. Varyansların Homojenliğini Test Etmede Kullanılan Bazı Yöntemlerin I. Tip Hata ve Testin Gücü Bakımından İrdelenmesi (Doktora tezi), Ankara: Ankara Üniversitesi; 2002.
2. Steel, R. G. D. Torrie J. H. and Dickey D. A., 1997. Principles and Procedures of Statistics, 3rd. ed. McGraw Hill, New York, 1997.
3. Sokal, R. R. and Rohlf, F. J., Biometry. Thirt Edition, W. H. Freeman and Company. New York, USA. 887p. 1995.
4. Cochran, W.G. and Cox, G. M., Experimental Design. John Willey and Sons Inc. New York, USA. 455p. 1957.
5. Bishop, T. A. and Dudewicz. E. J. Exact analysis of variance with unequal variances. Tests procedures and tables. Technometrics, 1978, 20(4); 419-424.

6. Anonymous, Microsoft Developer Studio, Microsoft Corporation, USA.,1995
7. Zar, J. H., Biostatistical Analysis. Fourt Edition, Prentice Hall Inc. Simon and Schuster/A Viacom Company. New Jarsey USA. 663p. 1999.
8. Loh, W. Y., Some modification of Levene's test of variance homogeneity. J. Statist. Comput. Simul., 1987, 28; 213-226.
9. Gartside, P. S., A study of methods for comparing several variances. JASA, 1972, 67(338); 342-346.
10. Overall, J. E. and Woodward, J. A., A simple test for heterogeneity of variance in complex factorial designs. Psychometrika. 1974, 39(3); 311-318.
11. Geng, S., Wang, W. J. and Miller, C., Small sample size comparisons of tests for homogeneity of variances by monte-carlo. Commun. Statist.-Simula. Computa., 1979.B8(4); 379-389.
12. Keskin, S. ve Gürbüz, F., Küçük örneklerde Bartlett ve Levene test istatistiklerinin I. tip hata bakımından incelenmesi. Tarım Bilimleri Dergisi. 2002, 8(1); 51-58.
13. Conover, W. J., Johnson, M. E. and Johnson, M. M., A comparative study of tests for homogeneity of variance, with application to the outer continental shelf bidding data. Technometrics, 1981, 23(4); 351-361.
14. Sharma, S. C., A new Jackknife test for homogeneity of variances Commun. Statist.-Simula, 1991, 20(2&3); 479-495.

POSTER BİLDİRİLER

TIP EĞİTİMİNDE BİYOİSTATİSTİK

İlker Ercan, Güven Özkaya, Gökhan Ocakođlu, Bülent Ediz, Berna Yazıcı, İsmet Kan

*Uludađ Üniversitesi, Tıp Fakóltesi, Biyoistatistik A.D.
Anadolu Üniversitesi Fen Fakóltesi, İstatistik Bölümü*

Çalışma tıp eğitiminde, biyoistatistik eğitiminin hangi aşamalarda verilmesi gerektiğinin, farklı eğitim düzeyindekilerin biyoistatistik hakkındaki görüşlerinin ve hangi istatistik testler hakkında bilgi sahibi olduklarının incelenmesi amacıyla yapılmıştır.

Araştırmamızda I. sınıf öğrencilerden profesörlük ünvanına kadar 845 denek yer almıştır. Araştırmaya katılanların % 24,7'si sadece uzmanlığın ilk yıllarında, % 15,2'si sadece VI. sınıfta, % 10,2'si ise sadece I. sınıfta dersin verilmesinin uygun olduğunu düşünmektedirler. Dersin I. sınıf ve I. sınıfla birlikte diğer eğitim seviyelerinde de verilmesi gerektiğini düşünenler ise %19,9'dur. Araştırmamızda biyoistatistiğın işlevi hakkında büyük ölçüde bilgisi olduğunu belirtenlerin % 49,7'si sağlık alanında yapılan bir araştırmaya katılmıştır. Bu oran hiç ve kısmen bilgisi olduğunu ifade edenlerin çalışmada bulunma oranlarından anlamlı düzeyde yüksektir ($p<0.001$). Dersi almanın hekimlere mesleğinde yarar sağlayacağı ve biyoistatistiğın tıp biliminde önemli olduğu düşünceleri bakımından hekim ve öğrenci görüşlerinde anlamlı farklılıklar vardır ($p<0.001$). Biyoistatistiğın işlevi hakkında bilgi düzeyi arttıkça biyoistatistiğın tıp biliminde önemli olduğu ve dersi almanın hekimlik mesleğinde yarar sağlayacağı düşüncelerinde artış görülmektedir ($p<0.001$). Biyoistatistiğe verilen önem arttıkça hakkında bilgi sahibi olunan test sayısı da artmaktadır ($p<0.001$). İstatistiksel testler derste anlatılanlar ve anlatılmayanlar, parametrik ve parametrik olmayan, tek değişkenli ve çok değişkenli, örnekleme yöntemleri ve yaşam analizi yöntemleri olarak gruplandırılarak farklı eğitim seviyelerine göre bilinme oranları incelendiğinde anlamlı farklılıklar bulundu.

Sonuçta, öğrencilik yıllarında biyoistatistiğın önemi yeterince anlamamaktadır ve eğitim sürecinin ilerleyen yıllarında biyoistatistiğın işlevi hakkındaki bilgi düzeyi arttıkça verdikleri önem de artmaktadır. Testler hakkında bilgi farklı eğitim seviyelerine göre farklılıklar göstermektedir. Ayrıca biyoistatistik eğitiminin ilk yılda verilmemesi ve uzmanlık sürecinin ilk yıllarında verilmesi gerektiği çoğunluk tarafından düşünülmektedir.

DEĞİŞKEN YAPILARINA GÖRE İSTATİSTİKSEL YAKLAŞIM

Zeki Akkuş, M. Yusuf Çelik

Dicle Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı

e-mail:zakkus@dicle.edu.tr

Özet

Değişken yapıları istatistiksel çalışmalarda birinci adımda dikkate alınması gereken önemli bir durumdur. Araştırmacılar adimsal çözümlmeyi dikkate almadan yaptıkları çözümlmelerde bir çok yanlışlıklarla dolu yayınlar yapmaktadırlar. Değişken yapıları seçilecek olan istatistiksel yöntemi belirler. Konuyu incelikleri ile bilmeyen araştırmacıların değişken yapılarına dikkat etmeleri olası değildir.

Bu çalışmada istatistikte en güçlü varsayım kabul edilen değişken yapılarıyla istatistik yöntemler arasındaki güçlü ilişkiler incelenecektir. Araştırmacılar çalışmalarda bazen tek değişken bazen de çok değişken yapıyla ilgilenirler. Günümüzde çok değişkenli etkenlerin daha çok gözlemlendiği bir gerçektir. Tek değişkenli veya çok değişkenli Biyoistatistik yöntemlerin tümü değişkenlerle ilgili bazı varsayımlar öne sürerler.

Parametrik testlerde ana varsayım normal dağılımdır. Normal dağılımın en önemli özelliği sürekli değişkenlerin bir dağılışı olmasıdır. Tek değişken (univariate) veya çok değişken (multivariate) parametrik yöntemlerin ön gördüğü değişken yapısı sürekli (continuous) değişkenlerdir. Sürekli değişkenlerin normal dağılım gösterdiği bilinmektedir. Sürekli olmayan değişkenler normal dağılım göstermedikleri için "dağılıştan bağımsız" yöntemlerin seçilmesi gerekir.

Parametrik olmayan (non-parametric) yöntemlerin en önemli özelliği araştırmacıya değişken konusunda bir sınırlama getirmemesidir. Araştırmacılar için bu özellik istatistiksel çözümlemede bir kolaylık sağlar. Bu avantajları doğru kullandıkça yöntemlerin sonuçları daha güçlü ve daha duyarlı olacağı bir gerçektir.

Çözümlemeye başlamadan önce değişkenlerin yapıları açık bir şekilde tanımlanması yöntem seçiminde doğru bir sonuca gitmemize olanak tanır.

YAPAY SINIR AĞLARI VE İSTATİSTİK YÖNTEMLERLE İLİŞKİLERİ

Ayşe Canan Yazıcı¹, Ersin Ögüş¹, Handan Çamdeviren², Sinan Canan³

Başkent Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı¹

Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı²

Başkent Üniversitesi Tıp Fakültesi, Fizyoloji Anabilim Dalı³

e-mail: acyazici@baskent.edu.tr

Özet

Yapay sinir ağları (YSA); insan beyninin çalışma mekanizmasını taklit ederek beynin öğrenme, hatırlama, genelleme yapma yolu ile yeni bilgiler türetebilme gibi temel işlevlerini gerçekleştirmek üzere geliştirilen mantıksal yazılımlardır. Çeşitli avantajlar sunan ve sürekli gelişmekte olan bu teknoloji ile istatistik yöntemler arasında önemli ilişkiler söz konusudur. Bu çalışmada YSA'nın genel yapısı ele alınarak istatistik yöntemlerle ilişkileri araştırılmıştır. Pek çok yapay sinir ağı modelinin bazı klasik istatistik yöntemlerin genelleştirilmiş hali olduğu görülmüştür. İstatistikte doğrusal olmayan modellerde kullanılan yöntemler ile ileri beslemeli YSA algoritmaları eğitilerek sonuç üretilebilmektedir. İstatistikte yapay sinir ağları genellikle esnek olarak doğrusal olmayan regresyon analizi veya sınıflandırma modelleri olarak kullanılmaktadırlar. Pek çok YSA algoritmasının parametrik olmayan istatistiksel yöntemlerle yakın ilişkisi vardır.

Anahtar kelimeler: Yapay sinir ağları, doğrusal olmayan modeller, optimizasyon modelleri.

TÜRKİYE'DEKİ ÖZÜRLÜ GRUPLARIN YAPISININ UYUM ANALİZİ İLE İNCELENMESİ

Şengül Cangür, Deniz Sığırlı, Bülent Ediz, İlker Ercan, İsmet Kan

Uludağ Üniversitesi, Tıp Fakültesi, Biyoistatistik A.D.

e-mail: cangur@uludag.edu.tr

Özet

Uyum Analizi (Correspondence Analysis, CA); bir veri matrisini, satır ve sütunları noktalarla gösterilen özel bir grafiksel gösterim tipine dönüştüren açıklayıcı çok değişkenli bir analiz tekniğidir. Uyum analizi, satır profillerindeki homojenlik eksikliği ya da kontenjans tablosunun satır ve sütunları arasındaki bağımlılığı (ya da etkileşimi) daha düşük boyutta açıklama imkanı sağlamaktadır. Ülkemizde nüfusun % 12.29' unu özürlü bireyler oluşturmaktadır. Özürlü bireylerin fiziksel ve psiko-sosyal yapılarının toplum içinde sorunsuz bir şekilde gelişmesine uygun ortam yaratmak, toplumun üzerinde durması gereken en önemli sosyal konulardan birisidir. Ergonomik boyutlarının araştırılmasına yönelik bundan sonra yapılacak projelere ışık tutmak amacıyla, 2002 yılında Devlet İstatistik Enstitüsü ile işbirliği içerisinde gerçekleştirilen Türkiye Özürlüler Araştırması'ndan elde edilen verilere sözü edilen analiz tekniği uygulanarak, özürlü kişilerin yaşamlarını daha çağdaş ve daha insani ortamlarda sürdürebilmeleri amacıyla çeşitli özürlü gruplarının yapıları incelenmiştir.

Anahtar kelimeler: Uyum Analizi; Çoklu Uyum Analizi; Homojenlik Analizi; Özürlü Bireyler.

Kaynaklar

1. **Die, (2004).** Türkiye İstatistik Yıllığı, Ankara: T.C. Başbakanlık Devlet İstatistik Enstitüsü.
2. **Eye Von, A., (1990).** Statistical Methods in Longitudinal Research Volume II, London: Academic Pres.
3. **Gifi, A., (1990).** Nonlinear Multivariate Analysis, New York: John Wiley&Sons.

4. **Greenacre, M.&Hastie, T., (1987)**. The Geometric Interpretation of Correspondence Analysis. *Journal of The American Statistical Association*, 82, 398, 437-447.
5. **Higgs, N.T., (1990)**. Practical and Innovative Uses of Correspondence Analysis. *The Statistician*, 40, 2, 183-194.
6. **Hill, M.O., (1974)**. Correspondence Analysis: A Neglected Multivariate Method. *Applied Statistics*, 23, 3, 340-354.
7. **Jambu, M., (1991)**. *Exploratory and Multivariate Data Analysis*, San Diego: Academic Press, Inc.
8. **Özdamar, K., (2004)**. *Paket Programlar ile İstatistiksel Veri Analizi*, Eskişehir: Kaan Kitabevi.
9. **Palmer, M.W., (1993)**. Putting Things in Even Better Order: The Advantages of Canonical Correspondence Analysis. *Ecology*, 74, 8, 2215-2230.
10. **Panagiotakos, D.B.&Pitsavos, C., (2004)**. Interpretation of Epidemiological Data Using Multiple Correspondence Analysis and Log-linear Models. *Journal of Data Science*, 2, 75-86.

TÜRKİYE VE AVRUPA BİRLİĞİ'NE ÜYE ÜLKELERİN SAĞLIK DÜZEYİ ÖLÇÜTLERİNİN ÇOK BOYUTLU ÖLÇEKLEME ANALİZİ İLE İNCELENMESİ

Deniz Sığırılı, Şengül Cangür, Bülent Ediz, İlker Ercan, İsmet Kan

Uludağ Üniversitesi, Tıp Fakültesi, Biyoistatistik A.D.

e-mail: sigirli@uludag.edu.tr

Özet

Çok Boyutlu Ölçekleme (Multidimensional Scaling-MDS) Analizi, nesne ya da birimler arasında gözlenen benzerlikler ya da farklılıklardan oluşan uzaklık değerlerine dayalı olarak bu nesnelerin tek ya da çok boyutlu uzaydaki gösterimini elde etmeyi amaçlayan, böylece nesnelere arasındaki ilişkilerin belirlenmesini sağlayan bir yöntemdir. p boyutlu bir uzayda n nokta arasındaki benzerlikleri kullanarak bu noktaların birbirlerine yakınlıklarını uzaydaki konumları ile birlikte değerlendirmek ve grafiksel olarak sunumunu sağlamak amacıyla bu yöntemden yararlanılmaktadır. Ülkelerin ulaşmak istedikleri kalkınmışlık düzeyi, günümüzde yeni bir yaklaşımla ele alınmakta ve özellikle sağlık konusu bu yaklaşımda önemli bir yer tutmaktadır. Sağlık sektörünün kalkınma üzerindeki rolünü ön plana çıkaran bu yeni yaklaşım, sektörün önemini daha da artırmakta ve ülkelerin gelişmişlik düzeyinin belirlenmesinde çeşitli sağlık göstergelerinin de kullanılmasına yol açmaktadır. Toplumsal gelişmenin yansımaları olarak kabul edilen bu sağlık göstergeleri bir yandan mevcut durumu yansıtırken öte yandan da ülkelerin önceliklerine ilişkin yol gösterici olmaktadır. Sağlık alanındaki talebin saptanması ve karşılanması, toplumun sağlık sorunlarının belirlenmesi ve bunların çözümü, sağlık hizmetlerinin planlanması ve oluşturulacak sağlık politikalarına yol göstermesi amacıyla; ülkemizin diğer Avrupa Birliği üye ülkeleriyle sağlık göstergeleri bakımından karşılaştırılması ve nasıl bir konumda olduğunun belirlenmesi gerekmektedir. Bu doğrultuda, Dünya Sağlık Örgütü (World Health Organization-WHO) tarafından seçilen sağlık düzeyi ölçütleri dikkate alınarak Avrupa Birliği'ne üye ve üyeliğine başvuruda bulunmuş bazı ülkelere ilişkin verilere çok boyutlu ölçekleme yöntemi uygulanarak, ülkelerin ilgili sağlık göstergeleri bakımından birbirlerine göre konumları incelenmiş ve aralarında var olabilecek benzerlikler/farklılıklar ortaya konulmuştur.

Anahtar kelimeler: Çok Boyutlu Ölçekleme, Sağlık Göstergeleri, Öklid Uzayı.

**KLİNİK ARAŞTIRMALARDA, “ÜSTÜNLÜK”,
“EŞDÜZEYLİLİK” VE “DAHA DÜŞÜK OLMAMAK”
DEĞERLENDİRMELERİ İLKELERİ**

F. N. Turan¹, M. Şenocak²

T.Ü. Tıp Fak. Biyoistatistik ABD.¹

İ.Ü. Cerrahpaşa Tıp Fak. Biyoistatistik BD.²

Özet

Klinik araştırmalar genellikle yeni bir ilacın, tedavinin o ana dek kullanılan standart yaklaşımdan daha üstün (superiority) olduğunu kanıtlamak için yapılırlar. Ancak bazı durumlarda da bunun tam tersi olan yeni bir ilacın esasen standart yaklaşıma eşdüzey olduğu (equivalent), ya da bu standart yaklaşımdan daha düşük etkinlikte (non-inferiority) olmadığı kanıtlanmaya çalışılır.

Klasik sıfır varsayımı yaklaşımı ve buna dayalı “p” değerlendirmesi özellikle eşdüzey etkinlik ve daha düşük etkinlikte olmamak durumlarını kanıtlamada yetersiz kaldığından, yargılamaların güvenilir yapılabilmesi için tasarlanan kavram ve yaklaşımlardan birisi de güven aralıkları yaklaşımıdır. Eğer sıfır hipotezi doğru ise oluşan gözlenen olasılık ya da çıktılarıdaki fark yani “p” değeri, klinik deneme sonuçlarını belirtmede sıklıkla kullanılan bir istatistiktir. Ancak “p” değeri bazı durumlarda yetersiz kaldığı için yorumlamada sorunlar yaşanmaktadır. Bu yüzden güven aralığı yöntemi üstünlük, özellikle de eşdüzey etkinlik ve daha düşük etkinlikte olmamak yargılamalarında “p” değerinden daha iyi bilgi verir.

Klinik araştırmalarda yeni bir ilacın ya da tedavinin o ana dek kullanılan standart yaklaşımdan daha üstün olduğunu, en azından eşdüzey etkinlikte olduğunu ya da daha düşük etkinlikte olmadığını yargılamanın en iyi yöntemi klasik sıfır varsayımı yaklaşımı olan “p” değeri yönteminden ziyade güven aralıkları yöntemidir.

PATH ANALİZİNİN, KORELASYON ANALİZİ VE ÇOK DEĞİŞKENLİ REGRESYON ANALİZİ İLE KARŞILAŞTIRILMASI

Rojan Arcak¹, Hasan Akbayın², Tuba Biçer³

Dicle Üniversitesi Ataürk Sağlık Hizmetleri M.Y.O¹

Dicle Üniversitesi Eğitim Fakültesi Ortaöğretim Fen ve Matematik Alanlar Bölümü²

Dicle Üniversitesi Ziraat Fakültesi Tarla Bitkileri Bölümü³

Bu çalışmanın amacı Path Analizi, Çok Değişkenli Regresyon Analizi ve Korelasyon Analizi yöntemlerini karşılaştırmak ve mercimekte tane verimini etkileyen bazı morfolojik değişkenleri kullanarak açıklamaya çalışmaktır.

Bilindiği gibi Path Analizi değişkenler arasındaki korelasyonu bileşenlerine ayıran ve değişkenler arasındaki doğrudan ve dolaylı etkileri belirleyen bir analiz yöntemidir. Değişkenler arasındaki ilişkinin ölçüsü korelasyon katsayısıdır. İki değişken arasındaki korelasyon katsayısı bu iki değişkenin birlikte değişim derecesidir. Yani iki değişken arasında hesaplanan korelasyon katsayısı yüksek ise, bu iki değişkenin birbirine bağlı olduğunu ve birlikte değiştiğini söyleyebiliriz. Ancak, iki değişken arasında hesaplanan korelasyon katsayısı başka bir değişken ya da değişkenler tarafından etkileniyor ise, yani iki değişken arasındaki sebep-sonuç ilişkisi üçüncü bir değişkenin etkisine bağlı ise, korelasyon katsayısı bu ilişkiyi açıklamada yeterli değildir. Ayrıca sistemde bunlar ile ilişkili olduğu düşünülen başka değişkenlerin etkisi olabilir. Değişkenler arasında hesaplanan korelasyon katsayısında diğer değişkenler ile olan ilişkiden kaynaklanan kısımların bulunması istendiğinde Path Analizi kullanılmaktadır. Path katsayıları kısmi regresyon katsayılarının hesaplanmasıyla bulunur.

Değişkenler arasındaki sebep-sonuç ilişkilerinin belirlenmesi ve bu ilişkiler sisteminden elde edilen sonuçların yorumlanmasına Path Analizi tekniği önemli bir katkı sunmaktadır.

Bu çalışmada Path Analizi tekniğinin matematiksel alt yapısının hemen her alandaki araştırmacılar tarafından anlaşılabilir ve yorumlanabilir hale getirilmesine ve Korelasyon Analizi ile Çok Değişkenli Regresyon'la ilişkileri içinde ele alınarak açıklanmasına çalışılmıştır.

Path Analizi'ni bir uygulamayla örneklendirmek amacıyla, mercimeğin bazı morfolojik değişkenlerinden hareketle mercimekte tane veriminin tespitinde Path Analizi yöntemi kullanılmış ve sonuçlar detaylı olarak verilmiştir. Önce Path şeması kurulmuş, sonra her bir değişkenin bitki tane verimi ile arasındaki korelasyon Doğrudan (DE), Dolaylı (IE), S ve U etkilerine ayrılmıştır. Kullanılan morfolojik değişkenlerin hangi yollarla ne kadar etkilendiği tespit edildikten sonra, bu sonuçlar her değişken için yorumlanmıştır.

Sonuç olarak bitki boyu ile bitki tane verimi arasındaki korelasyon 0,539 iken bunun sadece % 32'sinin bitki boyundan kaynaklandığı, iki değişken arasındaki ilişkinin başka değişkenler üzerinden olduğu belirlenmiştir. Dal sayısı ile bitki tane verimi arasında pozitif bir 0,371'lik korelasyon olmasına karşın, dal sayısı bu korelasyona -0,063'lük Doğrudan Etki (DE) ile negatif bir etki yapmaktadır.

Bakla sayısı ile verim arasında 0,929'luk yüksek ve istatistiksel olarak önemli bir korelasyon bulunmasına karşın, Path Analizi sonuçları bize, bakla sayısının Doğrudan Etkisi'nin sadece 0,357 olduğunu, en fazla etkinin tane sayısı ile arasındaki U etkisinden kaynaklandığı görülmektedir. Path Analizi sonuçları, iki değişken arasındaki ilişkinin başka değişkenler üzerinden gerçekleştiğini göstermektedir. Çok değişkenli bir analiz olan Path Analizi yöntemini kullanmayıp sadece korelasyon analizi ile yetinmiş olsaydık, epeyce yanıltılmış olacaktık. Bu nedenle mümkün ve uygun olduğunda tek değişkenli tekniklerin kullanılması araştırmalardan elde edilecek sonuçlar açısından daha sağlıklı olacaktır.

Anahtar kelimeler: Path Analizi, Korelasyon Analizi, Çoklu Regresyon Analizi.

DÖNGÜSEL SÜREÇLERDE OLAY KESTİRİMİ: MARKOV MODEL YAKLAŞIMI

Necdet Süt¹, Mevlüt Türe², Mustafa Şenocak¹

¹*İstanbul Üniversitesi, Cerrahpaşa Tıp Fakültesi, Biyoistatistik B.D.*

²*Trakya Üniversitesi, Tıp Fakültesi, Biyoistatistik A.D.*

e-mail: necdetsut@yahoo.com

Özet

Bu çalışmada Markov modelin tıpsal karar vermede ve kestirimde bulunmada kullanımının Akut Lenfoblastik Lösemiye ilişkin örnek bir uygulama ile gösterilmesi amaçlandı.

Markov model, olayların zamanlaması önemli olduğunda, zaman aşımı karar probleminde risk yaratıyorsa ve olaylar birden fazla kez tekrar edebiliyorsa (recurrence - nüks gibi) yararlı olan bir modellemedir. Markov Modellerin karar ağacı modellerinden farkı, belirsiz olayları şans düğümlerinde modelleme yerine tanımlı sağlık durumları arasında geçişler olarak modellemesidir. Ayrıca modelin döngüsel yapısı gelecekte olabilecek olayları tahmin etmek için oldukça faydalıdır. Model matris cebiri şeklinde Kohort yada Monte Carlo benzetimi olarak değerlendirilebilir.

Sonuç olarak, Markov modeller yardımıyla (alternatif) klinik stratejilerinin maliyet, etkinlik ve yaşam kalitesi ölçütlerini sentez edilerek, bunların sonucunda yaşam beklentisi, kalite düzeltmeli yaşam beklentisi ve yaşam maliyeti hesaplanabilir.

Anahtar Kelimeler: Markov model, döngüsel süreç, tıpsal karar.

Giriş

Modelleme kavramı farklı disiplinlerde kullanılmaktadır. "Araştırma Çıktılarının Değerlendirilmesinde (Outcomes Research)" model, tedavi yöntemlerinin hasta/toplum açısından çeşitli durumlarda sağlık ve ekonomik sonuçlarının gösterildiği matematiksel yapı olarak ifade edilebilir. Diğer bir deyişle ilgilenilen özel bir fenomenin (hastalık, tedavi vb.) aslına çokbenzeyen bir durumunu yaratmak için değişkenler arası ilişkilerin sistematik yapılarıyla ilgilenildiği analitik yapıdır. Model yapılarıyla fenomenin ana parametrelerinin gelecekteki değerleri tahmin edilebilir ve bu parametreler

arasındaki ilişkiler ortaya konulabilir, ayrıca modelde zaman boyutu da yer alabilir (1, 2, 3).

Beck&Pauker 1983 yılında karar ağacı modellerinden farklı olan Markov modellerinin tıbbi uygulamalarda prognozu belirlemek için uygulanabilirliğini tanımladıklarından itibaren Markov modeller artan bir hızla sağlık alanında kullanılır bir konuma gelmiştir (3).

Dünya literatüründe tıpsal karar vermede ve kestirimde bulunmada oldukça yaygın olarak kullanılan ancak ülkemiz literatürü tarandığında tıpsal alanda uygulaması bulunamayan Markov modelin döngüsel süreçlerde tıpsal karar vermede ve geleceğe yönelik kestirimde bulunmada kullanımını Akut Lenfoblastik Lösemiye ilişkin örnek bir uygulama ile göstermeyi amaçladık.

Yöntem

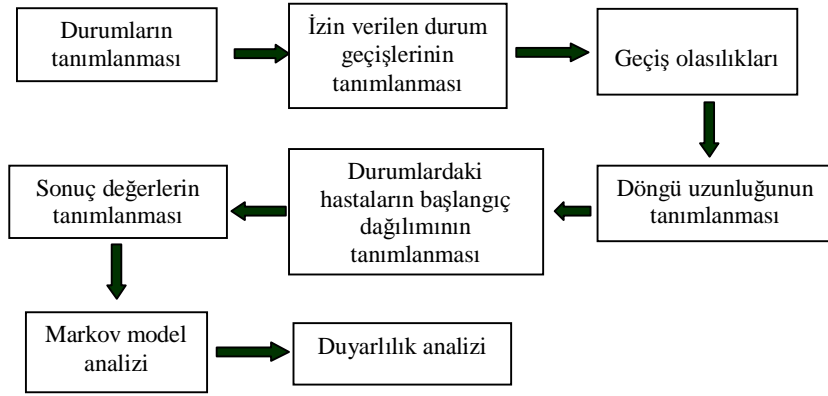
Markov Model, sonlu sayıda durumla ifade edilebilen bir olasılıksal karar sürecinde bu durumlar arası geçişlerin olasılık değerleriyle –sürecin bir durumundan diğer bir durumuna geçmesi olasılığı– ifade edildiği yapıdır. Amaç incelenen sorunun beklenen sonucuna ilişkin optimum yapıyı belirlemektir (3, 4).

Model hastayı kesikli sağlık durumunda ele almaktadır ve bu kesikli durumlar 'Markov durumları' olarak adlandırılır. Durum kavramıyla, Markov sisteminde var olan olası tüm sağlık sonuçları ifade edilmektedir. Markov sistemi belli bir zamanda sadece bir sağlık durumunu göstermektedir. Markov modelinde durumlar dairesel sembollerle gösterilmektedir. Markov modelinde geçişler, dairesel sembollerle gösterilen Markov durumlarının geçiş yönünü gösteren oklarla birleştirilmesiyle ifade edilmektedir. Geçiş, Markov diyagramında bir sağlık durumundan diğer bir sağlık durumuna geçilmesi olasılığını göstermektedir. İyi bir sağlık durumundan kötü bir sağlık durumuna geçiş 'başarısızlık oranıyla', kötü bir sağlık durumundan iyi bir sağlık durumuna geçiş 'onarım (iyileşme, şifa) oranıyla' ifade edilmektedir (3, 5, 6, 7, 8, 9). Markov model bir Kohort ya da Monte Carlo benzetimi gibi cebirsel olarak matrisle değerlendirilebilir. Ayrıca Markov modeller yardımıyla klinik olayların gösteriminde bir döngü ağacı kullanılarak 'Markov döngü ağacı' bir Kohort ya da Monte Carlo benzetimi gibi değerlendirme yapılabilir (3, 5).

- 1- **Kohort benzetimi:** Hastalık durumlarının herhangi birinde hipotetik bir kohort hasta grubu ile başlanan modelde kohort model süresince izlenir. Zamanın herhangi bir anında herhangi bir durumda kohort oranı ve her bir durumun ortalama süresi hesaplanabilir.
- 2- **Monte-Carlo benzetimi:** Çok sayıda hipotetik hasta model sürecinde tek tek geçer ve bunların geçiş yolları kaydedilir. Monte-

Carlo yaklaşımının avantajı, ayrılma (dispersiyon) süresinin tahmininin tek tek simülasyon verilerinden elde edilebilir olmasıdır.

Markov model, zamana bağlı olarak tekrarlayan olayların gösterimini ve bu olaylarla ilgili klinik durumların olasılık ve yararlarının daha hızlı gösterimini sağlamaktadır (3, 5, 10, 11, 12). Markov modelin, geliştirme aşamaları aşağıda Şekil 1'de gösterilmektedir.



Şekil 1. Markov model geliştirme aşamaları

Herhangi bir hastalığın K farklı aşamasında seyrinin N bağımsız birimde incelendiğini düşünelim. Bunu yapabilmek için K aşamalı olasılıksal sürecin (stochastic process) düzenlenmesi gerekmektedir.

Gözlemler serisinde her bir birim i ile simgelenerek aşağıdaki gibi ifade edilir,

$$D_{ij} = (X_{ij}, T_{ij})_{\substack{i \in \{1, \dots, N\} \\ j \in \{1, \dots, n_i\}}}$$

formüldeki

n_i : i bireyinin örnek (model) sayısıdır,

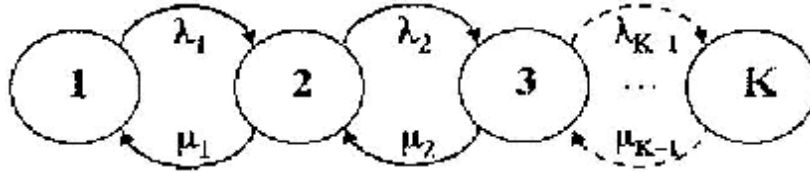
X_{ij} : j. örnekte (modelde) i bireyinin gözlemlenen durumunu

T_{ij} : j.vizitte X_{ij} aşamasında gözlemlenen i bireyi zamanını

ifade etmektedir.

$\theta = (\lambda, \mu)$ olarak ele alındığında, $\lambda(\lambda = \lambda_1, \dots, \lambda_{K-1})$ ve $\mu(\mu = \mu_1, \dots, \mu_{K-1})$

ardışık aşamalar arasındaki geçiş olasılıklarını ifade etmektedir. Homojen zaman dilimlerinde izlem yapılmış K geriye dönülebilir aşamalı uzunlamasına bir Markov modellemesinin geçiş grafiği Şekil-2'de gösterilmiştir (6).



Şekil 2. K aşamalı geriye dönülebilir Markov modelinde geçişlerin grafiksel gösterimi

Şekil-2'deki düğümler ilgilenilen olayın (hastalık, tedavi v.b.) K aşamasını, oklar ilgilenilen geçiş oranlarının olası geçiş yönünü göstermektedir. λ_k , k durumundan $k+1$ durumuna geçiş oranını ve μ_k $\{1, \dots, (K-1)\}$ 'deki tüm k değerleri için $k+1$ durumundan k durumuna onarım (iyileşme, şifa) oranını ifade etmektedir.

K durumdan oluşan $S = \{1, 2, \dots, K\}$ bir Markov modelde düğümler arası geçiş ve onarım oranlarının yer aldığı geçiş matrisi Λ 'nın açık şekilde ifade edilişi aşağıdaki gibidir (6),

$$\Lambda = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \dots & 0 \\ \mu_1 & -(\mu_1 + \lambda_2) & \lambda_2 & 0 & \vdots \\ 0 & \mu_2 & -(\mu_2 + \lambda_3) & \lambda_3 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 0 \\ \vdots & \vdots & \vdots & 0 & \mu_{k-2} & -(\mu_{k-2} + \lambda_{k-1}) & \lambda_{k-1} \\ 0 & \dots & \dots & 0 & 0 & \mu_{k-1} & -\mu_{k-1} \end{pmatrix}$$

Matrisin (l, m) elemanı aşama l 'den aşama m 'ye geçiş yoğunluğunu göstermektedir.

$P(t) = (P_{l,m}(t))_{l,m}$ geçiş olasılık matrisidir. Matristeki $P_{l,m}(t)$ ifadesi $\{1, \dots, K\}$ 'daki tüm l, m için t zamanında l durumundan m durumuna geçiş olasılıklarını gösterir. Geçiş olasılık matrisi $P(t)$ ve geçiş yoğunluk matrisi Λ arasındaki ilişki Kolmogorov ileri diferansiyel denklemiyle,

$$\frac{\partial P(t)}{\partial t} = P(t)\Lambda$$

şeklinde ifade edilir (6). Diferansiyel denkleminin tekil çözümü bu sistemde $P(t) = \exp(\Lambda t)$ olarak ifade edilebilir, sınır koşullara bağlı olarak $P(0) = I$ formüldeki I birim matristir. Homojenitenin ve sürecin tersinebilirliğinin benimsenmesiyle Kolmogorov denkleminin çözümü aşağıdaki denklemle verilir,

$$P(t) = P \text{diag}\{\exp(\rho_1 t), \dots, \exp(\rho_k t)\} P^{-1}$$

Geçiş matrisi Λ 'nın özdeğerleri $(\rho_k)_k$ ile gösterilir, P , k sütunlu kare matristir, ρ_k için sağ özvektör ve $\text{diag}\{\exp(\rho_k t)\}_k$ tüm k 'lar için $\{1, \dots, K\}$ 'da $(\exp(\rho_k t))_k$ katsayılarıyla diagonal matrise uygundur.

Bu nedenle geçiş yoğunluklarına göre geçiş olasılıklarının fonksiyonel şekli

$$P_{l,m}(t) = \frac{\sum_{k=1}^K A'_{lm}(\rho_k) \exp(\rho_k t)}{\prod_{n=1, n \neq k}^K (\rho_k - \rho_n)} \quad \forall k, l, m, n \in \{1, \dots, K\}$$

formülü ile ifade edilir. Formüldeki $A'_{lm}(\rho_k)$, $(\rho_k I - \Lambda)$ transpose matrisinin (l, m) elemanının kofaktörüdür (eşçarpan). Λ geçiş matrisinin özdeğerleri $-\Lambda$ karakteristik polinomunun kökleri- yaklaşık olarak ikiye bölünür, özdeğerlerin birinin sıfır olduğu bilindiğinden diğerleri gerçek sayı ve negatiftir. Bu olasılıklar t sonsuz eğilimde olduğunda bir eşik sınır değere doğru yakınsaktırlar (6).

Bulgular

Akut Lenfoblastik Lösemiye ilişkin Markov model kurgulaması

Markov modelin sağlık alanında uygulanabilirliğini göstermek amacıyla literatür bilgileri (13, 14) ve uzman görüşleri doğrultusunda St Jude Toplam Terapi XIII B protokolünün Akut Lenfoblastik Lösemi (Acute Lymphoblastic Leukemia - ALL) hastalığında tedavi etkinliğine ilişkin bir model öngörülmüştür.

Model oluştururken öncelikle ALL'ye ilişkin sonuç durumlar saptanmıştır (13).

- I- *Remisyon*
- II- *Relaps*
- III- *Dirençli seyir*
- IV- *Ex*

Literatür bilgileri (13) ve uzman görüşleri doğrultusunda hastalığın sonuç durumları arasında izin verilen durum geçişleri aşağıdaki şekilde olabilmektedir.

Remisyon	→	Remisyon / Relaps / Ex
Relaps	→	Remisyon / Ex
Dirençli seyir	→	Remisyon / Ex

Geçiş matrisi Δ 'yı belirleyebilmek için ALL hastası bir kohortun 5 yıllık izlem sonucu elde edilen bulguları kullanılmıştır. Geçiş matrisi aşağıda ifade edilmektedir.

	Remisyon	Relaps	Dirençli Seyir	Ex
Remisyon	0,65	0,27	0	0,08
Relaps	0,24	0	0	0,76
Dirençli Seyir	0	0	0	1
Ex	0	0	0	1

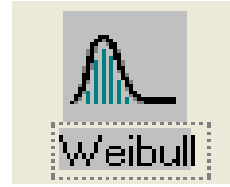
Kurguladığımız çözümlemde Markov modelin döngü uzunluğu olarak 5 yıl öngörülmüştür. Ayrıca modele girecek ALL hastalarının başlangıç durum dağılımı olarak Weibull dağılımı öngörülmüştür.

Weibull Dağılımı

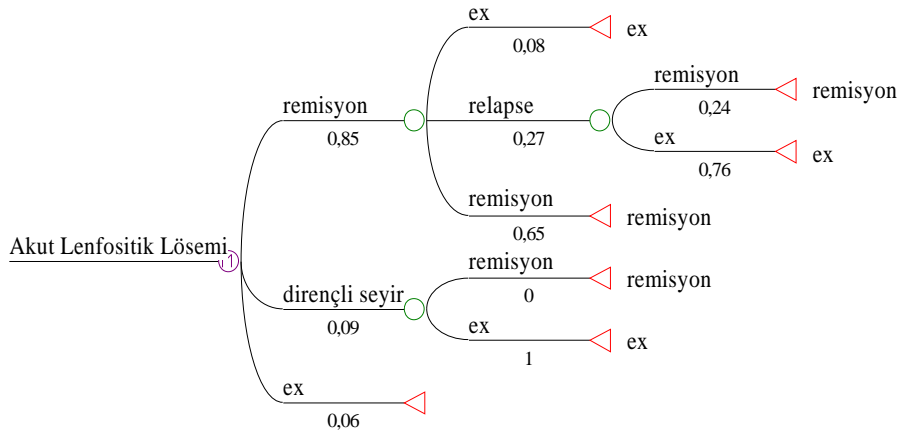
$$f(x; a, c) = 1 - e^{-(x/a)^c}$$

$$x > 0; a, c > 0$$

a, c; Weibull dağılımının parametreleri



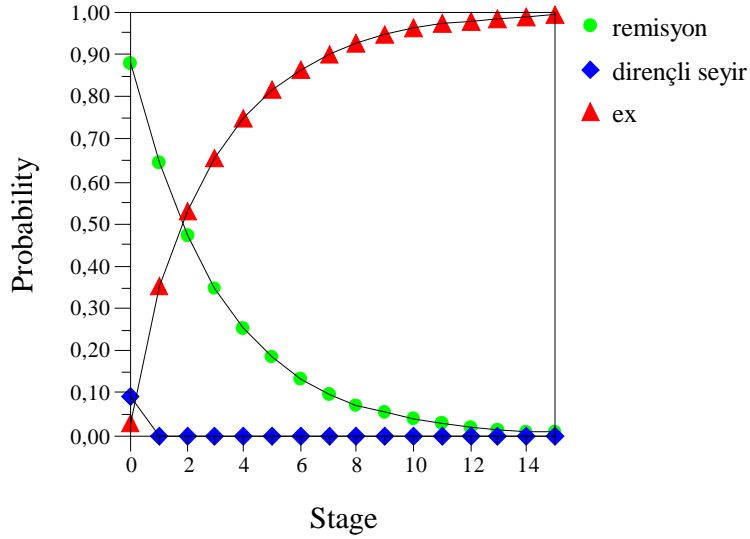
ALL'ye ilişkin Markov Döngü ağacı aşağıdaki şekil 3'te gösterilmektedir. Şekilde ALL kohortunun her bir döngüdeki dağılımı, sonuç durumlar arasındaki olası geçiş yolları ve geçiş olasılıkları yer almaktadır.



Şekil 3. Markov döngü ağacı (M, markov düğümü)

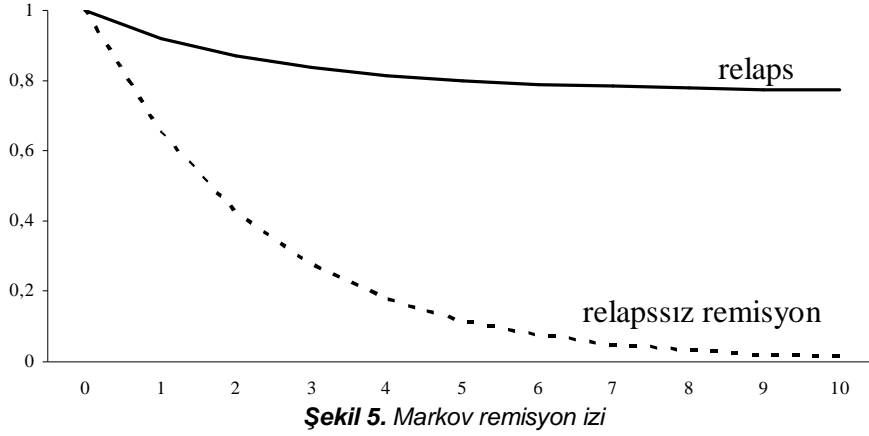
Şekil-4, Markov döngü ağacının (Şekil-3) çözümlemesi sonucu markov durumlarına ilişkin olarak elde edilen ilk 15 döngüye ilişkin kohortun her bir döngüdeki dağılımının grafiksel gösterimini ifade etmektedir. Şekle göre döngü uzunluğu arttıkça (>2) söz konusu kohortun ex olma olasılığının remisyonda kalma olasılığından daha yüksek olduğu gözlenmektedir. Dirençli seyreden olguların tümü izlem döneminde remisyona girmeden ex oldukları için bu markov durumunun diğer döngülerdeki olasılık değerleri 0 olarak yer almaktadır.

Markov Probability Analysis



Şekil 4. ALL Kohortunun Markov olasılık çözümlemesi

ALL hastalığında amaç olguyu relaps ya da ex olmadan remisyonda tutabilmektir. İzlenen kohortun verileri doğrultusunda başlangıç durumu remisyon olan 10000 denemelik bir kohort benzetim çözümlemesi düzenlenerek 10 döngü sürecinde olguların remisyonda kalma beklentisi hesaplanmıştır. Şekil 5'te söz konusu tedavi uygulaması ile döngülerde Relaps olmadan remisyonda kalma ve relaps olma olasılıkları yer almaktadır. Şekle göre 2. döngüde herhangi bir olgunun relapssız remisyonunda kalma olasılığı % 43'tür, bu döngüden sonra olasılık artan bir hızda düşmektedir.



Tartışma

Tedavi yöntem/rejimlerine bağlı olarak hastalıkların gelecekteki seyri çok büyük önem taşımaktadır. Hastalık sonuç değerlerinin gelecekte ne olabileceği –incelenen olay ölümcül sonuç değer oluşturabiliyorsa yada toplumda nadir gözleniyorsa, AIDS gibi– yönünde öngörülerde bulunabilmek tıpsal karar vericiler ve hastalar için büyük önem taşımaktadır. Bu nedenle Markov modeller, zamana bağlı olarak tekrarlayan olayların ve bu olaylarla ilgili klinik durumların olasılık ve yararlarını ele almak için tıpsal karar vericilere yol gösteren bir araç olarak kullanılabilirler. Model, hastalıklar zaman sürecinde birbirini izleyerek gözlemleniyorsa –örneğin stabil anjinal hastaların akut MI geçirmesi yada lokal meme kanserli hastalarda kanserin yeniden nüks etmesi gibi– bu hastalıklara ilişkin kestirimlerde bulunmaya olanak tanır. Modelin döngüsel yapısı yardımıyla belli bir zamanda veya zaman sürecinde meydana gelecek olaylar tahmin edilebilir.

Sonuç olarak Markov modeller yardımıyla (alternatif) klinik stratejilerin yaşam beklentisi, kalite düzeltmeli yaşam beklentisi (QALE-quality-adjusted life expectancy) ve yaşam maliyetleri hesaplanarak, maliyetler, etkiler ve sağlıkla ilgili yaşam kaliteleri değerlendirilebilir. Ayrıca model Monte Carlo ve Kohort benzetimleri yardımıyla nadir gelişen olayları makul bir zaman süreci içerisinde ele almak için uygun bir yöntemdir.

Kaynaklar

1. Drummond M.F., McGuire A. Economic Evaluation in Health Care, Oxford University Press, 2001.

2. Buxton M.J., Drummond M.F., Van Hout B.A., Prince R.L., Sheldon T.A., Szucs T., Vray M. Modelling in Economic Evaluation: An Unavoidable Fact of Life Health Economics, 6: 217–227, 1997.
3. Sonnenberg F.A., Beck J.R. Markov Models in Medical Decision Making: A Practical Guide, Medical Decision Making, 13:322-338, 1993.
4. Taha H.A. Operation Research an Introduction, Prentice-Hall, NJ, 1997.
5. Hunink M., Glasziou P. Decision making in health and medicine, Cambridge University Pres, 2001.
6. Kousignian I., Autran B., Chouquet C., Calvez V., Gomard E., Katlama C., Riviere Y., the IMMUNOCO Study group, Costagliola D. Markov modelling of changes in HIV-specific cytotoxic T-lymphocyte responses with time in untreated HIV-1 infected patients, Statistics in Medicine, 22:1675 –1690, 2003.
7. Karnon J. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation, Health Economics, 12: 837–848, 2003.
8. Briggs A., Sculpher M. An introduction to Markov modelling for economic evaluation, Pharmacoeconomics, 13(4):397-409, 1998.
9. Michaels JA, Galland RB. Management of asymptomatic popliteal aneurysms: the use of a Markov decision tree to determine the criteria for a conservative approach. Eur J Vasc Surg 7:136-43, 1993.
10. McLaughlin F.S., Pickhardt R.C. Quantitative Techniques for management decisions, Houghton Mifflin Company, Bostyon, 1997.
11. Plunkett BA, Grobman WA. Routine hepatitis C virus screening in pregnancy: A cost-effectiveness analysis, American Journal of Obstetrics and Gynecology, 192: 1153–1161, 2005.
12. Empanan C., Wolters H., Laukötter M., Dame C., Senninger N. Cost-Effectiveness Analysis of Basixilimab Induction and Calcineurin-Sparing Protocols in “Old to Old” Programs Using Markov Models, *Transplantation Proceedings*, 35:1324-1325, 2003.
13. Hoelzer D., Gökbüget N., Ottmann O., Pui CH., Relling MV., Appelbaum FR., van Dongen JJM. and Szczepański T. Acute Lymphoblastic Leukemia, Hemathology, 162-192, 2002.
14. Pui CH, Sandlund JT, Pei D, Campana D, Rivera GK, Ribeiro RC, Rubnitz JE, Razzouk BI, Howard SC, Hudson MM, Cheng C, Kun LE, Raimondi SC, Behm FG, Downing JR, Relling MV, Evans WE. Improved outcome for children with acute lymphoblastic leukemia: results of Total Therapy Study XIIIB at St Jude Children's Research Hospital, Blood, 1;104(9): 2690-6, 2004.

18 YAŞ ÜZERİ SAĞLIKLI BİREYLERDE KANTİTATİF ULTRASON ÖLÇÜMLERİNİN İNCELENMESİNDE DEĞİŞİM KATSAYISININ KULLANILMASI

Başak Gürtekin, Sevda Özel, Rian Dişçi

*İstanbul Üniversitesi İstanbul Tıp Fakültesi, Biyoistatistik ve Demografi Bilim
Dalı*

e-mail: bgurtek@istanbul.edu.tr

Özet

Çalışmamızda, Tıp'ta, Fizik Tedavi alanında kemik kalitesini belirlemede topuktan yapılan tekrarlı ölçümlerle elde edilen Kantitatif Ultrason (KUS) değerlerindeki değişkenliğin ve tekrarlı ölçüm sayısının, değişim katsayısı kullanılarak belirlenmesi amaçlanmıştır.

Çalışma 20.10.2004-27.12.2004 tarihleri arasında İstanbul Üniversitesi Fizik Tedavi ve Rehabilitasyon Yüksekokulu'nda gerçekleştirilmiştir. Topuk ölçümleri Fizik Tedavi ve Rehabilitasyon Yüksekokulu'na ait Sahara Klinik Kemik Sonometresi (Hologic, Waltham, MA, USA) ile 18 yaşın üzerinde 20 kadın ve 20 erkek bireye iki ayrı gözlemci tarafından yapılmıştır. Ölçümler, ilk 5'i 1. gözlemci, ikinci 5'i 2. gözlemci tarafından olmak üzere tek bir topuktan toplam 10 kez tekrarlanmıştır.

İki gözlemci tarafından gerçekleştirilen ölçümlerde KUS değişkenlerinden biri olan BUA için bireyin kısa dönem kesinliği ($SD_{kesinlik}$) ve bu kesinliğin yüzde olarak ifadesi (CV_{SD}) kadınlarda ve erkeklerde ayrı ayrı hesaplanmıştır.

Bu çalışmada iki gözlemci arasında fark olduğu ve tekrarlı ölçüm sayısının artırılmasının değişkenliği azaltmada etkili olmadığı, uygun tekrarlı ölçüm sayısının ise 2 olmasının yeterli olacağı kanısına varılmıştır.

Anahtar kelimeler: Değişim katsayısı, kantitatif ultrason, topuk.

Giriş

Bir serinin tüm değerlerini içeren dağılım ölçülerinden biri de değişim katsayısıdır. Standart sapmanın büyük olması popülasyonu meydana getiren değişkenlerin kendi ortalamalarından olan sapmalarının fazla olması anlamını taşımaktadır. Ancak bazı hallerde, standart sapma, dağılımı gerçek

anlamda yansıtmayabilir. Bu durum, özellikle iki popülasyon dağılımının birbiriyle karşılaştırılmasında karşımıza çıkmaktadır.

İki popülasyondan birini meydana getiren gözlemlerin değer bakımından diğer popülasyonu teşkil edenlerden daha büyük olması yanında standart sapması da diğerinden daha büyük olacaktır. Halbuki gerçekte belki de, o popülasyonu meydana getiren gözlemlerin kendi ortalamalarından olan sapmaları diğer popülasyondakinden daha büyük değildir. İşte standart sapmanın bu sakıncasını gidermek için, dağılımın diğer bir ölçüsü olarak σ/μ şeklinde ifade edilen “değişim katsayısı” kullanılmaktadır.

Değişim katsayısı kavramı, istatistikçi Karl Pearson (1857-1936) tarafından 1895’te ortaya atılmıştır

(1). Örnekten hesaplandığında ve % olarak ifade edilmek istendiğinde

$$D.K. = \frac{s}{\bar{x}} \times 100 \quad (1)$$

şeklinde formüle edilebilir (2).

s : standart sapma

\bar{x} : aritmetik ortalama

Kısaca, değişim katsayısı birim sayıları ve ölçü birimleri birbirlerinden farklı olan değişkenlerin ortalamaya göre yayılışlarını karşılaştırmak için yararlanılan bir dağılım ölçüsüdür. Ölçü birimi yoktur.

- DK % < %5 ise değişkenin; homojen, dar bir değer aralığında yayılma gösterdiği ve X’in türdeş dağılıma sahip olduğu varsayılır.

- % 5 ≤ DK % < %10 ise X’in normal genişlikte sayılabilecek bir dağılıma sahip olduğu varsayılır.

- % 10 ≤ DK % < %20 ise değişkenin türdeş dağılımının kaybolmaya başladığı varsayılır.

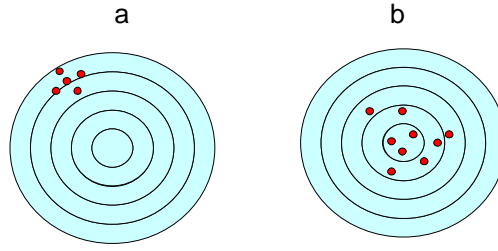
- DK% ≥ %20 ise değişkenin türdeş olmayan bir dağılıma sahip olduğu ya da verilerin hatalı toplandığı varsayılır (3).

Değişim katsayısı, Tıp’ta Fizik Tedavi alanında kemik kalitesini belirlemek amacıyla topuktan yapılan Kantitatif Ultrason ölçümlerindeki değişkenliği dolayısıyla ölçümün kesinliğini belirlemek için kullanılmaktadır.

Aynı bireyde topuktan yapılan ölçümler her seferinde farklı sonuçlar verdiği için ölçüm birkaç kez yinelenmektedir. Çalışmamızda da bu kesinliği belirlemek ve kaç ölçümde en iyi sonucun alınacağını bulmak amaçlandığı için değişim katsayısı kullanılmıştır.

Kesinlik (Presizyon) Hataları

Kesinlik hataları (precision errors), tanısal bir tekniğin tekrarlanabilirliğinin belirlenmesi için tanımlanır. Doğruluk hataları da (Accuracy errors), ölçülen sonuçların “gerçek” değerlerden ne derecede saptığını yansıtır. Kesinlik ve doğruluk arasındaki farkı anlatmak için sıklıkla bir okçunun performansı örneği verilir (Şekil 1a, b).



Şekil 1. a) Kesinlik

b) Doğruluk

Yukarıdaki şekilde de (Şekil 1b) görüldüğü gibi bir okçu, hedefe yakın dairelere ok atıyor ve dairelerin dışına çıkmıyorsa doğruluğu yüksek ancak kesinliği düşük atışlar yapıyor. Okçu hedefe uzak bir yere ok atıyor ama atışlarında hep aynı noktayı tutturuyorsa doğruluğu düşük ancak kesinliği yüksek atışlar yapıyor (Şekil 1a).

Bireyin Kısa Dönem Kesinliği

Bir bireydeki tekrarlanan ölçümlerin rassal değişimlerinin normal dağıldığı varsayıldığında kesinlik, Gaussian olasılık dağılımında σ parametresinin tahmini ile temsil edilir. Kısa dönem kesinlik (SD), j deneğindeki $i = 1, \dots, n_j$ 'ye kadar olan tekrarlı ölçümlerin standart sapması olarak tanımlanır(4).

$$SD_j = \sqrt{\frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{n_j - 1}} \quad (2)$$

n_j : ölçüm sayısı

x_{ij} : j deneğinde yapılan i. ölçümün sonucu

\bar{x}_j : j deneğindeki tüm x_{ij} 'lerin ortalaması

Ölçümün gerçek ortalaması bilinmediğinden ve n kere tekrarlanan ölçümlerin ortalamasından tahmin edileceğinden SD^2 'nin Gaussian olasılık dağılımında σ^2 parametresinin sapmasız tahmincisi olarak değerlendirilebilmesi için paydanın (n_j-1) olarak ifade edilmesi gerekir. Bu uyarlama SD 'nin ölçümlerinin tekrarlanma sayısından bağımsız olduğunu garanti eder. Paydadaki (n_j-1) yani (tekrarlanan ölçüm sayısı-1) değeri bu değerlendirme ile ilişkili serbestlik derecesi (s.d.)'dir.

Kesinlik hataları, tam sayılarla ya da tekrarlanan ölçümlerin tipik olarak yüzdelerle verilen değişim katsayısı (CV) formülü ile ifade edilir (4).

$$CV_{SD_j} = \frac{SD_j}{\bar{x}_j} \times 100\% \quad (3)$$

\bar{x}_j : Tüm x_{ij} 'lerin ortalaması

Genel uygulamanın aksine kesinlik hatasının doğru değerlendirmesi, deneklerdeki bireysel kesinlik hatalarının aritmetik ortalaması ile değil varyanslarının (SD^2) aritmetik ortalaması ile tanımlanır (4).

$$SD^2 = \left(\sum_{j=1}^m SD_j^2 \right) / m \quad (4)$$

İstatistiksel olarak bunun nedeni, standart sapmanın değil varyans (SD^2)'nin Gaussian normal dağılımının σ^2 parametresinin sapmasız tahmini olarak sayılabileceğidir. Bunun sonucunda her bir m deneği için (2) numaralı denklem ile hesaplanmış olan kesinlik hataları aşağıdaki ifade ile verilmektedir (4).

$$SD = \sqrt{\left(\sum_{j=1}^m SD_j^2 \right) / m} \quad (5a)$$

Her bir birey için iki ölçüm yapılırsa (birinci ve ikinci sonuç arasındaki fark d_j olarak gösterilirse) aşağıdaki formül uygulanır;

$$SD = \sqrt{\left(\sum_{j=1}^m d_j^2 \right) / 2m} \quad (5b)$$

Kesinliđi yüzde olarak göstermek istersek řu formülün kullanılması önerilir;

$$C V_{SD} = \left(SD / \left[\left(\sum_{j=1}^m \bar{x}_j \right) / m \right] \right) \times 100 \% \quad (6)$$

Gereç ve Yöntem

Bu çalışmada topuktan yapılan ölçümler (20.10.2004 – 27.12.2004) tarihleri arasında İstanbul Üniversitesi Fizik Tedavi ve Rehabilitasyon Yüksekokulu'nda gerçekleştirilmiştir. Çalışmaya 18 yaşın üzerinde 20 kadın ve 20 erkekten oluşan toplam 40 kişi katılmıştır. Ölçümlerin yapılmasını sağlayan topuk KUS aleti İstanbul Üniversitesi Fizik Tedavi ve Rehabilitasyon Yüksekokuluna ait Sahara Clinical Bone Sonometer (Hologic, Waltham, MA, USA) idi (Şekil 2).



Şekil 2. Sahara Clinical Bone Sonometer
(Sahara Klinik Kemik Sonometresi) resmi

Bu çalışmada bireylere konu hakkında bilgi vererek çalışmaya gönüllü olarak katılmaları sağlanmıştır. Ölçümlere geçmeden önce bireyde dominant olma durumuna bağlı olarak hangi ayaktan ölçümler yapılacaksa o ayak çıplak olarak KUS cihazında ayak platosuna yerleştirilmiştir (Şekil 3) ve

ayağın hareket etmesini önlemek için şekil 3'te de görüldüğü gibi ayak bir kayış ile bağlanmıştır.



Şekil 3. KUS ile yapılan ölçümün önden görünüşü

Kantitatif Ultrason aletindeki kauçuk pedler (problar) topuğun iki yanından ayağa temas etmektedir. Pedlere yağ bazlı ultrason temas jeli sürülür. Her bireyde boy uzunluğunun farklılık göstermesinden dolayı oturma pozisyonları KUS cihazına bağlı olarak ayarlanmıştır. Daha sonra ölçümler iki ayrı gözlemci tarafından, belirlenen tek bir topukta toplam 10 kez tekrarlanmış olup, ilk 5 ölçüm 1. gözlemci, ikinci 5 ölçüm ise 2. gözlemci tarafından yapılmıştır. Her bir birey için ölçüm işlemi ortalama 20 dakika sürmüştür ve çıkan sonuçlar kaydedilmiştir.

Bulgular

Ölçüm yapılan 40 deneğin 19'u (% 47,5) 18 – 20 yaş grubunda; 21'i (% 52,5) 21 – 27 yaş grubunda idi.

Tablo 1. Deneklerin yaş, boy, kilo ve beden kitle indeksi (BKİ) değerleri

Cinsiyet	n	Yaş (yıl) $\bar{x} \pm SD$	Boy (cm) $\bar{x} \pm SD$	Kilo (kg) $\bar{x} \pm SD$	BKİ (kg/m ²) $\bar{x} \pm SD$
Kadın	20	21,05 ± 2,46	164,35 ± 6,32	54,55 ± 4,54	20,20 ± 1,32
Erkek	20	20,95 ± 1,76	177,00 ± 6,21	68,90 ± 6,75	22,03 ± 2,33
Toplam	40	21,00 ± 2,11	170,68 ± 8,90	61,73 ± 9,22	21,11 ± 2,09

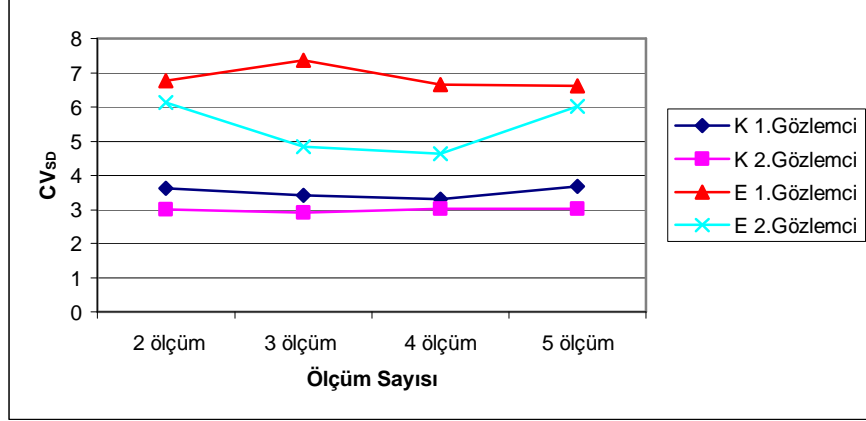
Tablo 2. BUA deęişkeninin cinsiyete, gözlemciye ve ölçüm sayılarına göre tanımlayıcı istatistikleri

	Ölçüm Sayısı	1. GÖZLEMÇİ			2.GÖZLEMÇİ		
		\bar{x}	\pm	SD	\bar{x}	\pm	SD
Kadın (n=20)	1	78,74	\pm	12,98	78,16	\pm	13,39
	2	78,75	\pm	12,86	77,91	\pm	13,20
	3	78,81	\pm	12,93	78,16	\pm	13,30
	4	78,82	\pm	12,53	78,17	\pm	13,16
	5	78,77	\pm	12,60	78,15	\pm	13,27
Erkek (n=20)	1	81,89	\pm	17,85	80,63	\pm	14,08
	2	81,71	\pm	15,22	81,55	\pm	15,61
	3	82,10	\pm	15,43	81,13	\pm	15,31
	4	82,23	\pm	15,80	80,90	\pm	15,01
	5	82,48	\pm	15,67	81,13	\pm	15,50

Tablo 3: Kadınlarda ve erkeklerde BUA deęişkeni için iki gözlemci tarafından gerçekleştirilen ölçümlerde bireyin kısa dönem kesinlięi ($SD_{kesinlik}$) ve bu kesinlięin yüzde olarak ifadesi (CV_{SD})

Cinsiyet	Ölçüm sayısı	1.Gözlemci		2.Gözlemci	
		$SD_{kesinlik}$	CV_{SD}	$SD_{kesinlik}$	CV_{SD}
Kadın (n=20)	2	2,848	3,617	2,345	3,010
	3	2,689	3,412	2,279	2,916
	4	2,608	3,309	2,363	3,023
	5	2,904	3,687	2,370	3,032
Erkek (n=20)	2	5,531	6,769	4,995	6,125
	3	6,047	7,365	3,932	4,847
	4	5,475	6,658	3,749	4,634
	5	5,457	6,616	4,878	6,013

Tablo 3 ve Grafik 1 de görüldüğü gibi 2. gözlemciye göre hem kadınlarda hem de erkeklerde daha düşük kal-maktadır. Ölçüm sayısının artması deęişkenlięi etkilememektedir. Ölçüm sayısının 2 olması yeterli olacaktır.



Grafik 1. BUA değişkenine ait CV_{SD} değerlerinin cinsiyete, gözlemciye ve ölçüm sayısına göre aldığı değerler

Tartışma ve Sonuç

KUS değişkenleri için yapılan ölçümlerin kesinliğini belirlemede değişim katsayısı sıklıkla kullanılmaktadır (5, 6, 7, 8, 9). Araştırmacılar, değişkenliği azaltabilmek yani kesinliği yüksek bir sonuç elde edebilmek için ölçüm sayılarını yüksek tutmaktadırlar. Gereksiz yere tekrar sayısının yüksek olması araştırmayı zaman yönünden olumsuz olarak etkiler. Bu nedenle çalışmamızda KUS değişkenleri için tekrar sayısının kaç olması gerektiğinin belirlenmesi amaçlandı. Çalışmada, iki gözlemci tarafından gerçekleştirilen 5'er tekrarlı ölçümün sonuçları incelendiğinde; gözlemcilerin elde ettiği sonuçların farklı olduğu ve ölçüm sayısını arttırmakla daha iyi sonuçların alınmadığı görüldü. Kantitatif Ultrason (KUS) değişkenlerinden BUA için 2. gözlemcinin elde ettiği değerlerin 1. gözlemciye göre daha iyi olduğu gözlemlendi. Söz konusu farkın kullanılan jel miktarından ve ayakların sabitlenme durumundan kaynaklandığı düşünüldü. Ölçümün fazla sayıda tekrarlanması gereksiz olduğu ve tekrar sayısının 2 olmasının yeterli olacağına karar verildi.

Kaynaklar

1. Zar JH. Biostatistical Analysis. 3rd edition., Prentice Hall, New Jersey; 1996.
2. Yurtsever N. Deneysel İstatistik Metotlar, 1. Baskı, Toprak ve Gübre Araştırma Enstitüsü Müdürlüğü Yayınları, Ankara; 1984: 40-41.

3. Özdamar K. SPSS ile BİYOİSTATİSTİK, 4. baskı, Kaan Kitabevi. Eskişehir; 2001.
4. Glüer CC, Blake G, Lu Y, Blunt BA, Jergas M, Genant HK. Accurate Assessment of Precision Errors: How to Measure the Reproducibility of Bone Densitometry Techniques. *Osteoporosis Int.* 1995; 5:262 – 270.
5. Drozdowska B. Quantitative ultrasound at the calcaneus in corticosteroid – treated male patients with bronchial asthma. *Ultrasound in Med. and Biol.* 2004; 30 (8): 1057 – 1061.
6. Sosa M, Saavedra P, Torres M, Alegne J, Gomez C, Macias J, Guanabens N, Hawkins F, Lozano C, Martinez M, Mosquera J, Cano R, Quesada M, Salas E and the GIUMO Study Group. Quantitative Ultrasound Calcaneus Measurements: Normative Data and Precision in the Spanish Population *Osteoporos Int* 2002; 13: 487 – 492.
7. Lee M, Czerwinski SA, Choh AC, Towne B, Demerath EW, Chumlea WC, Sun SS, Siervogel RM. Heritability of calcaneal quantitative ultrasound measures in healthy adults from the Fels Longitudinal Study. *Bone* 2004; 35: 1157 – 1163.
8. Frost ML, Blake GM, Fogelman I. Contact Quantitative Ultrasound: An Evaluation of Precision, Fracture Discrimination, Age – Related Bone Loss and Applicability of the WHO Criteria. *Osteoporos Int* 1999; 10: 441 – 449.
9. Hadji P, Hars O, Bock K, Albert US, Beckmann MW, Emons G, Schulz KD. Age Changes of Calcaneal Ultrasonometry in Healthy German Women. *Calcif Tissue Int* 1999; 65:117 – 120.

MEVSİMSEL İNSİDANS HIZI FARKLARININ ANLAMLILIK TESTİ VE BİR UYGULAMA

Ahmet Dirican, Ömer Uysal, Sevim Purisa, Ayşegül Bostan,
Mustafa Şenocak

İstanbul Üniversitesi, Cerrahpaşa Tıp Fakültesi, Biyoistatistik BD.

e-mail: adirican@istanbul.edu.tr

Özet

Epidemiyolojik çalışmalarda söz konusu hastalığın insidans hızının mevsimlik veya daha farklı periyodik dönemlere ait eğilimlerini anlamaya gerek duyulur. Çalışmanın verileri uygulama bireyleri üzerinden toplanırken, belli periyotlara ait özel kategoriler olarak bireyin ölüm, doğum, nüks veya hastaneye yatışları kaydedilir.

Çalışmanın başlıca amacı bir toplumda ele alınan hastalığın potansiyel eğilimlerini tanımlamaktır. Böylece hastalığın patofizyolojik mekanizmalarının altında yatan temel nedenler ortaya çıkarılır ve diğer önemli çevresel faktörler tanımlanarak, özellikle kronik hastalıkların nüks veya mortalite risklerinin hangi zamansal kesitte yükseleceği tahmin edilir.

Sonuçta hastalıkların mevsimler veya daha küçük periyotlardaki epidemiyolojik verilerin biyoistatistiksel anlamlılığı Roger's testi ile denetlenerek uygulama alanında hastalığın risk dönemlerine uygun ve yeterli önlemleri almamıza olanak sağlar.

Tıbbi önlemlerin geliştirilmesine ve bilimsel uygulayıcıların eğitim stratejilerinin belirlenmesine de temel oluşturan bu çalışma bir klinik uygulama örneği üzerinde ayrıntılı olarak irdelenmiştir.

Anahtar kelimeler: Mortalite; İnsidans hızı; Epidemiyolojik veriler; Mevsimler.

SAĞLIK PERSONELİNİN EMPATİK EĞİLİMLERİNİN VE MESLEKİ DOYUMLARININ ÇOK DEĞİŞKENLİ KOVARYANS ANALİZİ (MANCOVA) İLE DEĞERLENDİRİLMESİ

Semra Akgöz¹, İsmet Kan¹, Alis Özçakır², Yusuf Sivrioğlu³,
Kamuran Tombul⁴, Yıldız Altınsoy⁴

Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı¹

Uludağ Üniversitesi Tıp Fakültesi Aile Hekimliği Anabilim Dalı²

Uludağ Üniversitesi Tıp Fakültesi Psikiyatri Anabilim Dalı³

Uludağ Üniversitesi Tıp Fakültesi Sağlık Uygulama ve Araştırma Merkezi⁴

Özet

Kaliteli sağlık hizmetlerinin yürütülmesinde empatik iletişim çok önemlidir. Sağlık personelinin, hizmetlerin sunumunda daha etkin ve verimli çalışabilmeleri için ise, mesleki doyumunu önem taşımaktadır. Bu çalışma, Uludağ Üniversitesi Sağlık Uygulama ve Araştırma Merkezi'nde çalışan sağlık personelinin mesleki doyum ve empatik eğilim düzeylerinin demografik özellikleri, meslekleri ve kendileri ile ilgili görüşleri açısından farklılık gösterip göstermediğini saptamak amacıyla yapılmıştır.

Tanımlayıcı nitelikteki bu çalışmanın örneklemini, Uludağ Üniversitesi Sağlık, Uygulama ve Araştırma Merkezi'nde (SUAM) en az 6 aydır uzman veya araştırma görevlisi olarak çalışan 304 hekim ve 479 hemşire oluşturmuştur. Araştırmaya katılma oranı % 82,4'dir (N=645).

Araştırmada, Musal ve arkadaşlarının benzer konularda yapılan araştırmalardan ve Minnesota İş Doyum Ölçeğinden yararlanarak oluşturdukları 22 sorudan oluşan Mesleki Doyum Ölçeği ile Dökmen tarafından geliştirilmiş, kişilerin günlük yaşamlarında empati kurma potansiyellerini ölçmeyi amaçlayan 20 sorudan oluşan Empatik Eğilim Ölçeği kullanılmıştır. Beşli Likert tipindeki ölçeklerin güvenilirlik analizi sonucunda Cronbach'ın Alfa Katsayıları sırasıyla 0.9199 ve 0.8920 ($p < 0.0001$) olarak bulunmuştur.

Empatik eğilim ölçeği için ortalama skor 71.02 ± 0.36 , mesleki doyum için 4.22 ± 0.06 'dır. Empatik eğilim ve mesleki doyum arasında pozitif, anlamlı korelasyon ($r = 0.298$), yaş ile empatik eğilim ($r = 0.163$) ve mesleki doyum ($r = 0.283$) arasında da mevcuttur ($p < 0.001$).

Çok değişkenli kovaryans analizine dahil edilen bağımsız değişkenler temelinde, (mesleki bilgi ve becerisini kullandığını düşünme, geçmişe dönebilse mesleğini yeniden tercih edecek olma ve çevresindekilere duygularını rahatlıkla ifade edebilme) 2x2x2 faktöriyel dizaynda 8 grup oluşmuştur ve grupların her birisi önerilen minimum örneklem büyüklüğüne sahiptir ($n \geq 20$). Normallik analizinden sonra her bir bağımlı değişkenin hata varyansının gruplarda eşit olduğu görülmüştür (Levene test $p > 0.05$). Box's M test p-değeri 0.554'dür, böylece varyans-kovaryans matrislerinin homojenliği varsayımı karşılanmıştır. Bağımlı değişkenlerin korelasyonu için Bartlett küresellik testi anlamlı bir p-değerine (< 0.001) sahiptir. Pillai's Trace, Wilk's lambda ve Roy's largest Root istatistikleri etkileşimlerin etkisinin anlamlı olmadığını ($p > 0.05$) gösterirken, tüm bağımsız değişkenler ve kovaryet değişken "yaş" için anlamlı sonuçlar vermiştir ($p < 0.001$). "Mesleki bilgi ve becerisini kullandığını düşünme" empatik eğilim ve mesleki doyum için ($p < 0.01$), "çevresindekilere duygularını rahatlıkla ifade edebilme" empatik eğilim için ($p < 0.001$), "geçmişe dönebilme imkanı olsa yeniden mesleğini tercih edecek olma" mesleki doyum için ($p < 0.001$) gruplar arasında anlamlı farklılık göstermiştir. Kovaryet değişken "yaş" ile bağımlı değişkenler arasında anlamlı ilişki saptanmıştır ($p < 0.001$).

Çalışmamızda sağlık personelinin mesleki doyumunun ve empatik eğiliminin yaş ile birlikte arttığı görülmüştür. Geçmişe dönebilse mesleğini yeniden tercih edecek olanların empatik eğilimleri, mesleki bilgi ve becerilerini kullanabildiğini düşünenlerin ise hem mesleki doyumları hem de empatik eğilimleri anlamlı olarak daha yüksektir. Çevresindekilere duygularını güçlkle ifade edenler ise anlamlı olarak daha düşük empatik eğilim düzeylerine sahiptir.

DÖRT FARKLI NORMALLİK TESTİNİN TIP I HATA VE GÜÇ BAKIMINDAN DEĞİŞİK DAĞILIMLARDA İNCELENMESİ

Derya Öztuna, Atilla Halil Elhan, Ersöz Tüccar

Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı
[e-mail:dgokmen2001@yahoo.com](mailto:dgokmen2001@yahoo.com)

Özet

Sürekli bir verinin normal dağılıp dağılmadığını araştırmak amacıyla geliştirilmiş birçok yöntem vardır. Bunlardan bir bölümü **grafik**, bir bölümü de **test** yöntemleridir. Ancak, grafik yöntemleri dağılım yapısı hakkında bilgi sahibi olmamıza olanak sağlarken; dağılımın normal olduğunu garanti etmemektedir. Bu nedenle verinin dağılımının normal dağılıma (çan eğrisi) uygun olup olmadığına karar vermek için en iyi yol, grafikleri uygun test yöntemleri ile birlikte incelemektir.

Dağılımların normal dağılıp dağılmadığını test etmek amacıyla kullanılan en yaygın test yöntemleri; Ki-kare uyum iyiliği testi, Kolmogorov Smirnov (K-S) D testi, Lilliefors düzeltmeli Kolmogorov Smirnov (K-S) testi, Anderson-Darling testi, Cramer-von Mises testi, Shapiro-Wilk W testi, D'Agostino çarpıklık testi, Anscombe-Glynn basıklık testi, D'Agostino-Pearson K^2 testi ve Jarqua-Bera LM testidir. Çalışmamızda, Tip I hata ve güç bakımından Lilliefors düzeltmeli Kolmogorov Smirnov (K-S) testi, Shapiro-Wilk W testi, D'Agostino-Pearson K^2 testi ve Jarqua-Bera LM testinin performansları karşılaştırılmıştır. Sonuçlar genel olarak incelendiğinde, normal dağılıma sahip örneklerde Jarqua-Bera LM testinin, normal dağılıma sahip olmayan örneklerde ise Shapiro-Wilk W testinin diğer testlere göre daha iyi performansa sahip oldukları belirlenmiştir.

Anahtar Kelimeler: Lilliefors ve Kolmogorov-Smirnov testi, Shapiro-Wilk testi, D'Agostino-Pearson testi, Jarqua-Bera testi, Tip I hata, Güç.

Kaynaklar

1. D'Agostino R.B., Belanger A., D'Agostino Jr. (1990). *A Suggestion for Using Powerful and Informative Tests of Normality*. The American Statistician, 44 (4), 316-321.

2. Mendes M. ve Pala A. (2003). *Type I Error Rate and Power of Three Normality Tests*. Pakistan Journal of Information and Technology, 2 (2), 135-139.
3. Jarqua C.M. ve Bera A.K. (1987). *A test for normality of observations and regression residuals*. International Statistical Review, 55(5):163-172.

ÜLKEMİZDE AĞRI ve KAPI İĞNE ZAMANINI, ULUSLARARASI KLAVUZLARIN ÖNERDİĞİ SÜREYE UYGUN MUDUR?

Hatice Bulut¹, Gülşah Seydaoğlu², Fevziye Türemez²

¹ Çukurova Üniversitesi Tıp Fakültesi Kardiyoloji AD

² Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik AD

e-mail : türemezf@yahoo.com

Özet

Amaç: Uluslararası kılavuzlar akut miyokard infarktüsü (AMI) geçiren hastalarda, hastanın acile girişinden iğnenin yapıldığı ana kadar geçirilen süreyi 30 dakika, ağrının başladığı andan iğnenin yapıldığı ana kadar geçen süreyi de 60 dakika olarak belirlemiştir. Bu çalışmada bölgemizde, AMİ geçiren hastalarda trombolitik tedavi uygulamasına kadar geçen sürenin belirlenmesi, bu sürenin mortalite ile ilişkisi ve tedavi gecikmesini etkileyen faktörler araştırıldı.

Yöntem: Bu çalışmada Ocak 2003 ile Aralık 2003 arasında Çukurova Üniversitesi acil ünitesine başvuran ve AMİ olduğu saptanan 107 hasta bir ay süresince izlenmiştir. Hastanın acile girişinden iğnenin yapıldığı ana kadar geçirilen süre; kapı iğne zamanı-KİZ, ağrının başladığı andan iğnenin yapıldığı ana kadar geçen süre; ağrı iğne zamanı-AİZ olarak tanımlanmıştır.

Bulgular: Çalışmaya 83 erkek 24 kadın olmak üzere 107 hasta (yaş ortalaması 55.3±10.1) alınmıştır. Tüm hastalarda ağrının başlaması ile acil servisine başvuru arasında geçen sürenin ortalaması (başvuru gecikmesi) 223.3±175.1 dk (medyan 180 dk), acilde gecikme süresi ortalama 26.2±13.9 dk (medyan 25 dk), koroner yoğun bakım ünitesinde (KYBÜ) gecikme süresinin ortalaması ise 16.4±11.7 dk (medyan 10 dk) olarak belirlenmiştir. Kapı iğne zamanı 42.7±17.2 dk (medyan 42 dk), ağrı iğne zamanı ortalaması 266.1±177.6 dk (medyan 235 dk) olarak saptanmıştır. KİZ, 36 hastada (% 33.6) 30 dakikanın altında iken AİZ sadece 4 hastada (% 3.7) 60 dakikanın altındadır. Cinsiyet, yaş grubu, yaşadığı yer, ağrının başlama saati, daha önce bir sağlık kurumuna başvurma faktörlerinin AİZ için risk faktörü olduğu saptanmıştır. Tedavi sonrasında % 86.9'unda patolojik Q, % 57.0'ında % 50 ST çökmesi saptanan hastaların 3'ü 28 günden önce, 4'ü ise 28 gün

sonrasında kaybedilmiştir. Gecikme sürelerinin, ölen hastalarda sağ kalanlara oranla anlamlı olarak daha uzun olduğu belirlenmiştir.

Sonuç: Kapı-ağrı iğne zamanlarındaki gecikmenin AMİ hastalarının mortalitesini etkilediği gözlenmiştir. Bölgemizde elde edilen kapı-ağrı iğne zamanların uluslararası kılavuzların önerdiği (KİZ; 30 dk ve AİZ; 60 dk) sürelerinden uzun olduğu belirlenmiştir. Bunu etkileyen önemli faktörlerin hastaların geç başvurusu ve acil ünitesindeki gecikme olduğu düşünülmektedir.

GENELLEŞTİRİLMİŞ TAHMİN DENKLEMLERİ: GTD

Seval Kul, E. Arzu Kanık

Mersin Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı

e-mail: sevalkul@mersin.edu.tr

Özet

Bu çalışmada genelleştirilmiş tahmin denklemlerinin kullanım alanları, quasi-likelihood fonksiyonunun elde edilişi, bağlantı fonksiyonlarının ve üzerinde durulan ilişki matrislerinden en uygun olanın tercih edilmesi ve genelleştirilmiş tahmin denklemlerinde karşılaşılan over dispersion gibi bazı sorunlar tartışılmıştır. Örnek bir veri seti üzerinde GTD yönteminin uygulanması R paket programı üzerinde gösterilmiştir.

Genelleştirilmiş tahmin denklemlerinde anahtar karakteristik aynı deneklerden elde edilen ölçümler ilişkilidir ve dolayısıyla bağımsızlık varsayımı altında çalışılan birçok istatistiksel yöntemin kullanılması mümkün olmayacaktır. Bu ilişkinin görmemezlikten gelinmesi ve standart istatistik yöntemlerinin kullanılması yanlışlığı ortaya çıkararak yetersiz tahminlere neden olabilecektir. Standart tekrarlanan ölçümlü çalışmalarda ise bireylerin birbirinden bağımsız olduğu varsayılır. Fakat bu varsayım ölçümlerin aynı kümedeki deneklerden elde edildiği birçok klinik ve epidemiyolojik çalışmada bozulur. Kümeler; aileler, kardeşler, ikizler, eşler, sınıflar ve aynı hasta odasında kalan hastalar gibi çok çeşitli olabilir. Bu tür özelliklere sahip çalışmalarda GTD kullanılabilir alternatif bir yöntemidir.

Genelleştirilmiş tahmin denklemlerinde katsayıların tahminleri quasi-likelihood fonksiyonundan yararlanılarak elde edilir.

$$\sum_{i=1}^K \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{S}_i = \mathbf{0}$$

Quasi-likelihood fonksiyonunun bileşenlerine göz atacak olursak;

$\mathbf{S}_i = y_i - \mu_i$; Her bireyin tekrarlanan ölçümlerinin kendi ortalamalarından farkıdır.

\mathbf{D}_i : Bağlantı fonksiyonunun modelde tahmin edilecek her bir parametre için kısmi türevinin alınması ile elde edilir.

Varyans sandviç tahmini (sandwich estimation) adı verilen V_i ile tahmin edilir.

$$V_i = A_i^{1/2} R_i(\alpha) A_i^{1/2}$$

A_i , $n_i \times n_i$ boyutlu tekrarlanan ölçümlere ait Varyans kovaryans matrisinin köşegenlerinden elde edilen ve varyanslardan oluşan bir köşegen matrisi, $R_i(\alpha)$ üzerinde durulan ilişki matrisidir. Aynı küme içindeki gözlemler arasındaki ilişki $R_i(\alpha)$ $n_i \times n_i$ boyutlu simetrik bir matris ile ifade edilir. Ve bu matrise “Üzerinde Çalışılan İlişki Matrisi” adı verilir. Genelleştirilmiş tahmin denklemlerinde birçok ilişki matrisinden söz edilebilir. Bunlardan bazıları; Yapılandırılmamış (Unstructured) korelasyon, Bağımsızlık (Independence) korelasyon yapısı, Değişebilir (Exchangeable) korelasyon yapısı, Otoregresif (Auto-regressive) korelasyon yapısı, Sabit (Fixed) korelasyon yapısı, M-bağımlı (M-dependent) korelasyon yapısıdır.

BİR KENTTE 1995-2005 YILLARI ARASINDA TUTULAN 63.831 ÖLÜM SERTİFİKASI BİZE NE ANLATIYOR?

Gülşah Seydaoğlu, Fevziye Türemez, Nilgün Arpacı, Nazan Alparslan

Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik AD

ÖZET

Bu çalışmanın amacı Adana ilinde toplanan ölüm sertifikalarının değerlendirilmesi ve kalite kontrolünün yapılmasıdır.

Yöntem

Çalışmada 1995-2002 yılları arasında toplanan ve bilgisayar ortamına ölüm sebebi kodlanmamış olarak girilmiş olan 47.665 ölüm sertifikası 150 başlıklı hastalık kodlamasına göre kodlandı. 2003-2005 (ilk 6 ay) yıllarında kodlanarak girilmiş 16.241 ölüm sertifikası ile birleştirilerek değerlendirildi.

Bulgular

Tüm ölümlerin % 58.1'i erkek, % 41.9'u kadındır, cinsiyet bilgisi olmayan form sayısı sadece 48'dir. Yaş bilgisi olmayan form sayısı 151 (% 0.2), medeni durum bilgisi olmayan 3733 (% 5.8), öğrenim durumu bilgisi olmayan 3919 (% 6.1) meslek bilgisi olmayan 4192'dir (% 6.6). Yerleşim yeri % 68,8'i il merkezi, % 19.8'i ilçe merkezi, % 10.8'i köy olan ölümlerin 407'sinin (% 0.6) ikametgah bilgisi eksiktir. Vakaların sadece 3853'üne (% 6.0) otopsi yapılmıştır. Ölüm sebeplerine göre yapılan sınıflamada 13483 (% 21.1) formda uygun olmayan ölüm sebebi (semptom, son veya ara neden gibi) veya okunaksız açıklamalar yazıldığı saptanmıştır. Bu uygunsuz formların % 24.4'ünde solunum yetmezliği, % 16.2'sinde kardiyak arrest, % 13.9'unda sepsis, % 11.1'inde beyin kanaması, % 7.9'unda böbrek yetmezliği gibi temel neden olamayacak açıklamalar yazmaktadır. Bunların dışında kalan nedenler içinde en sık ölüm nedeni kalp hastalıklarına bağlı ölüm (% 30.6) olarak görülmektedir. Fakat bu sınıflamada da şüpheler vardır, ölüm nedeni olarak yalnızca "kalp yetmezliği" yazılmış % 11.4 formda temel neden kalp hastalığı mıdır yoksa son neden midir belli değildir. Buna rağmen ölüm verileri yıllara göre nüfusa oranlanarak çeşitli ölüm hızları

saptanmaya çalışılmış (yaşa, cinsiyete, hastalıklara göre) ve DİE ve sağlık bakanlığı verileri ile karşılaştırılmıştır.

Sonuç

Bilgisayar teknolojisinin yaygınlaştığı, hekim sayısının arttığı, öğretim elemanları sayısının yeterli olduğu ülkemizde hala ölümler ve hastalıklar ile ilgili tutulan kayıtlarda önemli eksiklik ve yanlışlıklar olduğu görülmektedir. Bu konuda ülke çapında hekim eğitimi seferberliği oluşturulmasında biyoistatistik uzmanlarının görev ve sorumluluğunun fazla olduğu inancındayız.

FAKTÖR ANALİZİ MODELİNDE ORTAK VARYANSLARI BELİRLEMEDE GENELLEŞTİRİLMİŞ TERSLERİN KULLANIMI

Seval Süzölmüş¹, Sadullah Sakallıođlu²

*Çukurova Üniversitesi, Osmaniye Meslek Yüksekokulu OSMANİYE¹
Çukurova Üniversitesi Fen-Ed. Fak. İstatistik Bölümü Balcalı-ADANA²
e-mail:suzulmus@mail.cu.edu.tr*

Özet

Bu çalışmada, faktör analizi modelinde ortak varyansların tahmini üzerinde çalışıldı. Daha önceki çalışmalarda varyans-kovaryans matrisi Σ 'nın pozitif tanımlı olduđu kabul edilmiştir. Σ 'nın non-negatif tanımlı olduđu kabul edilerek genelleştirilmiş ters kullanımı içeren az sayıda çalışma bulunmaktadır. Bu çalışmada Σ 'nın non-negatif tanımlı olduđu kabul edilerek genelleştirilmiş ters kullanımı ele alınıp faktör analizi modelindeki ortak varyansların tahmini incelenmiştir.

Anahtar Kelimeler: Faktör Analizi, Ortak Varyans, Genelleştirilmiş Ters.

FUZEL YAĞINDAN İZOAMİL ASETATIN SÜREKLİ ÜRETİMİNDE OPTİMUM KOŞULLARIN BELİRLENMESİ İÇİN İKİ YÖNTEMİN BİR ARADA KULLANILMASI: YANIT YÜZEYLERİ VE YAPAY SINIR AĞLARI

Ömer Utku Erzengin¹, Havva Elemzade², Gamze (Yalçın) Erzengin², Afife Güvenç²

*Marmara Üniversitesi, Tıp Fakültesi, Biyoistatistik ABD¹
Ankara Üniversitesi, Mühendislik Fakültesi, Kimya Mühendisliği Bölümü²*

Biyoteknoloji ve eczacılıkta kimyasal üretim aşamalarında kullanılan optimizasyon yöntemlerinden biri yanıt yüzeyleridir. Kimyasal deneyler pahalı ve zaman alıcı olduğundan, önceden belirlenmiş değişkenler ve noktalarda veri elde edip modellemelerin yapılmasıyla en uygun koşulların belirlenmesi amaçlanmaktadır. Elde edilen modeller ışığında, deneyleri yapılmayan nokta-lar ve aralıklar için üretim aşamasında tahminlerde bulunmaktadır.

Fuzel yağı etil alkol üretiminde elde edilen bir yan üründür. Dünyada özellikle etil alkolün büyük ölçeklerde üretildiği ülkelerde yan ürün probleminin önemli bir kısmını oluşturmaktadır. Bazı ülkelerde endüstrinin enerji ihtiyacının bir kısmını karşılamak için yakılmaktadır. Türkiye’de ise, etkin bir şekilde değerlendirilememektedir.

Fuzel yağının değerlendirilmesi kapsamında yapılan kaynak araştırması sonucu, bu maddeden gıda, farmasötik ve kozmetik endüstrilerinde önemli tat ve koku bileşenlerinin üretildiği görülmüştür. İlaçlarda (şurup, tatlandırıcı vb) kullanılan çilek ve muz aromaları olan etil bütirat ve izoamil asetat sözü edilen tat ve koku bileşenlerine örnek olarak verilebilir.

Fuzel yağı başlıca izoamil alkol, propil alkol, izobütil alkol, etil alkol ve n-bütil alkol gibi düşük molekül ağırlıklı alkoller ve az miktarda sudan oluşmaktadır. Fuzel yağından izoamil asetatın (y) üretimi için sıcaklık (x_1), basınç (x_2), substrat / CO₂ miktarı (x_3), besleme akış hızı (x_4) değişkenleri altında denemeler yapılmıştır. Elde edilen veri yanıt yüzeyi (response surface methodology) yöntemiyle modellenmektedir.

Yanıt yüzeyleri metodolojisi (Response Surface Methodology, RSM) iki ve/veya daha fazla değişken yardımıyla tanımlanan yüzeylerin matematiksel ve istatistiksel tekniklerin yardımı sayesinde modellenmesidir.

Yapay sinir ağırları birçok modellemeye alternatif olarak ortaya atılmış bir modelleme tekniğidir. Yapay sinir ağırları (Artificial Neural Network, ANN) model olarak beyindeki sinir topluğunun bilgiyi iletme prensibini esas alır.

Yanıt yüzeylerine karşılık gelen yapay sinir ağırlarının kullanımı literatürde de yenidir. Vaithiyalingam ve arkadaşları 2002 (kasım ayı) yılında selüloz asetat bütirat yayılımı üzerine, Dutta ve arkadaşları 2004 yılında Pseudomonas sp izolasyonu üzerine, Sun ve arkadaşları 2003 yılında kontrollü ilaç salınım sistemleri üzerine, Takayama ve arkadaşları 2004 yılında ilaç formüllerinin optimizasyonu üzerine yanıt yüzeyleri ve yapay sinir ağırları ile karşılaştırmalı makaleler yayınlamışlardır.

Yapılacak sunumda fuzel yağından izoamil asetatın sürekli üretiminde optimum koşulların belirlenmesi için yanıt yüzeyleri ve yapay sinir ağırları bir arada kullanılacak ve iki yöntem birbiri ile karşılaştırılacaktır.

Kaynaklar

1. Anderson M ve arkadaşları. Design-Expert. Software Version 6 User's Guide. Minneapolis: Stat-Ease, Inc; 2003.
2. Christopher D, Himmel and Gary S. Advantages of Plasma Etch Modeling Using Neural Networks Over Statistical Techniques. Ieee Transactions On Semiconductor Manufacturing 1993; 6: 2.
3. Cundy AL. Use of Response Surface Metamodels in Damage Identification of Dynamic Structures (Yüksek Lisans Tezi). Virginia: Virginia Polytechnic Institute and State University 2003.
4. Dutta JR, Dutta PK, Banerjee R. Optimization of culture parameters for extracellular protease production from a newly isolated Pseudomonas sp. using response surface and artificial neural network models. Process Biochemistry 2004; 39:2193-2198.
5. Demuth H, Beale M. Neural Network Toolbox For Use with MATLAB. Eighth printing. Apple Hill Drive: The MathWorks, Inc; 2004.
6. Papila NU, Neural Network And Polynomial-Based Response Surface Techniques For Supersonic Turbine Design Optimization (Doktora Tezi). Florida: University Of Florida 2001.
7. Sarle WS. Neural Networks and Statistical Models. Proceedings of the Nineteenth Annual SAS Users Group International Conference. 1994.
8. Seung S. Introduction to Neural Networks (Massachusetts Institute of Technology Elektronik ortamda ders notları) <http://hebb.mit.edu/>; 2002.
9. Subramanian N, Yajnik A, Murthy RSRamachandra. Artificial Neural Network as an Alternative to Multiple Regression Analysis in Optimizing

Formulation Parameters of Cytarabine Liposomes. AAPS Pharm Sci Tech 2004; 5.

10. Sun Y, Peng Y, Chen Y, Shukla AJ. Application of artificial neural networks in the design controlled release drug delivery systems. *Advanced Drug Delivery Reviews* 2003; 55: 1201–1215.
11. Takayama K, Fujikawa M, Obata Y, Morishita M. Neural network based optimization of drug formulations. *Advanced Drug Delivery Reviews* 2003; 55:1217-1231.
12. Vaidyanathan R, Papila Ni, Shyy W, Tucker PK, Griffin(2) LW. Neural Network And Response Surface Methodology For Rocket Engine Component Optimization. *AIAA* 2000; 4880.
13. Vaithiyalingam S, Khan MA. Optimization and characterization of controlled release multi-particulate beads formulated with a customized cellulose acetate butyrate dispersion. *International Journal of Pharmaceutics* 2002; 234:179-193.

TÜRKİYE'DE İL VE İLÇE MERKEZLERİ ÖLÜM İSTATİSTİKLERİYLE MORTALİTENİN YAŞA, CİNSİYETE VE ZAMANA GÖRE DEĞİŞİMİ (1990-2003)

R. Erol Sezer¹, Hafize Sezer²

Cumhuriyet Üniversitesi Tıp Fakültesi Aile Hekimliği Anabilim Dalı¹

Cumhuriyet Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı²

Giriş

1963'den beri düzenli aralıklarla yapılan nüfus ve sağlık araştırmaları, bebek, 1-4 yaş ve beş yaş altı ölüm hızları ile bu hızların değişim özelliklerinin bilinmesine olanak sağlamıştır (1, 2). Türkiye'nin 1963'de binde 208 olarak tahmin edilen bebek ölüm hızı, 1966-67 döneminde binde 170, 1973 de binde 100-150 olarak tahmin edilmiştir (1). Beş yaş altı ölüm hızı, 1973 araştırmasında(1) binde 218 olarak bildirilmiştir. Her iki hız da zaman içinde sürekli düşmüştür. 1988 araştırmasında(3) bebek ölüm hızı binde 82, beş yaş altı ölüm hızı binde 98 olarak bildirilirken, 2003 araştırmasında bu hızlar sırayla, binde 29 ve binde 37 olarak bulunmuştur (2). Bu araştırmalar, bebek, 1-4 yaş ve beş yaş altı ölüm hızlarının bölgelere, kent-kır olma durumuna ve bazı demografik değişkenlere göre değişimini inceleme olanağı vermektedir. Özetle Türkiye'de beş yaş altı dönemle ilgili bebek ve çocuk ölümlülüğü konusunda güvenilir bilgiler son kırk yıl için vardır. Fakat 5 yaş üstü nüfusun ölümlülüğü konusundaki bilgiler çok azdır. Nüfus Etütleri Enstitüsü tarafından beş yılda bir yapılan nüfus ve sağlık araştırmaları beş yaş üstü nüfusun ölümlülüğünü inceleme kapsamı dışında tutmaktadır. Devlet İstatistik Enstitüsü'nün (DİE) 1989 Türkiye Nüfus Araştırması (1989 TNA) kent-kır ayırımı olacak şekilde bölgelere göre kaba ölüm hızlarıyla son bir yıl için yaşa ve cinsiyete özgü ölüm hızlarını tahmin etmiştir (4). İzleyen dönemde Türkiye'nin beş yaş üstü nüfusunun ölümlülüğüyle ilgili yayınlanmış bir araştırmaya ulaşamamıştır. Türkiye'de ölüm istatistikleri, il ve ilçe merkezlerini kapsayacak şekilde Devlet İstatistik Enstitüsü tarafından yayınlanmaktadır. Bu istatistiklerin gerek nitelik gerekse nicelik yönünden sorunlu olduğu belirtilmiştir (5, 6).

Koroner kalp hastalığı, serebrovasküler hastalık ve kanser, dünyada ve Türkiye'de 35 yaş üstü nüfusun en önemli üç ölüm nedenidir. Türkiye'de 1990'lı yıllarda yapılan bir izleme çalışmasına göre 20 yaş üzeri ölümlerin, % 42.5'i koroner kalp hastalığı, % 25'i kanser, % 16.2'i serebrovasküler hastalık nedeniyledir (7). Yirminci yüzyılın ikinci yarısı, bu hastalıkların önlenmesi ve tedavisi konusunda önemli keşiflere sahne olmuştur. Gün-

müzde bu hastalıklar artık önlenabilir hastalıklar arasındadır. Kardiyovasküler hastalığın tedavisi konusunda yakın geçmişte ölüm oranlarını önemli ölçüde düşürücü gelişmeler yaşanmıştır. Koroner kalp hastalığı ve serebrovasküler hastalık mortalite hızları gelişmiş ülkelerde hızla düşmektedir (8). Sigarayla mücadele programları etkili olan ülkelerde akciğer kanseri ölüm hızları da azalmaya başlamıştır (9).

Bu çalışma şu sorulara yanıt aramak amacıyla yapılmıştır: Türkiye’de mortalitenin durumu, değişim yönü ve hızı nedir? Türkiye’nin il ve ilçe merkezleri ölüm istatistikleri bu sorunun yanıtına katkı sağlayabilir mi?

Yöntem

İnceleme dönemi olarak 1990-2003 yılları seçilmiştir. Bunun nedenlerinden birisi ölümlerin yaşa ve cinsiyete göre dağılımını içeren en son yapılmış ulusal araştırmanın 1989 TNA olması, diğeri ise DİE’nin en yeni ölüm verilerinin 2003 yılına ait olmasıdır.

İnceleme döneminin ölümle ilgili verileri, DİE tarafından her yıl yayınlanmakta olan ve il ve ilçe merkezlerinde olan ölümleri kapsayan ölüm istatistiklerinden sağlanmıştır (10). DİE bu bilgileri “Ölü gömme izin kağıdının” bir parçası olan ve sağlık müdürlükleri tarafından DİE’ye gönderilen “Ölüm istatistik formları’ndan” sağlamaktadır.

İnceleme dönemi ölümlüğünün incelenmesinde yaşa ve cinsiyete özgü hızlar ile yaşa ve cinsiyete göre direkt standardize ölüm hızları kullanılmıştır. Direkt standardizasyonda standart nüfus olarak, kurgu bir nüfus olan Avrupa standart nüfusu alınmıştır. Yaşa ve cinsiyete özgü ölüm hızlarının hesabında yıl ortası nüfus olarak ilgili yılların tahmini yıl ortası nüfusları kullanılmıştır. 1990 nüfus sayımı sonuçlarına göre nüfusun yaş ve cinsiyet dağılımı (11), 1985-90 dönemi için il ve ilçe merkezleri nüfus artış oranı, 2000 yılı nüfus sayımına göre il ve ilçe merkezleri nüfusu, 1990-2000 nüfus sayımları arasında il ve ilçe merkezleri nüfusunun artış oranı, 2003 Türkiye Nüfus ve Sağlık Araştırması’nın nüfusun cinsiyete ve yaş gruplarına göre yüzde dağılımı, yıl ortası nüfusların hesaplanmasında kullanılmıştır. Yıl ortası nüfus hesabı, $P = P_0 * e^{hz}$ formülü ile yapılmıştır. İl ve ilçe merkezleri nüfus artış oranları 1985- 1990 dönemi için binde 33.83, 1990-2000 dönemi için binde 26,81 olarak DİE tarafından hesaplanmıştır. Türkiye’nin 2003 yılı hızları, gelişmiş ülke örneği olarak İngiltere’nin hızlarıyla karşılaştırılmıştır. Çünkü bu ülkenin ölüm istatistikleri güvenilir ve internet ile ulaşılabilir özelliktedir.

Bulgular

1-4 no’lu tablolar, il ve ilçe merkezlerinin 1990 ve 2003 yıllarına ait, yaşa ve cinsiyete özgü ölüm hızları ile, direkt standardize ölüm hızlarını göstermektedir. Bu tablolar, yıl ortası nüfusun yaş ve cinsiyete göre dağılımıyla, ölümlerin yaş ve cinsiyete göre dağılımını da içermektedir.

Bildirimlerin yeterliliği: Tablo 1 ve Tablo 2'deki veriler birlikte kullanıldığında, 1990 yılı için il ve ilçe merkezlerinin kaba ölüm hızı binde 4.56 olarak bulunur. 1989 Türkiye Nüfus Araştırması'na (1989 TNA) göre Türkiye'nin Ekim 1989'dan önceki bir yıla ait kaba ölüm hızı binde 7.79, kentlerde (Nüfusu 20 bin ve üzerinde olan yerleşim birimlerinde) binde 7.07, nüfusu daha az olan yerleşim birimlerinde (kır) binde 8.48 olarak bulunmuştur. Bu çalışmada bulunan binde 4.56'lık kaba ölüm hızı, 1989 TNA kaba ölüm hızı olan binde 7.07'den % 35 daha düşüktür.

Yaşa, cinsiyete ve zamana göre ölüm hızlarının değişimi: Ölüm hızları her iki cinsiyette de yaşla önemli düzeyde değişmektedir. Ölümlülüğün en yüksek olduğu dönemler, sırayla; 75 yaş ve üstü dönem, 65-74 yaş dönemi, bebeklik dönemi ve 45-64 yaş dönemidir. Ölümlülüğün en düşük olduğu yaş grupları da sırasıyla 5-14 ve 15-24 yaş gruplarıdır. Gerek 1990, gerekse 2003 yılı yaşa özgü ölüm hızları, her yaş grubu için kadınların ölüm riskinin erkeklerden düşük olduğunu göstermektedir. Tüm yaş gruplarını kapsayacak şekilde yaşa göre direkt standardize ölüm hızı 1990'da erkekler için binde 11.62, kadınlar için binde 8.70'dir. Aynı hızlar 2003'de erkekler için binde 7.97, kadınlar için binde 6.07'dir. Tüm yaş gruplarını kapsayan yaşa göre standardize hızlar için 1990 ile 2003 arasında gerçekleşen düşme düzeyi erkeklerde % 31.4, kadınlarda % 30.2'dir.

Ölümlülük, 35-64 yaş grubu için incelendiğinde; yaşa göre direkt standardize ölüm hızı, 1990'da erkekler için binde 8.70, kadınlar için binde 4.31; 2003'de erkekler için binde 6.07, kadınlar için binde 3.08'dir. Bu yaş aralığı için standardize ölüm hızlarının 1990 ve 2003 arasında gerçekleşen düşme düzeyleri, erkeklerde % 30.2, kadınlarda % 28.5'dir.

Tablo 1. Türkiyede il ve ilçe merkezlerinde erkekler için ölüm hızları, 1990

Yaş	Ölüm sayısı 1	Yıl ortası nüfus 2	Ölüm hızı (milyonda) 3 (1:2)	Avrupa stan- dard nüfusu 4	Beklenen ölüm sayısı 3*4
0	12154	321396	37816	1600	60.51
1-4	1895	1362571	1391	6400	8.90
5-14	1594	4082604	390	14000	5.47
15-24	2272	3664295	620	14000	8.68
25-34	3000	2908068	1032	14000	14.44
35-44	4637	2041983	2271	14000	31.79
45-54	8612	1208941	7124	14000	99.73
55-64	17221	911560	18892	11000	207.81
65-74	15402	354950	43392	7000	303.74
75-79	18497	96074	105233	2000	420.93
80-84		48359		1000	
85+		31339		1000	
Toplam	85284 ^a	17032140	Binde 5.01	100000	1162 ^b

^aYaşı belli olmayan 305 kişi, grup büyüklüğüyle orantılı biçimde, yaş gruplarına dağıtılmıştır.
^bKurgu Avrupa Standart Nüfusu, standart nüfus alındığında 1990'da Türkiye'de erkekler için il ve ilçe merkezleri direkt standardize ölüm oranı binde 11.62, 35-64 yaş için direkt standardize ölüm hızı: Binde 8.7'dir.

Tablo 2. Türkiye'de il ve ilçe merkezlerinde kadınlar için ölüm hızları, 1990

Yaş	Ölüm sayısı 1	Yıl ortası nüfus 2	Ölüm hızı (milyonda) 3 (1:2)	Avrupa stan- dart nüfusu 4	Beklenen ölüm sayısı 3*4
0	9789	306182	31971	1600	51.15
1-4	1596	1293779	1234	6400	7.90
5-14	1111	3760619	295	14000	4.12
15-24	1195	3213959	372	14000	5.21
25-34	1529	2719440	562	14000	7.87
35-44	2269	1871539	1212	14000	16.97
45-54	4007	1174815	3411	14000	47.75
55-64	8741	929391	9405	11000	103.46
65-74	10790	423393	25485	7000	178.30
75-79	23981	120693	94460	2000	377.84
80-84		78165		1000	
85+		55017		1000	
Toplam	65008 ^a	15946992	Binde 4.08	100000	800.57 ^b

^aYaşı belli olmayan 196 kişi, grup büyüklüğüyle orantılı biçimde, yaş gruplarına dağıtılmıştır.
^bKurgu Avrupa Standart Nüfusu, standart nüfus alındığında 1990'da Türkiye'de kadınlar için, il ve ilçe merkezleri direkt standardize ölüm oranı binde 8, 35-64 yaş için direkt standardize ölüm hızı: Binde 4.31'dir.

Tablo 3. Türkiye'de ölüm kayıtlarına göre il ve ilçe merkezlerinde erkekler için ölüm hızları, 2003

Yaş	Ölüm sayısı 1	Yıl ortası nüfus 2	Ölüm hızı (milyonda) 3 (1:2)	Avrupa stan- dart nüfusu 4	Beklenen ölüm sayısı 3*4
0	7503	455524	16471	1600	26.35
1-4	1387	1822095	761	6400	4.87
5-14	1160	4485514	259	14000	3.63
15-24	2344	4415792	531	14000	7.43
25-34	2803	3904490	718	14000	10.05
35-44	5018	3160778	1588	14000	22.23
45-54	11751	2393822	4909	14000	68.73
55-64	17860	1347979	13249	11000	145.74
65-74	27021	836674	32296	7000	226.07
75-79	26235	232409	70552	2000	281.81
80-84		139446		1000	
85+				1000	
Toplam	103082	23194523	Binde 4.4	100000	796.91

Kurgu Avrupa Standart Nüfusu, standart nüfus alındığında;
35-64 yaş erkekler için direkt standardize ölüm hızı: Binde 6.069

Tablo 4. Türkiyede il ve ilçe merkezlerinde kadınlar için ölüm hızları, 2003

Yaş	Ölüm sayısı 1	Yıl ortası nüfus 2	Ölüm hızı (milyonda) 3 (1:2)	Avrupa standart nüfusu 4	Beklenen ölüm sayısı 3*4
0	5375	410286	13101	1600	20.96
1-4	1041	1641145	634	6400	4.06
5-14	840	4561416	184	14000	2.58
15-24	1412	4633820	305	14000	4.27
25-34	1660	4199399	395	14000	5.53
35-44	2738	3354692	816	14000	11.42
45-54	5639	2485851	2268	14000	31.75
55-64	9630	1375665	7000	11000	77.00
65-74	19402	989514	19608	7000	137.26
75-79	33511	265478	69426	2000	277.70
80-84		217210		1000	
85+				1000	
Toplam	81248	24134476	Binde 3.4	100000	572.53

Kurgu Avrupa Standart Nüfusu, standart nüfus alındığında Türkiye'nin 35-64 yaş kadınlar için direkt standardize ölüm hızı: Binde 3,081

Tartışma

Bu çalışmada il ve ilçe merkezleri için binde 4.56 olarak bulunan 1990 yılı kaba ölüm hızı, 1989 TNA'sının, binde %7.09 olarak bulunduğu kentlere ait kaba ölüm hızından % 35 daha düşüktür. 1989 TNA, nüfusu 20 binden az olan yerleşim yerlerini kır olarak kabul etmiştir. Oysa nüfusu 20 binden az olan birçok ilçe merkezi vardır ve 1989 TNA kırsal kesim için kaba ölüm hızını binde 8.48 olarak bulmuştur. Bu nedenle bu çalışmada bulunan hızların en az % 35 düşük saptanmış olduğu düşünülebilir. 1989 TNA, bebek ölüm hızını kentler için binde 51.6 olarak bulmuştur. Bu çalışmanın il ve ilçe merkezleri için, her iki cinsiyet birlikte alındığında binde 34.96 olarak bulunduğu bebek ölüm hızı da, 1989 TNA değerinden % 32.2 oranında daha düşüktür. Bu çalışmada hızların düşük çıkmasında, ölüm istatistik formlarının eksik toplanması önemli bir etmen olmalıdır. Nüfus sayımlarında yerleşme yeri nüfusunu fazla gösterme gayretiyle yapılabilecek hileler ile diğer başka sorunlu durumlar da bu eksikliğe katkıda bulunmuş olabilir.

Türkiye il ve ilçe merkezleri için 2003 yılına ait erkeklerde binde 7,97 olarak bulunan yaşa göre standardize ölüm hızı, İngiltere'nin Avrupa standart nüfusu kullanılarak erkekler için hesaplanmış 2003 yılı yaşa göre standardize ölüm hızına (binde 7.99) benzer bir değerdir (12). Kadınlar için bu çalışmada bulunan 2003 yılı yaşa göre standardize ölüm hızı (binde 5,73) da İngiliz kadınlarının 2003 yılı yaşa göre standardize ölüm hızına (binde 5.58) benzemektedir. İngiltere'de, kalite kontrol mekanizmaları olan, tüm

ülkeyi kapsayan, mevzuatı ve örgütü olan, iyi işleyen bir ölüm kayıt sistemi vardır. Bu sistemin ayrıntılarına İngiltere'nin ulusal istatistikler internet sitesinin ilgili sayfalarından ulaşılabilir. İngiltere istatistiklerinin tam olduğu, Türkiye ölüm istatistikleriyle bulunan hızların ise gerçeği en az % 35 eksik yansıttığı düşünülürse Türkiye'de ölümlülüğün genel olarak İngiltere'den en az % 35 daha yüksek olduğu söylenebilir.

35-64 yaş aralığındaki ölümlülük bu çalışmanın odaklandığı konulardan biridir. Türkiye il ve ilçe merkezlerinin 35-64 yaş aralığındaki erkeklere ait yaşa göre standardize hızı bu çalışmada binde 6.07 olarak bulunmuştur. Oysa bu değer İngiltere'de 2003 yılında binde 4.63'tür. Türkiye'nin bu değeri, İngiltere'nin ilgili değerinden % 31 daha fazladır. Türkiye değerini binde 6.07 olarak değil de bunun % 35 fazlası olarak düşünürsek, Türkiye değeri İngiltere değerinden % 77 daha yüksek olur. Bu durum, 35-64 yaş aralığındaki Türk erkeklerinde ölümlülüğün yüksek olduğunu göstermekte, kronik hastalıkların önlenmesi, erken teşhisi ve tedavisine yönelik ulusal kontrol programlarına önem verilmesini gerektirmektedir.

Türkiye il ve ilçe merkezlerinin 35-64 yaş aralığındaki kadınlara ait yaşa göre standardize ölüm hızı bu çalışmada binde 3.08 olarak bulunmuştur. Bu değer İngiltere'de 2003 yılında binde 2.91'dir. Türkiye'nin bu değeri, İngiltere'nin bu değerinden % 6 kadar daha yüksektir. Türkiye'de kadınların sigara içme oranlarında geçen on yıllık dönemde ortaya çıkan hızlı artış(2), gerekli müdahale yapılmazsa, ileride Türk kadınların durumunu, Türk erkeklerinin yukarıda tanımlanan olumsuz durumuna doğru çekebilir.

1990 hızları ile 2003 hızlarının karşılaştırılması: 1990 ve 2003 yılları, yaşa ve cinsiyete özgü ölüm hızları ile karşılaştırıldığında, hem kadınlarda hem erkeklerde her yaş grubunda ölüm hızlarının düştüğü görülmektedir. En fazla düşme her iki cinsiyette de bebeklik dönemindedir. 1-4 yaş ve 55-64 yaş grupları da her iki cinsiyette ölüm hızlarının bebeklik döneminden sonra en çok düştüğü gruplardır. Ölüm hızlarının en az düştüğü yaş grubu, her iki cinsiyette de 15-24 yaş grubudur. Yaşa göre standardize hızlarla, 2003 yılı ölüm hızı, 1990 ölüm hızından erkeklerde % 31, kadınlarda % 28 daha düşüktür. 35-64 yaş aralığı için yaşa göre standardize hızlarla 2003 yılı ölüm hızı, 1990 yılı ölüm hızından, erkeklerde % 30, kadınlarda % 29 daha düşüktür. Bu bulgular, 1990 ve 2003 yılları ölüm kayıtlarının aynı eksiklik düzeyinde olması sayılıyla, Türkiye'de ölümlülüğün son 13 yılda yaklaşık % 30 düzeyinde düştüğünü düşündürmektedir.

Sonuç

Türkiye'de ölümlülüğü değerlendirme ve izleme yönünden önemli bir eksiklik vardır. Ölüm kayıt sistemi yetersizdir, mevcut ölüm kayıtları yaklaşık üçte bir oranında eksiktir. Düzenli aralıklarla yapılan ve ölümlülüğü ulusal düzeyde değerlendiren ve izleyen araştırmalar yoktur. Bu durum sağlık hiz-

metlerinin etkili sunumunu engelleyen temel eksikliklerden biridir. Türkiye’de bebek ve çocuk ölümlülüğünün yanı sıra 35 yaş sonrası dönem ölümlülüğü de yüksek düzeydedir. Bu nedenle bu yaş grubunun ölüm nedenleri olan koroner kalp hastalığı, serebrovasküler hastalık ve kansere yönelik ulusal müdahale programlarının başlatılmasına ve güçlendirilmesine gerek vardır.

Kaynaklar

1. Hacettepe Üniversitesi Nüfus Etüdüleri Enstitüsü. Türkiye’de Nüfus Yapısı ve Nüfus Sorunları 1973 Araştırması. Hacettepe Üniversitesi Yayınları D-25, Ayyıldız Matbaası, Ankara, 1978.
2. Hacettepe Üniversitesi Nüfus Etüdüleri Enstitüsü. Türkiye Nüfus ve Sağlık Araştırması 2003. Hacettepe Üniversitesi Nüfus Etüdüleri Enstitüsü, Ankara, 2005.
3. Hacettepe Üniversitesi Nüfus Etüdüleri Enstitüsü. Türkiye Nüfus ve Sağlık Araştırması 1988. Hacettepe Üniversitesi Nüfus Etüdüleri Enstitüsü, Ankara, 1989.
4. T.C. Başbakanlık Devlet İstatistik Enstitüsü. 1989 Türkiye Nüfus Araştırması. Devlet İstatistik Enstitüsü Matbaası, Ankara, 1991.
5. Sümbüloğlu V, Sümbüloğlu K. Sağlık Hizmetlerinde Veri Toplama ve Değerlendirme Yöntemleri. Aydoğdu Ofset, Ankara, 1995.
6. Fişek HN. Halk Sağlığına Giriş. Çağ Matbaası, Ankara, 1983.
7. ONAT A. Erişkinlerimizde kalp hastalıkları prevalansı, yeni koroner olaylar ve kalpten ölüm sıklığı. ONAT A (editör):TEKHARF Türk erişkinlerinde kalp sağlığı, risk profili ve kalp sağlığı. Yenilenmiş ikinci baskı, Ohan Matbaacılık, s:16-23, 2000.
8. U.S. DHHS, National Institutes of Health, National Heart, Lung, Blood Institutes. Morbidity and Mortality: 2004 Chart Book on Cardiovascular, Lung and Blood Diseases. May 2004.
9. United States Department of Health and Human Services. The health consequences of smoking: A report of the Surgeon General. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2004.
10. T.C. Başbakanlık Devlet İstatistik Enstitüsü. Ölüm İstatistikleri İl ve İlçe Merkezlerinde 1990. Devlet İstatistik Enstitüsü Matbaası, Ankara, 1993.
11. T.C. Başbakanlık Devlet İstatistik Enstitüsü. 1990 Genel Nüfus Sayımı: Nüfusun Sosyal ve Ekonomik Nitelikleri. Devlet İstatistik Enstitüsü Matbaası, Ankara, 1993.
12. Office for National Statistics, UK. Mortality Statistics: Review of the Registrar General on Deaths by Cause, Sex, and Age in England and Wales, 2003. Series DH2 no.30, <http://www.statistics.gov.uk>

DİSKİRMINANT (AYIRMA) ANALİZİ İLE EL KULLANIMI TAYİNİ

Bülent Ediz¹, Şengül Cangür¹, Gökhan Ocakoğlu¹, İlker Ercan¹,
N.Şimşek Cankur², İhsaniye Coşkun², İsmet Kan¹

Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik AD.¹

Uludağ Üniversitesi Tıp Fakültesi Anatomi AD.²

Bilimin temel yöntem bilimlerinden birisi de karmaşık durumları daha makul duruma indirgeyerek sınıflandırmaktır. Nesnelerin sınıflandırılması belki de tüm bilimsel çabaların temelini oluşturan bir çalışmadır. Sınıflandırma problemi, araştırmacının bir birey üzerinde bireyin çeşitli özellikleri bakımından ölçüm yapması ve bu bireyi elde ettiği ölçümlere dayanarak sayısı önceden bilinen gruplardan birisine yerleştirmek istemesiyle ortaya çıkmıştır. Botanik ve antropolojideki sınıflandırma problemleri “**irksal benzerlik katsayısı**” ve “**ayırma fonksiyonu**” ‘nun, dolayısıyla çok değişkenli istatistikte “**ayırma analizi(AA)**”nin doğuşuna neden olmuştur.

En çok kullanılan elin saptanması amacıyla, Uludağ Üniversitesi Tıp Fakültesi Anatomi Anabilim Dalı koleksiyonunda bulunan, 120 humerus'ta (62 sol, 58 sağ) sulcus intertubercularis ölçümleri yapıldı. Ölçüm parametreleri olarak sulcus intertubercularis'e ait genişlik, derinlik, iç duvar açıları ve dış duvar açıları alındı. Elde edilen ölçümlere ayırma (diskriminant) analizi uygulandı. Ayırma analizi ile iç ve dış duvar açılarına ait değerlerin, sağ ve sol el ayırımında anlamlı etkisinin olduğu saptandı ($P<0,001$). Ancak, genişlik ve derinlik değerleri el ayırımında anlamlı etkiye sahip değildi ($P>0,05$).

Anlamlı bulunan değişkenlerle elde edilen ayırma fonksiyonunun doğru sınıflama oranı % 85,8 olarak bulundu. Bu sonuçlar iç ve dış duvar açılarının ölçümlerinin, sağ el kullanımının baskınlığını belirlemede değerli kriterler olduğunu düşündürdü.

Anahtar Kelimeler: Diskirminant (ayırma) analizi, El tercihi, sulcus intertubercularis.

İKİ YANIT DEĞİŞKENLİ DEĞİŞİK VERİ SETLERİNDE LOJİSTİK REGRESYON VE DİSKİRMİNANT ANALİZİ'NİN DOĞRU SINIFLANDIRMA ORANLARI:

Bir Simülasyon Çalışması

**Bülent Ediz, Güven Özkaya, Deniz Sığırlı, Şengül Cangür,
Gökhan Ocakoğlu, İlker Ercan, İsmet Kan**

Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik AD

e-mail: ediz@uludag.edu.tr

Gözlenen verilerin analiziyle ilgilenildiğinde, kuramsal istatistik bir modelin matematik fonksiyonlarla ifade edilmesi gereği ortaya çıkar. Bu fonksiyonlar, gözlenen verilerden hareketle gelecekteki olaylar hakkında tahmin yapılmasına ve olaylara etki eden faktörlerin belirlenmesine olanak sağlarlar. Olaylara çözüm getirmede, aynı amaç için farklı modeller gerekebilir. Ancak, aynı amaç için farklı modellerin kullanılması bazı risklere yol açabilir. Tercih belirleyecek etken, riski en az olan yol seçilerek bu riskleri minimum yapan daha etkin modelin kullanılması olmalıdır. Gözlemleri, verilerin yapısında bulunan olası gruplara atamak için kullanılan yöntemlerden ikisi, ayırma (discriminant analysis) analizi (**AA**) ve lojistik regresyon analizleridir (Logistic Regression Analysis)(**LRA**).

Bu amaçla biri iki yanıt düzeyli bağımlı (binary), 10'da bağımsız değişkenden oluşan 11 değişkenli veri setleri oluşturuldu. Verilerin üretiminde GENSTAT 8.0 for Windows programı kullanıldı. Veri setleri; büyüklükleri (n) 10, 20, 30, 50, 100, 300, 500, 1000, 2000, 5000 ve 10000, bağımlı değişkende ilgilenilen özelliğin ortaya çıkma olasılığı (p) 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 ve 0.9, bağımsız değişkenlerin tümü sürekli, tümü kesikli ve 5'i sürekli 5'i kesikli (karışık) değişkenlerden oluşacak şekilde veri setleri oluşturuldu. 11 farklı büyüklük, 9 farklı bağımlı değişken oranı ve 3 farklı veri tipinde toplam 297 veri seti incelendi.

Yapılan çalışmalar sonucunda LRA ve AA'nin $n < 20$ için kullanılması durumunda sınıflandırma sorunu yaşanmaktadır. Kesikli-Sürekli-Karışık bağımsız değişkenlerden oluşan veri setlerinde; $n=20$ olduğunda LRA ve AA tüm veri setlerinde çok yakın doğru sınıflandırma oranlarıyla gözlenmektedir ($P > 0.05$). $n=30$ için; kesikli veri setlerinde bağımlı değişkenin gözlenme olasılığı 0.1, 0.2, 0.4, 0.7, 0.8 ve 0.9 için LRA daha yüksek sınıflandırma

oranına sahipti ($P < 0.01$), sürekli veri setlerinde bağımlı değişkenin gözlenme olasılığı 0.6 iken LRA daha yüksek doğru sınıflandırma oranı bulundu ($P < 0.001$), karışık veri setinde ise 0.4-0.8 aralığında LRA AA'nden anlamlı derecede daha yüksek sınıflandırma oranına sahipti ($P < 0.01$). $n \geq 50$ iken tüm veri setlerinde bağımlı değişkenin gözlenme olasılığı $p = 0.5$ olması durumunda LRA ve AA doğru sınıflandırma oranları aynı ya da çok yakın çıkmaktadır ($P > 0.05$). Bağımlı değişkenin gözlenme olasılığı 0.5'ten uzaklaştıkça LRA'nin doğru sınıflandırma oranı AA'dan anlamlı derecede yüksektir ($P < 0.001$).

Anahtar Kelimeler: Lojistik Regresyon, Diskirminant (ayırma) analizi, Doğru Sınıflandırma Oranı, Simülasyon.

ARAŐTIRMA ETİK KURULLARINDA BİYOİSTATİSTİK UZMANININ ÖNEMİ

Elif Atıcı, Sezer Erer

Uludağ Üniversitesi, Tıp Fakültesi, Deontoloji Anabilim Dalı, 16059 Görükle-BURSA

e-mail: elifatici@uludag.edu.tr

Özet

Araştırmacı ya da hekimin, denek ya da hastanın haklarının gözetilmesi, yapılan tıbbi uygulama ve araştırmaların bilimsel standartları taşınması, yasalara ve ahlaki değerlere uygunluğunun denetlenmesi gereklidir. Araştırma etik kurulları, insan üzerinde yapılacak araştırmalarda etik standartların ve bilimsel değerin sağlanmasında önemli bir rol üstlenirler. Bilimsel değeri olmayan bir araştırma aynı zamanda etik ilkelere de aykırıdır. Biyomedikal araştırmalarda geçerli ve güvenilir sonuçların elde edilmesinde istatistik biliminden yararlanma bilimsel ve etik yükümlülüktür. Bu nedenle, biyoistatistiğin yoğun olarak kullanıldığı araştırmaları etik ve bilimsel açıdan değerlendiren araştırma etik kurullarında biyoistatistik uzmanının bulunması gereklidir.

Bu çalışmanın amacı, araştırma etik kurullarında bir biyoistatistik uzmanının bulunmasının, araştırmaların bilimsel değerlendirilmesinin yanı sıra etik değerlendirilmesinde de gerekli olduğunu vurgulamaktır.

Anahtar Kelimeler: Biyoistatistik, Etik, Araştırma Etik Kurulları.

Giriş

Araştırma etik kurullarının amacı, insanlar üzerinde yapılan biyomedikal araştırmaların (farmasötik araştırmalara, cerrahi yöntemler, medikal araçlar, tıbbi radyasyon, tıbbi kayıtlar, biyolojik örnekler, epidemiyolojik, sosyal ve psikolojik araştırmalar gibi) etik ve bilimsel standartlara uygunluğunu değerlendirmektir. Araştırma etik kurullarına sunulan araştırma projeleri, etik açıdan ve bilimsel açıdan değerlendirilirken istatistik yöntemlerinin doğru kullanılıp kullanılmadığı da göz önüne alınmalıdır. Bunun değerlendirilmesi amacıyla araştırma etik kurullarında biyoistatistik uzmanının bulunması gereklidir, çünkü istatistiğin doğru kullanımı, araştırmacının bilimsel niteliğini artırır. Araştırmacının bilimsel kurallara uygun olarak planlanması, yürütülmesi ve değerlendirilmesi, aynı zamanda araştırmacının etik açıdan da kabul edilebilmesinin koşullarından birini oluşturur. Unutulmamalı ki, etik değerler ve bilimsel doğrular bilimin ayrılmaz parçalarıdır.

Bu çalışmayla amaçlanan, araştırma etik kurullarında bir biyoistatistik uzmanının bulunmasının, araştırmaların bilimsel değerlendirilmesinin yanı sıra etik değerlendirilmesinde de gerekli olduğunu vurgulamaktır.

Araştırma Etik Kurullarının Önemi ve Biyoistatistik Uzmanının Gerekliliği

Ahlaki ilkelerin, teknolojik ve bilimsel gelişmenin gerisinde kalması, yeni etik sorunların ortaya çıkmasına neden olur. Bu durumu engellemek için oluşturulan mekanizmalar öncelikle etik ilkelere sahip çıkılmasını sağlayacak etik eğitiminin verilmesi, etik yaklaşımın yasalar ile güvence altına alınması ve denetlemenin yapılmasıdır. Geçmişte insanlar üzerinde yapılan etik dışı tıbbi denemelerin, bilim dünyasında ve toplumda yarattığı rahatsızlık ve tartışmalar denetleme mekanizmalarının oluşturulmasını gerekli kılmıştır. Tarihsel sürece bakıldığında etik değerlerin belirlenmiş olmasına karşın, uygulamada sorunlar yaşandığı görülmektedir. Araştırmacı ya da hekimin, denek ya da hastanın haklarının gözetilmesi, yapılan tıbbi uygulama ve araştırmaların bilimsel standartları taşıması, yasalara ve ahlaki değerlere uygunluğunun denetlenmesi gerekli hale gelmiştir. Bu nedenle klinik uygulama ve ilaç araştırmalarının bilimsel niteliğinin değerlendirilmesinde danışmanlık hizmeti veren etik kurullar oluşturulmuştur (1). "Araştırma"; insanın bilgi edinmesinde, teknolojiye ilerlemenin sağlanmasında çevreyi tanıyarak ondan en iyi biçimde yararlanmasında ve sorunlarını çözmesinde başvurduğu önemli bir etkinliktir. Bilinmeyen olay ve etmenleri ortaya çıkarabilmek, bilinenleri geliştirmek, herhangi bir konuyu aydınlatmak, bir soruna çözüm getirmek, belirli kavram, kuram ya da yasalara ulaşabilmek amacıyla yapılan bilimsel bir çalışma olarak da tanımlanabilir. "Araştırma süreci" ise, çağın bilim çevrelerince uygunluğu kabul edilmiş bilimsel yöntemlerin kullanılmasıyla, yeni bulgular elde etmesine yönelik bir sorun çözme etkinliği olarak kabul edilebilir. Bilimsel bir araştırmada en önemli konu araştırmanın bilimsel yöntemlerle planlanması, yürütülmesi ve sonuçlandırılmasıdır. Ancak bilimsel yöntemlerden yararlanılarak yapılan araştırmanın sonucu geçerli ve güvenilir olabilir (2).

Araştırma etiği yalnızca bilimle sınırlı soyut bir kavram olarak görülmemeli, genel etik değerlerin özel bir alandaki yansımaları oluşturduğu da göz önüne alınmalıdır. Bilginin kendisi yansızdır. Ancak, toplumsal uygulamada ve bilginin teknolojiye dönüşümünde bilgi yansız olmaktan çıkar. Bu nedenle bilimsel bilgiyi ortaya çıkaracak olan araştırmacının etik olarak toplumca onaylanabilirliğinin de değerlendirilmesi gerekmektedir (3). Araştırmaların sonucunda elde edilen bilginin yansız olarak sunulabilmesi için araştırmanın planlanmasından değerlendirilmesine kadar geçen her aşamasında biyoistatistik ilke ve yöntemlerinin doğru bir şekilde kullanılması büyük önem

taşımaktadır. Bunun değerlendirmesini yapabilecek olan ise biyoistatistik uzmanıdır.

Araştırma etik kurulları, insan üzerinde yapılacak araştırmalarda etik standartların ve bilimsel değerin sağlanmasında önemli bir rol üstlenirler. İnsan üzerinde yapılan deneylerle ilgili etik sorunların kurumsallaşmasına karşın hasta ve denek için, araştırmacı ve araştırmacı kurum için ortaklaşa etik ve hukuksal denetim oluştururlar. Asıl amaç deneğin korunmasıdır. Koruma işleminden birinci derecede sorumlu olan araştırmacı, koruma işlemini denetleme ve yönlendirme durumunda olan yapılanmaların başında ise araştırma etik kurulları gelmektedir. Temel olarak araştırma etik kurullarının görevi, araştırma protokollerinin etik standartlara uygunluğunu kontrol etmektir (4).

Araştırma etik kurulları, üç önemli sorumluluğa sahiptir. İliki ve en önemlisi, kurulun denek haklarının korunduğundan emin olmasıdır. Bunun sağlanmasında deneklerin hiçbir baskı altında kalmadan, özgür iradeleri ile, araştırmacının yarar ve zararları hakkında anlaşılır ve tam olarak bilgilendirilmelerinin ve onamlarının alınması önem taşır. İkincisi, kurulun araştırmaya kaynak sağlayarak sonuçlardan etkilenecek olan topluma karşı olan sorumluluğudur. Üçüncü olarak kurul araştırmacıya karşı sorumludur. Bu sorumlulukların yerine getirilmesi ile araştırmacının etik ve bilimsel standartlarının yüksek tutulması sağlanır (4, 5).

İnsan denekler üzerinde biyomedikal araştırma yapmanın etik ve bilimsel standartlarının saptanması amacıyla Helsinki Bildirgesi, İnsan Deneklerle İlgili Biyomedikal Araştırmaların Uluslararası Etik Yol Gösterici Kuralları (CIOMS), Dünya Sağlık Örgütü (WHO) ve İyi Klinik Uygulamalar Konusunda Yol Gösterici Kurallar (ICH) gibi uluslararası yol gösterici kurallar oluşturulmuştur. Bu kuralların amacı, öncelikle deneğin haklarının korunmasını sağlamak, araştırmacının bilimsel ve etik açıdan değerlendirilmesinde yardımcı olmaktır.

Araştırma etik kurulları deneklerin güvenliği, sağlığı ve insan haklarının korunmasının, bilimin ve toplumun çıkarlarının hiçbir zaman deneğin sağlığı ile ilgili kaygıların üzerine çıkarılmamasının ve deneklerin olası bir kullanımının önlenmesinin temel güvencesidirler. Bu güvencenin sağlanması, Helsinki Bildirgesi ve ulusal düzenlemelerde tanımlandığı biçimde araştırma protokollerinin araştırma öncesi değerlendirilmesi ile olur. Araştırma etik kurulları, araştırma protokollerinin sadece etik ölçülere uygunluğunu denetlemekle kalmaz, aynı zamanda araştırmacının amacını, gerekçesini, yaklaşım ve metodolojisini göz önünde tutarak bilimsel açıdan da inceler. Bilimsel değeri olmayan bir araştırma etik ilkelere de aykırı olacağından araştırma projesinin etik yönleri ve bilimsel yönleri birbirinden ayrılamaz. Araştırmacının planlama, yürütülme ve değerlendirme aşamalarında yapılacak hataların, araştırmacının bilimselliğine gölge düşürebileceği göz önüne

alındığında araştırma projelerinin değerlendirilmesinde biyoistatistik uzmanına danışılması araştırmanın bilimsel düzeyinin değerlendirilmesinde önemli bir rol üstlenebilir. Bu nedenle araştırma etik kurullarında biyoistatistik uzmanına yer verilmelidir (1).

Türkiye’de Araştırma Etik Kurulları

Türkiye’de ilk araştırma etik kurulu 1986 yılında kurulmuştur. Araştırmaların etik kurul onayını aldıktan sonra başlamaları ve yine etik kurulun gözetiminde sürdürülmeleri gerekliliği ise yasal düzeyde 1993 ten beri bulunmaktadır. T.C. Sağlık Bakanlığı tarafından geliştirilen ve 23 Ocak 1993 tarih, 21480 sayılı resmi gazetede yayımlanan “İlaç Araştırmaları Hakkında Yönetmelik”, Türkiye’de etik kurulların oluşturulması sürecinde dönüm noktasını oluşturur. Adı geçen yönetmelikle birlikte 25 Mayıs 1994 tarih ve 21942 sayılı resmi gazetede yayınlanan “Farmasötik Müstahzarların Biyoyararlanım ve Biyoeşdeğerliliğinin Değerlendirmesi Hakkında Yönetmelik”, klinik araştırmaları düzenleyen esasları belirler. Sağlık Bakanlığınca 29 Aralık 1995 tarih ve 51748 sayılı resmi gazetede yayınlanan “İyi Klinik Uygulamaları Kılavuzu” ve “İyi Laboratuvar Uygulamaları Kılavuzu” ile araştırmaların uluslararası bilimsel ve etik standartlarda yapılmasını sağlamak üzere laboratuvar çalışmaları ve/veya klinik araştırmalarda uyulması gereken kurallar ve izlenecek yol belirlenmiştir.

İlaç Araştırmaları Hakkında Yönetmelik’te, ilaçla yapılacak klinik araştırma isteğinin incelenebilmesi, klinik araştırma ve çalışmaların etik açıdan değerlendirilmesi ve kontrolü amacıyla, Sağlık Bakanlığının merkezinde Bakanlıkça oluşturulan kurul *merkezi etik kurul*, araştırmayı yapacak kurum ya da kuruluşlarda oluşturulan ve bu yönetmelikte belirlenen görevleri yürüten kurullar ise *yerel etik kurullar* olarak tanımlanır. Herhangi bir ilaç konusunda bilimsel amaçla klinik araştırma yapmak isteyen kişi, kurum ya da kuruluşlar, bu yönetmelik hükümlerine uygun olarak yerel etik kurul aracılığı ile Sağlık Bakanlığına başvuruda bulunmak ve izin almak zorundadırlar. İnsan üzerinde araştırma yapılabilmesi için etik kurul onayının alınması yasal bir zorunluluktur (6).

Sağlık Bakanlığı bünyesinde kurulan merkezi etik kurulunun ve araştırmanın yapılacağı eğitim hastanelerinde kurulan yerel etik kurulların hangi üyelerden oluşacağı ve üyelerin özellikleri ilgili yönetmelikte belirtilmiştir. Bu üyeler klinisyen, farmakolog, eczacı, biyokimya uzmanı, patolog, yapılacak çalışmanın özelliğine uygun bir uzman ve mümkünse bir tıbbi deontoloji uzmanıdır. Üyelerin arasında biyoistatistik uzmanı yer almamaktadır.

Yasal düzenlemeler bulunmakla birlikte etik kurulların işleyişi ve yapılanmaları ile ilgili bazı sorunlar bulunmaktadır. Bunlar :

1. İlgili yönetmeliklerin sadece ilaçlarla ilgili insan üzerinde yapılacak arařtırmaları ele almıř olması,
2. Yerel etik kurulların eđitim hastanelerinde yer almalarında ele alınacak kriterlerin belirlenmemiř olması,
3. Karar verme sũrecinin iřleyiři aısından yerel etik kurullarca gũrũř birliđine varılmıř bir yũntemin olmaması,
4. Etik kurul kararlarında gecikmenin olabilmesi,
5. Bazı yerel etik kurullarda arařtırmalarda bilimsel niteliklerin dikkate alınmaması,
6. Etik kurullarda biyoistatistik uzmanının ȃngȃrũlmemesi.

Etik kurullarda ortak bir yapılanma ve iřleyiř biimi oluřturulması ve bunun tũm ũlkelerdeki arařtırma etik kurul yȃnergeleri ile uyum iinde olması kuřkusuz yararlı olacaktır. Son iki maddede yer alan sorunlar, ele aldığımız konu ile iliřkilidir. Arařtırmaların bilimsel niteliklerinin deđerlendirilmesinde planlamadan deđerlendirmeye kadar arařtırmanın her ařamasında biyoistatistik ilke ve yũntemlerinin dođru kullanımı ȃnem tařımaktadır. Bunun deđerlendirmesini yapabilecek bilgi ve yeteneđe sahip olan ise biyoistatistik uzmandır. Arařtırmaların bilimsel niteliklerinin deđerlendirilmesi aynı zamanda etik deđerlendirmenin kořullarından biridir ve bu nedenle etik kurullarda biyoistatistik uzmanı bulunmalıdır.

24 Temmuz 2003 tarih ve 25178 sayılı resmi gazetede yayınlanan *Avrupa Birliđi Mũktesebatının ũstlenilmesine İliřkin Tũrkiye Ulusal Programı'nda İla arařtırmaları hakkında Yȃnetmeliđin ve Farmasȃtik Mũstahzarların Biyoyararlanım ve Biyoeřdeđerliliđinin Deđerlendirmesi Hakkında Yȃnetmeliđin yũrũrlükten kaldırılması ve 2001/20/EC sayılı Avrupa Parlamentosu ve Konsey Direktifi'ne uyumlu yeni yȃnetmeliđin yayınlanması ȃngȃrũlmũřtũr. Bu amala Beřeri Tıbbi ũrũnlerin Klinik Arařtırmaları Hakkında Yȃnetmelik Taslađı hazırlanmıřtır. Bu taslak sadece ila arařtırmalarını kapsamamaktadır. Bȃylece yukarıda sıraladıklarımız arasındaki ilk sorun ȃzũmlenmektedir. Aynı zamanda bu taslađa gȃre, merkezi etik kurulda biostatistik uzmanı yer almaktadır. Ancak yerel etik kurullarda ne biyostatistikçi ne de etik uzmanı yer almamaktadır. Etik uzmanı ve biyostatistikçi arařtırmaların etik ve bilimsel deđerlendirilmesinde ȃnemli bir yere sahiptir. Bunların yerel etik kurullarda yer almamaları taslađın ȃnemli bir eksikliđidir.*

Arařtırma Etik Kurullarında Etik ve Bilimsel Deđerlendirmenin Yapılması

Arařtırmaya dayanan tıbbi geliřme, insan denekler ũzerinde deneme yapılmasını gerekli kılmaktadır. Bununla birlikte, insan denekler ũzerindeki biyomedikal arařtırmalar tanı koyucu, iyileřtirici ve koruyucu iřlemleri geliřtirme, hastalıđın etyopatogenezini anlama amalı olmalıdır. İnsan ũzerindeki arařtırmaların etik boyutunun erevesini izen Helsinki Bildirgesi'nde; arařtırmaların genel bilimsel ilkelere uyması gerektiđi, insan ũzerindeki

araştırmalarda bilimin ve toplumun yararının, hiçbir zaman deneğin sağlığına verilen önemden daha öncelikli olamayacağı belirtilmektedir. Helsinki Bildirgesi'nde son olarak 2000 yılında değişiklikler yapılmıştır. Bildirgenin 13. maddesinde "İnsan üzerinde yapılacak olan her araştırma niteliğindeki yöntemde, araştırmanın düzenlenmesi ve uygulanmasının bir araştırma protokolüne bağlı olması ve bu protokolün değerlendirme, yorumlama ve yol göstermesi amacı ile bu konuda özelleşmiş bir etik kurula sunulması ve bu kurulun araştırmacı, destekleyici ve her tür etkiden uzak, bağımsız bir kurul olması" öngörülmektedir. Yapılan bu son değişiklikler ile araştırmanın yöntemine de müdahale edilmesi söz konusu olmaktadır (7). Bu açıklamadan yola çıkarak biyomedikal araştırmaların etik kurullar tarafından iki yönden değerlendirilmesi gerektiğini söyleyebiliriz:

- 1- Etik değerlendirme
- 2- Bilimsel değerlendirme

Etik Değerlendirme

İnsan deneklerin yer aldığı biyomedikal bir araştırmada göz önüne alınması gereken etik ilkeler; özerkliğe saygı (buna bağlı olarak aydınlatılmış onam), yararlı olma (buna bağlı olarak zarar vermeme), adalet ve gizlilik ilkeleridir. Bu ilkeler insan üzerinde yapılan denemelerin ahlaki yönüne şekil verirler.

Özerkliğe saygı ilkesi, araştırmada yer alan deneklerin aydınlatılmış onamlarının alınmasını ve araştırmanın risklerinden denekleri korumayı amaçlar. Yararlılık ilkesi, araştırmanın amacına uygun kar-zarar oranının gösterilmesine dayanır. Adalet ilkesi araştırmanın zarar ve yararının eşit olarak dağıtılmasını gerektirir. Gizlilik ilkesinin gereği ise, deneklerin kişilik haklarına saygı göstererek, kimlik bilgilerinin gizli tutulmasıdır (8).

Önceden denenmemiş olan ve insan üzerindeki etkileri kesinlikle bilinmeyen uygulamaların tedavi ya da bilimsel amaçla sınırsız olarak uygulanması hasta-deneğin hayatına, dokunulmazlığına büyük ölçüde zarar verebilir. Tıbbi denemeleri uygulayan araştırmacının sorumsuzluğu geri dönülmez olumsuz sonuçlara neden olabilir. Kişinin beden bütünlüğü üzerindeki girişimlerin etik açıdan olduğu kadar yasalarca da geçerli olabilmesi için bazı temel koşulların sağlanması gerekir. Bu koşulların ilki, hasta-deneğin geçerli onamının olmasıdır. Hasta-deneğin hiçbir kuşkuya yer vermeden uygulama ve bunun doğuracağı sonuçlar ve riskler hakkında bilgilendirilmesi, onamı verirken herhangi bir baskı altında kalmaması, kişinin içinde bulunduğu maddî, manevî ve fiziksel koşullardan faydalanılmaması verilen onamı geçerli kılar. İkinci önemli temel koşul, tıbbi denemelerden hasta-deneğin zarar görmemesi ve yaşamının tehlikeye atılmamasıdır. Hastalığın tedavisi için bilinen, bilimsel tıp tarafından onaylanmış olan tedavi yöntemlerinin sonuç vermediğinden emin olmak gerekir. Tedavi etme amacıyla ya da

bilimsel bilgi sağlama amacıyla girişimde bulunacak araştırmacı-hekimin bu konuda yeterli bilgi, yetenek ve beceriye sahip olması beklenir (9).

Adalet ilkesi gereğince; yaş, cinsiyet, ekonomik statü, kültür ve etnik değerlendirmeler de dikkate alınarak araştırmanın yarar ve sıkıntıları toplumdaki tüm grup ve sınıflar arasında eşit dağıtılmalıdır (10).

Araştırmanın planlanmasından sunumuna kadar geçen her aşamada karşılaşılabilecek etik sorunları üç ana başlık altında toplamak olanaklıdır:

1. *Araştırmanın planlanması ve uygulanması sırasında gözlenen etik sorunlar:*

1.a. *Araştırmanın amacı:* Araştırmanın toplumsal ve bilimsel gereksinimlere ne derece hizmet ettiği karşılaşılan önemli etik sorunlardan biridir. İnsan üzerinde yapılan deneyler, kişiler hakkında bilgi toplamak ya da kişilerin araştırmaya fiziksel katılımları ile olanaklıdır. Birinci durumda, kişilerin özel yaşamlarına, kültürel yapılarına etki söz konusu iken ikinci durumda az ya da çok beden bütünlüğüne müdahale ön plandadır.

Bu aşamada ortaya çıkabilecek etik sorunlar, özerkliğe saygı ile bağlantılı olarak aydınlatılmış onam ve yararlı olma ilkeleri açısından ele alınabilir. Araştırmanın amacı öncelikle deneğin yararı gözetilerek, topluma ve bilime katkı sağlaması olmalıdır. Klinik denemenin planlanması ve yürütülmesindeki asıl zorluk, deneye katılan hastalara en iyi ve en uygun tedaviyi sağlarken aynı zamanda oluşturulan veri ve geliştirilen tedaviden değerli sonuçlar alma çabasıdır.

Araştırmalar iki nedenle yapılır: 1-Doğrudan deneğin tedavisi, 2-Geliştirilmek istenen tıbbi uygulama ya da tedavinin sağlıklı denekler üzerinde denenmesi. Her iki durumda da öncelikle dikkat edilmesi gereken nokta deneklerin aydınlatılmış onamlarının alınması olacaktır. Denekler uygulamanın sonuçları ve riskleri hakkında açıkça bilgilendirilmeli, hiçbir maddi, manevi baskı altında olmadan onamları alınmalıdır. Bilgilendirmenin yapılması tek başına yeterli değildir. Bunun denek tarafından doğru şekilde anlaşıldığından da emin olmak gerekir. Deneğin tedavisine yönelik bir araştırmada A tedavisi ile iyi sonuç alındığı bilindiği halde, araştırmanın konusu olan ancak kar-zarar oranı tam olarak bilinmeyen B tedavisinin uygulanması etik açıdan yararlı olma ve buna bağlı olarak zarar vermeme ilkelerine uygun olmayacaktır.

1.b. *Araştırma yönteminin yer, zaman denekler ve bilimsellik açısından etik ilkelere uygunluğu:* Araştırmanın amacı uygulanacak yöntemle uyum içinde olmalıdır. Araştırmaya katılacak deneklerin çalışmaya alınma ve alınmama kriterleri açık olarak belirtilmeli, seçilen kontrol grubunun gerçek bir kıyaslamaya uygun olması sağlanmalı, deneklere ait özellikler çalışma ve kontrol grupları arasında eşit olarak dağıtılmalı, gözlemler gerekli

nesnellik içinde yapılmalıdır. Araştırmalarda karşılaşılan önemli sorunlardan birisi de, araştırmacının isteyerek ya da istemeden yanlı davranmasıdır. Yanlı davranma araştırma yönteminin seçiminde, değerlendirme ve yayın aşamalarında görülebilir.

Bu aşamada yararlı olma ve adalet ilkelerine uygun olarak hareket etmeye dikkat etmek gerekir. Araştırma sonuçlarının istenilen şekilde çıkması amacıyla araştırmacının denek seçiminde yanlı davranması ile araştırma sonuçları bilimsellik ve gerçeklikten uzak olacağı gibi etik açısından da onaylanamaz.

1.c. *Araştırmanın mali desteği ile ilgili etik sorunlar:* Araştırma yapılırken öncelikle toplum çıkarları göz önüne alınmalıdır. Araştırmalara mali kaynak sağlayan şirketlerin yaklaşımları kar amacı doğrultusunda olduğunda, kendi çıkarlarıyla ya da beklentileriyle uyuşmayan araştırmalara destek olmama, istenmeyen sonuçlar çıktığında yayınlanmasını engellerek ya da geciktirerek yanlı davranma ve toplumu yanıltma gibi yaklaşımlarda bulunabilirler. Bu durumun engellenmesi için mali desteğin bir araştırmaya özgü değil genel bir araştırma fonuna verilmesi sağlanabilir. Bu konuda karşılaşılan diğer bir sorun ise tedavi hizmetlerine yönelik araştırmaların halk sağlığını korumaya yönelik araştırmalardan daha kolay mali kaynak bulabilmesidir. Bu durum, araştırmaların desteklenmesinde kar amacının ağırlığı olduğunun kanıtlarından biri olarak görülebilir.

Böyle bir durumda yararlı olma ilkesinin ihlali söz konusu olacaktır. İlaç geliştirilmesi ve pazarlanması trilyon dolarlık bir endüstri haline gelmiştir. Bu durum finansal taleplerin klinik denemelerin tasarlanmasının ve planlanmasının etkilenmesine neden olabilir. Şirketlerin desteklediği araştırmalarda sonuçların yanlı yansıtılması ve negatif sonuçların açıklanmaması araştırma konusunun yanlı yorumlanmasına, diğer araştırmacıların zaman kaybetmesine ve hastaların zarar görmesine neden olabilir.

2. *Araştırma sonuçlarının değerlendirilmesinde gözlenen etik sorunlar:* Sonuçların değerlendirilmesinde sık yaşanan sorunlardan biri çalışmaya alınmayan, çalışmadan çıkarılan ya da çalışma sırasında kayıtları kaybolan deneklere ait sonuçları hiç değerlendirmemek ya da bunlardan hiç söz etmemektir. Sonuçlarda bilerek değişiklik yapmak, sonuçları uydurmak, sonuçların bir kısmını çıkartmakla yanıltma ve sahtekârlık yapılması bilimsel açıdan olduğu kadar etik açısından da kabul edilemez.

Bilimsel yöntemlerin uygun olarak kullanılmaması, bilerek ya da bilmeyerek araştırma sonuçlarını etkileyebilecek hataların yapılması ile elde edilen sonuçlar yanıltıcı olabilir. Klinik araştırmaların değerlendirilmesinde iki önemli sorunun yanıtlanması gerekir: 1-Verilere güvenebilir miyim? 2-Elde edilen sonuçları hastalarımda kullanabilir miyim? Klinik araştırmalarda hatadan kaçınmak önemlidir. Yirminci yüzyılda istatistik yöntemlerinin kullanımında görülen artış ve bu yöntemlerin hassasiyeti hataya yol açan

sorunların anlaşılmasını kolaylaştırmıştır. Araştırmalarda istatistiksel yöntemlerin yersiz ya da yanlış kullanımı, sonuçların % 50 oranında farklı çıkmasına neden olabilmektedir. Böyle bir durumda, örneğin araştırmalara göre etkisi iyi bulunan bir ilacın kullanılması ile, beklenen yarar sağlanamayacaktır.

3. *Araştırmanın yayınlanması sırasında görülen etik sorunlar:* Araştırmanın birden çok yerde yayınlanması, tekrar yayınlanması, yayınlanmasının engellenmesi ya da geciktirilmesi ve yazarlık ile ilgili sorunlarla karşılaşabilmektedir (2-3, 11-15).

Bilimsel değerlendirmenin ve etik değerlendirmenin bir bütünün ayrılmaz parçaları olduğu görülmektedir. İnsan deneklerin yer aldığı araştırmaların etik açıdan değerlendirilmesi yapılırken aynı zamanda bilimsel açıdan değerlendirilmesi de yapılmaktadır. Bunun tersi de söylenebilir.

Bilimsel Değerlendirme

Biyomedikal araştırmalarda geçerli ve güvenilir sonuçlar elde edebilmek için istatistik biliminden doğru bir şekilde yararlanılması bilimsel ve etik bir zorunluluktur. Bilimsel değere sahip bir araştırma yüksek kalitede istatistiksel planlama ve veri analizini gerektirir. İstatistik, bilimsel yöntemin temelidir. Biyomedikal araştırmaların istatistiksel olarak uygun biçimde planlanması, yürütülmesi ve değerlendirilmesiyle bilimsel temele oturtulması araştırmaların etik açıdan kabul edilebilmesinin koşullarından birini oluşturur. Biyomedikal araştırmada tıp etiği ilkelerine uyulmaması araştırma sonuçlarını geçersiz kılmamakla birlikte biyoistatistik etiğine uyulmamasının araştırma sonuçlarını geçersiz kılacağı özellikle vurgulanması gereken noktadır (16).

Biyoistatistik ilke ve yöntemlerinin araştırmalarda doğru şekilde kullanılmasının önemini kavrayan araştırmacı: 1- çok yönlü algılama ve çok yönlü düşünme yeteneği geliştirir. Bu görüşü kazanan araştırmacı yanlış davranma hatasına düşmeyecektir. 2- elde edilen bulguların geçerli ve güvenilir olması için tek yolun bilimsel yaklaşım olduğu ve bunun da ancak biyoistatistik ilke ve yöntemlerinin yerinde ve doğru kullanımı ile sağlanabileceğinin farkına varır (2).

Bir araştırmanın üç aşaması vardır; planlama, yürütme ve değerlendirme. Yapılan araştırmanın bilimsel temele oturtulabilmesi için her üç aşama da biyoistatistik ilke ve yöntemleri kullanılarak düzenlenmelidir. Ne yazık ki yaygın olarak yapılan yanlış, araştırmacıların biyoistatistik yöntemlerini araştırmaya başlamadan önce değil, verileri topladıktan sonra kullanma alışkanlığının olmasıdır. Bunun nedenleri ise, araştırmacıların biyoistatistik ilke ve yöntemlerinin bilimsel araştırmadaki önemini yeterince kavrayamamaları ve doğru zamanda kullanmamaları, bilimsel dergi editörlerinin bu konuya gerekli duyarlılığı göstermemeleri, araştırmacıların akademik ünün çekiciliği ile verileri daha ilginç yapma ya da araştırmanın değerini artırma amaçları olarak sayılabilir (17). Bilimsel değer taşımayan, geçerli ve

güvenirliliği kuşkulu olan bu tür arařtırmalardan elde edilen bulgular yarar sağlayamayacağı gibi, kiřilere zarar da verebilir.

Biyomedikal arařtırmanın sonuçlarını etkileyebilecek önemli hataları řu řekilde sıralayabiliriz:

1. Uygun düzenlenmemiř bir arařtırmadan elde edilen bulgular bir tedavinin veya bir ilacın hatalı olarak yararlı olduđu sonucunu (yanlıř pozitif) veya tersini (yanlıř negatif) dođurabilir. Uygun olmayan alıřma düzeninden yanlılıđa açık ve beklenmeyen sonuçların ıkması kaçınılmazdır. Uygun olmayan düzenle yapılan deney sonuçları, daha sonra uygun düzenle yapıldıklarında deđişiklik gösterebilir.
2. Uygun örnekleme yönteminin kullanılmaması ve testin gücü dikkate alınmadan belirlenen küçük örneklem büyüklüđu, hipotezler hakkında yanlıř kararlar verilmesine neden olabilir. Örneklem büyüklüđünün yetersiz olması nedeniyle, tedavi yöntemleri arasındaki önemli farkları yakalama řansı kaybolabilir.
3. Verilerin ölçülmesinde kullanılan ölçü biriminin dođru seilmemesi yanlıř sonuçlara neden olabilmektedir. Örneđin, bir ilacın etkinliđini ölçmek için hangi ölçü birimi kullanılacaktır? Hastanın iyileřiip iyileřiemediđi mi yoksa iyileřieme süresi mi? Hangi ölçü biriminin kullanılacağına arařtırmanın bařında arařtırmacı tarafından karar verilmelidir. Bu ölçülerden hangisi arařtırma sorusunu karřılıyorsa o ölçü alınmalıdır.
4. Kontrol kullanılarak yapılan deney alıřmalarında, seilen deney grubunun kendisi aynı zamanda kontrol grubu olarak da kullanılabilir (eřileştirilmiř grup). Böyle bir yolun tercih edilmesinin gerekli olduđu durumlarda arařtırmacının bilerek kendi beklentilerine cevap vermesinden dolayı eřiştirilmiř grup yerine ayrı bir bađımsız kontrol grubu alıp hipotez testi yapması yanlıř sonuç dođurur.
5. Arařtırmalarda yapılabilecek en büyük hata yan tutmadır. Bir ilacın deney-kontrol gruplarına uygulanmasında hastaların hangi grupta olduklarını hissettirecek istenilen sonucun alınması, anket sorularının hazırlanması sırasında soruların istenilen cevabın alınmasına yönelik hazırlanması, ölçümleri bilerek yanlı olarak ölçmek, hipotez testlerinin seiminde uygun olanın deđil kendisine yarayacak olanın seilmesi gibi bilerek yapılan hatalar buna örnek olarak gösterilebilir (2, 17).

Arařtırma sonuçlarını etkileyen bu hatalar, arařtırmanın deđerinin ya da güvenirliliđinin azalmasına neden olan uygulamalardır ve bilimsel yanılma olarak adlandırılır. Arařtırmacının bilgi ve eđitim eksikliđinden kaynaklanan, farkında olmadan yanlıř sonuç alınmasına neden olan disiplinsiz

araştırma, bilimsel yanılmadan ayrı olarak değerlendirilmelidir. Burada söz konusu olan araştırmacının bilerek verileri uydurması, gerçekleri çarpıtmasıdır (18).

Araştırmalarda biyoistatistik yöntemlerinin bilinçsiz kullanımı, bulguların yanlış yorumlanması, sonuçların yanlış ve çarpıtılarak sunulması gibi bilim etiğine aykırı davranışların neden olacağı sonuçlar, tıp etiğinin yararlı olma, zarar vermeme ve adalet ilkelerinin ihlal edilmesini de beraberinde getirmektedir.

Tartışma ve Sonuç

Biyomedikal araştırmalarda yer alan uygulamalar sayısız istatistiksel ve etik güçlüklerle doludur. Araştırmaların bilimsel değerinin korunmasında araştırma etik kurulları, etik değerlendirmenin gerekli ve önemli bir parçasıdır. Deneklerin katıldığı araştırmalarda etik olarak kabul edilebilir amaçların ve bu amaçlara ulaşmak için uygun araçların seçilmesi gerekmektedir.

Biyomedikal araştırmalarda bilimsel yöntemlerin doğru şekilde uygulanması etik dışı olarak nitelendirilen olaylarda azalmaya neden olacaktır. Bu nedenle doğru uygulanan istatistiksel yöntemler ile araştırmanın en iyi şekilde planlanması ve veri analizinin yapılması etik olarak gereklidir.

Araştırmaların planlanması, yürütülmesi ve değerlendirilmesi aşamalarında karşılaşılabilen sorunlar :

1. Çalışmanın amacını iyi belirlenmemiş olması ya da amacı yöntemle uygunluk göstermemesi,
2. Deneklerin çalışmaya alınma ve alınmama kriterlerinin açık olarak belirtilmemesi,
3. Kontrol grubunun gerçek bir kıyaslama için uygun olmaması,
4. Klinik olarak önemli özelliklerin çalışma ve kontrol grupları arasında eşit olarak dağıtılmaması,
5. Gözlemlerin gerekli nesnellikle yapılmaması,
6. Çalışmadan çıkan ya da çalışma sırasında kayıtları kaybolan deneklere ait sonuçların hiç değerlendirilmemesi ya da araştırmada bundan söz edilmemesi olarak sıralanabilir.

Araştırmanın bilimsel kurallara uygun olarak planlanması, yürütülmesi ve değerlendirilmesi, aynı zamanda araştırmanın etik açıdan da kabul edilebilmesinin koşullarından birini oluşturur. Bu nedenle biyoistatistik uzmanının kasıtlı olmayan ya da yan tutularak yapılan hataları önlemede, ortaya çıkarmada ve araştırmaların sonuçlarının etkilerini ölçmede mutlaka yer almaları gereklidir.

Araştırma etik kurullarının multidisipliner bir yapıya sahip olması, araştırmaların tüm açılardan ve her aşamasında yeterli olarak değerlendirilmesine katkı sağlayacaktır. Her araştırma ekibinde bir biyoistatistik

uzmanının mutlaka bulunması gerektiği gibi araştırma projelerinin etik ve bilimsel açıdan değerlendirilmesi aşamalarında araştırma etik kurullarında da biyoistatistik uzmanının katılımı sağlanmalıdır.

Kaynaklar

1. Atıcı, E.: Etik Kurullar. In: Erdemir AD, Öncel Ö, Aksoy Ş (eds). Çağdaş Tıp Etiği. İstanbul: Nobel Kitab.; 2003: 409-36.
2. Sümbüloğlu V, Sümbüloğlu K. Sağlık Bilimlerinde Araştırma Yöntemleri. Ankara: Hatiboğlu Yayınevi; 1988.
3. Bilimsel Araştırmada Etik ve Sorunları. Ankara: Türkiye Bilimler Akademisi Yayınları; 2002.
4. Gelling L. Role of the Research Ethics Committee. Nurse Education Today 1999; 19(7): 564-569.
5. Haris J. Research Ethics Committees-the future. Notizie di Politeia 2002; 18(67): 123-138.
6. İlaç Araştırmaları Hakkında Yönetmelik. 29.01.1993 Tarih ve 21480 Sayılı Resmi Gazete.
7. World Medical Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects. JAMA. 2003; 284(23): 3043-3045.
8. Aaron D. Research Ethics. West Indian Med J. 1995; 44(4): 115-158.
9. Erdemir AD. The Place of Medical Experiments in Medical Ethics and Its Importance. Lectures on Medical History and Medical Ethics. İstanbul: Nobel Tıp Kitabevleri; 1995.
10. Operational Guidelines for Ethics Committees That Review Biomedical Research. WHO. Cenevre. 2000.
11. Macbeth F, Stephens R. Marketing Clinical Trials. Lancet 1996; 348: 111-112.
12. Chalmers I. Underreporting Research Is Scientific Misconduct. JAMA 1990; 263(10): 1405-1408.
13. Feinstein AR. Tıp Araştırmalarının Biyoistatistik Değerlendirmesinde Sık Yapılan Yanlışlar. In: Yazıcı H (çev.). Tıpta Bilimsel Yazım, Editörlük ve Denetleme. Ankara: TÜBİTAK; 1994: 61-7.
14. Gluud C. Trials and Errors in Clinical Research. Lancet 2000; 354:59.
15. Garattini S, Bertele V, Bassi LL. How Can Research Ethics Committees Protect Patients Better? BMJ 2003; 326(7400): 1119-1121.
16. Thall PF. Ethical Issues in Oncology Biostatistics. Stat Methods Med Res. 2002; 11: 429-448.
17. Marco CA, Larkin GL. Research Ethics:ethical issues of data reporting and the quest for authenticity. Acad Emerg Med 2000; 7(6): 691-694.
18. Buyse M, George SL, Evans S, et al. The Role of Biostatistics in the Prevention, Detection and Treatment of Fraud in Clinical Trials. Stat Med. 1999; 18(24):3435-3451.

OVER KANSERLERİNDE 5 YILLIK SAĞ KALIM ANALİZİ

E. Aslaner¹, M.A. Vardar², R. Burgut¹, A. Altıntaş²

*Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı¹
Çukurova Üniversitesi Tıp Fakültesi Kadın Hastalıkları ve Doğum Anabilim
Dalı²*

Amaç: Bu çalışmada over kanserinde, yaş, evre, histolojik tanı ve çeşitli tıbbi tedavilerin sağ kalım süresi üzerindeki etkilerini araştırmayı amaçladık.

Çalışmanın Yapıldığı Yer: ÇÜTF Balcalı Hastanesi Kadın Doğum Kliniği

Yöntem: Çalışma hastanemizde 1999-2004 yılları arasında yatarak tedavi gören 235 hastada yapılmıştır. Cerrahi evreler I ve IV arasında; % 29'u cerrahi, % 70'i cerrahi+KT ve % 1.3'ü diğer tıbbi tedavileri almıştır. Ca125 değerleri her evrede ölçüldü.

Bulgular: Yüksek Ca125 değerlerine Evre III ve IV'de rastlandı. Evre I'deki hastalar halen yaşamaktadırlar. Farklı cerrahi evreler istatistiksel olarak anlamlı derecede sağ kalım süresini değiştirmektedir.

Sonuç: Bu çalışmada Evre I'deki hastaların Evre III ve IV'e göre daha fazla yaşadıklarını gösterdi. Erken tanı, Over kanserinin tedavisi ve sağ kalım süresi üzerinde tartışılmayacak derecede önemlidir.

Anahtar Kelimeler: Over kanseri, Sağ kalım süresi, Erken tanı.

POSTER BİLDİRİLER

TIP EĞİTİMİNDE BİYOİSTATİSTİK

İlker Ercan, Güven Özkaya, Gökhan Ocakođlu, Bülent Ediz, Berna Yazıcı, İsmet Kan

*Uludađ Üniversitesi, Tıp Fakóltesi, Biyoistatistik A.D.
Anadolu Üniversitesi Fen Fakóltesi, İstatistik Bölümü*

Çalışma tıp eğitiminde, biyoistatistik eğitiminin hangi aşamalarda verilmesi gerektiğinin, farklı eğitim düzeyindekilerin biyoistatistik hakkındaki görüşlerinin ve hangi istatistik testler hakkında bilgi sahibi olduklarının incelenmesi amacıyla yapılmıştır.

Araştırmamızda I. sınıf öğrencilerden profesörlük ünvanına kadar 845 denek yer almıştır. Araştırmaya katılanların % 24,7'si sadece uzmanlığın ilk yıllarında, % 15,2'si sadece VI. sınıfta, % 10,2'si ise sadece I. sınıfta dersin verilmesinin uygun olduğunu düşünmektedirler. Dersin I. sınıf ve I. sınıfla birlikte diğer eğitim seviyelerinde de verilmesi gerektiğini düşünenler ise %19,9'dur. Araştırmamızda biyoistatistiğın işlevi hakkında büyük ölçüde bilgisi olduğunu belirtenlerin % 49,7'si sağlık alanında yapılan bir araştırmaya katılmıştır. Bu oran hiç ve kısmen bilgisi olduğunu ifade edenlerin çalışmada bulunma oranlarından anlamlı düzeyde yüksektir ($p<0.001$). Dersi almanın hekimlere mesleğinde yarar sağlayacağı ve biyoistatistiğın tıp biliminde önemli olduğu düşünceleri bakımından hekim ve öğrenci görüşlerinde anlamlı farklılıklar vardır ($p<0.001$). Biyoistatistiğın işlevi hakkında bilgi düzeyi arttıkça biyoistatistiğın tıp biliminde önemli olduğu ve dersi almanın hekimlik mesleğinde yarar sağlayacağı düşüncelerinde artış görülmektedir ($p<0.001$). Biyoistatistiğe verilen önem arttıkça hakkında bilgi sahibi olunan test sayısı da artmaktadır ($p<0.001$). İstatistiksel testler derste anlatılanlar ve anlatılmayanlar, parametrik ve parametrik olmayan, tek değişkenli ve çok değişkenli, örnekleme yöntemleri ve yaşam analizi yöntemleri olarak gruplandırılarak farklı eğitim seviyelerine göre bilinme oranları incelendiğinde anlamlı farklılıklar bulundu.

Sonuçta, öğrencilik yıllarında biyoistatistiğın önemi yeterince anlaşılamamaktadır ve eğitim sürecinin ilerleyen yıllarında biyoistatistiğın işlevi hakkındaki bilgi düzeyi arttıkça verdikleri önem de artmaktadır. Testler hakkında bilgi farklı eğitim seviyelerine göre farklılıklar göstermektedir. Ayrıca biyoistatistik eğitiminin ilk yılda verilmemesi ve uzmanlık sürecinin ilk yıllarında verilmesi gerektiği çoğunluk tarafından düşünülmektedir.

DEĞİŞKEN YAPILARINA GÖRE İSTATİSTİKSEL YAKLAŞIM

Zeki Akkuş, M. Yusuf Çelik

Dicle Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı

e-mail:zakkus@dicle.edu.tr

Özet

Değişken yapıları istatistiksel çalışmalarda birinci adımda dikkate alınması gereken önemli bir durumdur. Araştırmacılar adimsal çözümlmeyi dikkate almadan yaptıkları çözümlmelerde bir çok yanlışlıklarla dolu yayınlar yapmaktadırlar. Değişken yapıları seçilecek olan istatistiksel yöntemi belirler. Konuyu incelikleri ile bilmeyen araştırmacıların değişken yapılarına dikkat etmeleri olası değildir.

Bu çalışmada istatistikte en güçlü varsayım kabul edilen değişken yapılarıyla istatistik yöntemler arasındaki güçlü ilişkiler incelenecektir. Araştırmacılar çalışmalarda bazen tek değişken bazen de çok değişken yapıyla ilgilenirler. Günümüzde çok değişkenli etkenlerin daha çok gözlemlendiği bir gerçektir. Tek değişkenli veya çok değişkenli Biyoistatistik yöntemlerin tümü değişkenlerle ilgili bazı varsayımlar öne sürerler.

Parametrik testlerde ana varsayım normal dağılımdır. Normal dağılımın en önemli özelliği sürekli değişkenlerin bir dağılışı olmasıdır. Tek değişken (univariate) veya çok değişken (multivariate) parametrik yöntemlerin ön gördüğü değişken yapısı sürekli (continuous) değişkenlerdir. Sürekli değişkenlerin normal dağılım gösterdiği bilinmektedir. Sürekli olmayan değişkenler normal dağılım göstermedikleri için "dağılıştan bağımsız" yöntemlerin seçilmesi gerekir.

Parametrik olmayan (non-parametric) yöntemlerin en önemli özelliği araştırmacıya değişken konusunda bir sınırlama getirmemesidir. Araştırmacılar için bu özellik istatistiksel çözümlemede bir kolaylık sağlar. Bu avantajları doğru kullandıkça yöntemlerin sonuçları daha güçlü ve daha duyarlı olacağı bir gerçektir.

Çözümlemeye başlamadan önce değişkenlerin yapıları açık bir şekilde tanımlanması yöntem seçiminde doğru bir sonuca gitmemize olanak tanır.

YAPAY SINIR AĞLARI VE İSTATİSTİK YÖNTEMLERLE İLİŞKİLERİ

Ayşe Canan Yazıcı¹, Ersin Öğüş¹, Handan Çamdeviren², Sinan Canan³

Başkent Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı¹

Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı²

Başkent Üniversitesi Tıp Fakültesi, Fizyoloji Anabilim Dalı³

e-mail: acyazici@baskent.edu.tr

Özet

Yapay sinir ağları (YSA); insan beyninin çalışma mekanizmasını taklit ederek beyin öğrenme, hatırlama, genelleme yapma yolu ile yeni bilgiler türetebilme gibi temel işlevlerini gerçekleştirmek üzere geliştirilen mantıksal yazılımlardır. Çeşitli avantajlar sunan ve sürekli gelişmekte olan bu teknoloji ile istatistik yöntemler arasında önemli ilişkiler söz konusudur. Bu çalışmada YSA'nın genel yapısı ele alınarak istatistik yöntemlerle ilişkileri araştırılmıştır. Pek çok yapay sinir ağı modelinin bazı klasik istatistik yöntemlerin genelleştirilmiş hali olduğu görülmüştür. İstatistikte doğrusal olmayan modellerde kullanılan yöntemler ile ileri beslemeli YSA algoritmaları eğitilerek sonuç üretilebilmektedir. İstatistikte yapay sinir ağları genellikle esnek olarak doğrusal olmayan regresyon analizi veya sınıflandırma modelleri olarak kullanılmaktadırlar. Pek çok YSA algoritmasının parametrik olmayan istatistiksel yöntemlerle yakın ilişkisi vardır.

Anahtar kelimeler: Yapay sinir ağları, doğrusal olmayan modeller, optimizasyon modelleri.

TÜRKİYE'DEKİ ÖZÜRLÜ GRUPLARIN YAPISININ UYUM ANALİZİ İLE İNCELENMESİ

Şengül Cangür, Deniz Sığırlı, Bülent Ediz, İlker Ercan, İsmet Kan

Uludağ Üniversitesi, Tıp Fakültesi, Biyoistatistik A.D.

e-mail: cangur@uludag.edu.tr

Özet

Uyum Analizi (Correspondence Analysis, CA); bir veri matrisini, satır ve sütunları noktalarla gösterilen özel bir grafiksel gösterim tipine dönüştüren açıklayıcı çok değişkenli bir analiz tekniğidir. Uyum analizi, satır profillerindeki homojenlik eksikliği ya da kontenjans tablosunun satır ve sütunları arasındaki bağımlılığı (ya da etkileşimi) daha düşük boyutta açıklama imkanı sağlamaktadır. Ülkemizde nüfusun % 12.29' unu özürlü bireyler oluşturmaktadır. Özürlü bireylerin fiziksel ve psiko-sosyal yapılarının toplum içinde sorunsuz bir şekilde gelişmesine uygun ortam yaratmak, toplumun üzerinde durması gereken en önemli sosyal konulardan birisidir. Ergonomik boyutlarının araştırılmasına yönelik bundan sonra yapılacak projelere ışık tutmak amacıyla, 2002 yılında Devlet İstatistik Enstitüsü ile işbirliği içerisinde gerçekleştirilen Türkiye Özürlüler Araştırması'ndan elde edilen verilere sözü edilen analiz tekniği uygulanarak, özürlü kişilerin yaşamlarını daha çağdaş ve daha insani ortamlarda sürdürebilmeleri amacıyla çeşitli özürlü gruplarının yapıları incelenmiştir.

Anahtar kelimeler: Uyum Analizi; Çoklu Uyum Analizi; Homojenlik Analizi; Özürlü Bireyler.

Kaynaklar

1. **Die, (2004).** Türkiye İstatistik Yıllığı, Ankara: T.C. Başbakanlık Devlet İstatistik Enstitüsü.
2. **Eye Von, A., (1990).** Statistical Methods in Longitudinal Research Volume II, London: Academic Pres.
3. **Gifi, A., (1990).** Nonlinear Multivariate Analysis, New York: John Wiley&Sons.

4. **Greenacre, M.&Hastie, T., (1987)**. The Geometric Interpretation of Correspondence Analysis. *Journal of The American Statistical Association*, 82, 398, 437-447.
5. **Higgs, N.T., (1990)**. Practical and Innovative Uses of Correspondence Analysis. *The Statistician*, 40, 2, 183-194.
6. **Hill, M.O., (1974)**. Correspondence Analysis: A Neglected Multivariate Method. *Applied Statistics*, 23, 3, 340-354.
7. **Jambu, M., (1991)**. *Exploratory and Multivariate Data Analysis*, San Diego: Academic Press, Inc.
8. **Özdamar, K., (2004)**. *Paket Programlar ile İstatistiksel Veri Analizi*, Eskişehir: Kaan Kitabevi.
9. **Palmer, M.W., (1993)**. Putting Things in Even Better Order: The Advantages of Canonical Correspondence Analysis. *Ecology*, 74, 8, 2215-2230.
10. **Panagiotakos, D.B.&Pitsavos, C., (2004)**. Interpretation of Epidemiological Data Using Multiple Correspondence Analysis and Log-linear Models. *Journal of Data Science*, 2, 75-86.

TÜRKİYE VE AVRUPA BİRLİĞİ'NE ÜYE ÜLKELERİN SAĞLIK DÜZEYİ ÖLÇÜTLERİNİN ÇOK BOYUTLU ÖLÇEKLEME ANALİZİ İLE İNCELENMESİ

Deniz Sığırılı, Şengül Cangür, Bülent Ediz, İlker Ercan, İsmet Kan

Uludağ Üniversitesi, Tıp Fakültesi, Biyoistatistik A.D.

e-mail: sigirli@uludag.edu.tr

Özet

Çok Boyutlu Ölçekleme (Multidimensional Scaling-MDS) Analizi, nesne ya da birimler arasında gözlenen benzerlikler ya da farklılıklardan oluşan uzaklık değerlerine dayalı olarak bu nesnelerin tek ya da çok boyutlu uzaydaki gösterimini elde etmeyi amaçlayan, böylece nesnelere arasındaki ilişkilerin belirlenmesini sağlayan bir yöntemdir. p boyutlu bir uzayda n nokta arasındaki benzerlikleri kullanarak bu noktaların birbirlerine yakınlıklarını uzaydaki konumları ile birlikte değerlendirmek ve grafiksel olarak sunumunu sağlamak amacıyla bu yöntemden yararlanılmaktadır. Ülkelerin ulaşmak istedikleri kalkınmışlık düzeyi, günümüzde yeni bir yaklaşımla ele alınmakta ve özellikle sağlık konusu bu yaklaşımda önemli bir yer tutmaktadır. Sağlık sektörünün kalkınma üzerindeki rolünü ön plana çıkaran bu yeni yaklaşım, sektörün önemini daha da artırmakta ve ülkelerin gelişmişlik düzeyinin belirlenmesinde çeşitli sağlık göstergelerinin de kullanılmasına yol açmaktadır. Toplumsal gelişmenin yansımaları olarak kabul edilen bu sağlık göstergeleri bir yandan mevcut durumu yansıtırken öte yandan da ülkelerin önceliklerine ilişkin yol gösterici olmaktadır. Sağlık alanındaki talebin saptanması ve karşılanması, toplumun sağlık sorunlarının belirlenmesi ve bunların çözümü, sağlık hizmetlerinin planlanması ve oluşturulacak sağlık politikalarına yol göstermesi amacıyla; ülkemizin diğer Avrupa Birliği üye ülkeleriyle sağlık göstergeleri bakımından karşılaştırılması ve nasıl bir konumda olduğunun belirlenmesi gerekmektedir. Bu doğrultuda, Dünya Sağlık Örgütü (World Health Organization-WHO) tarafından seçilen sağlık düzeyi ölçütleri dikkate alınarak Avrupa Birliği'ne üye ve üyeliğine başvuruda bulunmuş bazı ülkelere ilişkin verilere çok boyutlu ölçekleme yöntemi uygulanarak, ülkelerin ilgili sağlık göstergeleri bakımından birbirlerine göre konumları incelenmiş ve aralarında var olabilecek benzerlikler/farklılıklar ortaya konulmuştur.

Anahtar kelimeler: Çok Boyutlu Ölçekleme, Sağlık Göstergeleri, Öklid Uzayı.

KLİNİK ARAŞTIRMALARDA, “ÜSTÜNLÜK”, “EŞDÜZEYLİLİK” VE “DAHA DÜŞÜK OLMAMAK” DEĞERLENDİRMELERİ İLKELERİ

F. N. Turan¹, M. Şenocak²

T.Ü. Tıp Fak. Biyoistatistik ABD.¹

İ.Ü. Cerrahpaşa Tıp Fak. Biyoistatistik BD.²

Özet

Klinik araştırmalar genellikle yeni bir ilacın, tedavinin o ana dek kullanılan standart yaklaşımdan daha üstün (superiority) olduğunu kanıtlamak için yapılırlar. Ancak bazı durumlarda da bunun tam tersi olan yeni bir ilacın esasen standart yaklaşıma eşdüzey olduğu (equivalent), ya da bu standart yaklaşımdan daha düşük etkinlikte (non-inferiority) olmadığı kanıtlanmaya çalışılır.

Klasik sıfır varsayımı yaklaşımı ve buna dayalı “p” değerlendirmesi özellikle eşdüzey etkinlik ve daha düşük etkinlikte olmamak durumlarını kanıtlamada yetersiz kaldığından, yargılamaların güvenilir yapılabilmesi için tasarlanan kavram ve yaklaşımlardan birisi de güven aralıkları yaklaşımıdır. Eğer sıfır hipotezi doğru ise oluşan gözlenen olasılık ya da çıktılarıdaki fark yani “p” değeri, klinik deneme sonuçlarını belirtmede sıklıkla kullanılan bir istatistiktir. Ancak “p” değeri bazı durumlarda yetersiz kaldığı için yorumlamada sorunlar yaşanmaktadır. Bu yüzden güven aralığı yöntemi üstünlük, özellikle de eşdüzey etkinlik ve daha düşük etkinlikte olmamak yargılamalarında “p” değerinden daha iyi bilgi verir.

Klinik araştırmalarda yeni bir ilacın ya da tedavinin o ana dek kullanılan standart yaklaşımdan daha üstün olduğunu, en azından eşdüzey etkinlikte olduğunu ya da daha düşük etkinlikte olmadığını yargılamanın en iyi yöntemi klasik sıfır varsayımı yaklaşımı olan “p” değeri yönteminden ziyade güven aralıkları yöntemidir.

PATH ANALİZİNİN, KORELASYON ANALİZİ VE ÇOK DEĞİŞKENLİ REGRESYON ANALİZİ İLE KARŞILAŞTIRILMASI

Rojan Arcak¹, Hasan Akbayın², Tuba Biçer³

Dicle Üniversitesi Ataürk Sağlık Hizmetleri M.Y.O¹

Dicle Üniversitesi Eğitim Fakültesi Ortaöğretim Fen ve Matematik Alanlar Bölümü²

Dicle Üniversitesi Ziraat Fakültesi Tarla Bitkileri Bölümü³

Bu çalışmanın amacı Path Analizi, Çok Değişkenli Regresyon Analizi ve Korelasyon Analizi yöntemlerini karşılaştırmak ve mercimekte tane verimini etkileyen bazı morfolojik değişkenleri kullanarak açıklamaya çalışmaktır.

Bilindiği gibi Path Analizi değişkenler arasındaki korelasyonu bileşenlerine ayıran ve değişkenler arasındaki doğrudan ve dolaylı etkileri belirleyen bir analiz yöntemidir. Değişkenler arasındaki ilişkinin ölçüsü korelasyon katsayısıdır. İki değişken arasındaki korelasyon katsayısı bu iki değişkenin birlikte değişim derecesidir. Yani iki değişken arasında hesaplanan korelasyon katsayısı yüksek ise, bu iki değişkenin birbirine bağlı olduğunu ve birlikte değiştiğini söyleyebiliriz. Ancak, iki değişken arasında hesaplanan korelasyon katsayısı başka bir değişken ya da değişkenler tarafından etkileniyor ise, yani iki değişken arasındaki sebep-sonuç ilişkisi üçüncü bir değişkenin etkisine bağlı ise, korelasyon katsayısı bu ilişkiyi açıklamada yeterli değildir. Ayrıca sistemde bunlar ile ilişkili olduğu düşünülen başka değişkenlerin etkisi olabilir. Değişkenler arasında hesaplanan korelasyon katsayısında diğer değişkenler ile olan ilişkiden kaynaklanan kısımların bulunması istendiğinde Path Analizi kullanılmaktadır. Path katsayıları kısmi regresyon katsayılarının hesaplanmasıyla bulunur.

Değişkenler arasındaki sebep-sonuç ilişkilerinin belirlenmesi ve bu ilişkiler sisteminden elde edilen sonuçların yorumlanmasına Path Analizi tekniği önemli bir katkı sunmaktadır.

Bu çalışmada Path Analizi tekniğinin matematiksel alt yapısının hemen her alandaki araştırmacılar tarafından anlaşılabilir ve yorumlanabilir hale getirilmesine ve Korelasyon Analizi ile Çok Değişkenli Regresyon'la ilişkileri içinde ele alınarak açıklanmasına çalışılmıştır.

Path Analizi'ni bir uygulamayla örneklendirmek amacıyla, mercimeğin bazı morfolojik değişkenlerinden hareketle mercimekte tane veriminin tespitinde Path Analizi yöntemi kullanılmış ve sonuçlar detaylı olarak verilmiştir. Önce Path şeması kurulmuş, sonra her bir değişkenin bitki tane verimi ile arasındaki korelasyon Doğrudan (DE), Dolaylı (IE), S ve U etkilerine ayrılmıştır. Kullanılan morfolojik değişkenlerin hangi yollarla ne kadar etkilendiği tespit edildikten sonra, bu sonuçlar her değişken için yorumlanmıştır.

Sonuç olarak bitki boyu ile bitki tane verimi arasındaki korelasyon 0,539 iken bunun sadece % 32'sinin bitki boyundan kaynaklandığı, iki değişken arasındaki ilişkinin başka değişkenler üzerinden olduğu belirlenmiştir. Dal sayısı ile bitki tane verimi arasında pozitif bir 0,371'lik korelasyon olmasına karşın, dal sayısı bu korelasyona -0,063'lük Doğrudan Etki (DE) ile negatif bir etki yapmaktadır.

Bakla sayısı ile verim arasında 0,929'luk yüksek ve istatistiksel olarak önemli bir korelasyon bulunmasına karşın, Path Analizi sonuçları bize, bakla sayısının Doğrudan Etkisi'nin sadece 0,357 olduğunu, en fazla etkinin tane sayısı ile arasındaki U etkisinden kaynaklandığı görülmektedir. Path Analizi sonuçları, iki değişken arasındaki ilişkinin başka değişkenler üzerinden gerçekleştiğini göstermektedir. Çok değişkenli bir analiz olan Path Analizi yöntemini kullanmayıp sadece korelasyon analizi ile yetinmiş olsaydık, epeyce yanıltılmış olacaktık. Bu nedenle mümkün ve uygun olduğunda tek değişkenli tekniklerin kullanılması araştırmalardan elde edilecek sonuçlar açısından daha sağlıklı olacaktır.

Anahtar kelimeler: Path Analizi, Korelasyon Analizi, Çoklu Regresyon Analizi.

DÖNGÜSEL SÜREÇLERDE OLAY KESTİRİMİ: MARKOV MODEL YAKLAŞIMI

Necdet Süt¹, Mevlüt Türe², Mustafa Şenocak¹

¹*İstanbul Üniversitesi, Cerrahpaşa Tıp Fakültesi, Biyoistatistik B.D.*

²*Trakya Üniversitesi, Tıp Fakültesi, Biyoistatistik A.D.*

e-mail: necdetsut@yahoo.com

Özet

Bu çalışmada Markov modelin tıpsal karar vermede ve kestirimde bulunmada kullanımının Akut Lenfoblastik Lösemiye ilişkin örnek bir uygulama ile gösterilmesi amaçlandı.

Markov model, olayların zamanlaması önemli olduğunda, zaman aşımı karar probleminde risk yaratıyorsa ve olaylar birden fazla kez tekrar edebiliyorsa (recurrence - nüks gibi) yararlı olan bir modellemedir. Markov Modellerin karar ağacı modellerinden farkı, belirsiz olayları şans düğümlerinde modelleme yerine tanımlı sağlık durumları arasında geçişler olarak modellemesidir. Ayrıca modelin döngüsel yapısı gelecekte olabilecek olayları tahmin etmek için oldukça faydalıdır. Model matris cebiri şeklinde Kohort yada Monte Carlo benzetimi olarak değerlendirilebilir.

Sonuç olarak, Markov modeller yardımıyla (alternatif) klinik stratejilerinin maliyet, etkinlik ve yaşam kalitesi ölçütlerini sentez edilerek, bunların sonucunda yaşam beklentisi, kalite düzeltmeli yaşam beklentisi ve yaşam maliyeti hesaplanabilir.

Anahtar Kelimeler: Markov model, döngüsel süreç, tıpsal karar.

Giriş

Modelleme kavramı farklı disiplinlerde kullanılmaktadır. "Araştırma Çıktılarının Değerlendirilmesinde (Outcomes Research)" model, tedavi yöntemlerinin hasta/toplum açısından çeşitli durumlarda sağlık ve ekonomik sonuçlarının gösterildiği matematiksel yapı olarak ifade edilebilir. Diğer bir deyişle ilgilenilen özel bir fenomenin (hastalık, tedavi vb.) aslına çokbenzeyen bir durumunu yaratmak için değişkenler arası ilişkilerin sistematik yapılarıyla ilgilenildiği analitik yapıdır. Model yapılarıyla fenomenin ana parametrelerinin gelecekteki değerleri tahmin edilebilir ve bu parametreler

arasındaki ilişkiler ortaya konulabilir, ayrıca modelde zaman boyutu da yer alabilir (1, 2, 3).

Beck&Pauker 1983 yılında karar ağacı modellerinden farklı olan Markov modellerinin tıbbi uygulamalarda prognozu belirlemek için uygulanabilirliğini tanımladıklarından itibaren Markov modeller artan bir hızla sağlık alanında kullanılır bir konuma gelmiştir (3).

Dünya literatüründe tıpsal karar vermede ve kestirimde bulunmada oldukça yaygın olarak kullanılan ancak ülkemiz literatürü tarandığında tıpsal alanda uygulaması bulunamayan Markov modelin döngüsel süreçlerde tıpsal karar vermede ve geleceğe yönelik kestirimde bulunmada kullanımını Akut Lenfoblastik Lösemiye ilişkin örnek bir uygulama ile göstermeyi amaçladık.

Yöntem

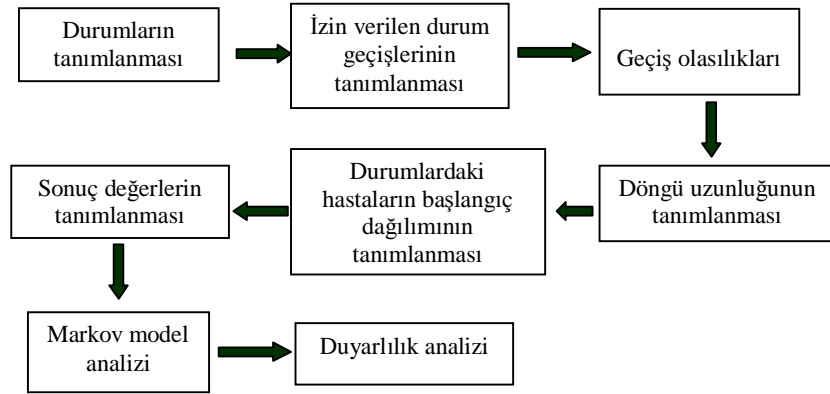
Markov Model, sonlu sayıda durumla ifade edilebilen bir olasılıksal karar sürecinde bu durumlar arası geçişlerin olasılık değerleriyle –sürecin bir durumundan diğer bir durumuna geçmesi olasılığı– ifade edildiği yapıdır. Amaç incelenen sorunun beklenen sonucuna ilişkin optimum yapıyı belirlemektir (3, 4).

Model hastayı kesikli sağlık durumunda ele almaktadır ve bu kesikli durumlar 'Markov durumları' olarak adlandırılır. Durum kavramıyla, Markov sisteminde var olan olası tüm sağlık sonuçları ifade edilmektedir. Markov sistemi belli bir zamanda sadece bir sağlık durumunu göstermektedir. Markov modelinde durumlar dairesel sembollerle gösterilmektedir. Markov modelinde geçişler, dairesel sembollerle gösterilen Markov durumlarının geçiş yönünü gösteren oklarla birleştirilmesiyle ifade edilmektedir. Geçiş, Markov diyagramında bir sağlık durumundan diğer bir sağlık durumuna geçilmesi olasılığını göstermektedir. İyi bir sağlık durumundan kötü bir sağlık durumuna geçiş 'başarısızlık oranıyla', kötü bir sağlık durumundan iyi bir sağlık durumuna geçiş 'onarım (iyileşme, şifa) oranıyla' ifade edilmektedir (3, 5, 6, 7, 8, 9). Markov model bir Kohort ya da Monte Carlo benzetimi gibi cebirsel olarak matrisle değerlendirilebilir. Ayrıca Markov modeller yardımıyla klinik olayların gösteriminde bir döngü ağacı kullanılarak 'Markov döngü ağacı' bir Kohort ya da Monte Carlo benzetimi gibi değerlendirme yapılabilir (3, 5).

- 1- **Kohort benzetimi:** Hastalık durumlarının herhangi birinde hipotetik bir kohort hasta grubu ile başlanan modelde kohort model süresince izlenir. Zamanın herhangi bir anında herhangi bir durumda kohort oranı ve her bir durumun ortalama süresi hesaplanabilir.
- 2- **Monte-Carlo benzetimi:** Çok sayıda hipotetik hasta model sürecinde tek tek geçer ve bunların geçiş yolları kaydedilir. Monte-

Carlo yaklaşımının avantajı, ayrılma (dispersiyon) süresinin tahmininin tek tek simülasyon verilerinden elde edilebilir olmasıdır.

Markov model, zamana bağlı olarak tekrarlayan olayların gösterimini ve bu olaylarla ilgili klinik durumların olasılık ve yararlarının daha hızlı gösterimini sağlamaktadır (3, 5, 10, 11, 12). Markov modelin, geliştirme aşamaları aşağıda Şekil 1'de gösterilmektedir.



Şekil 1. Markov model geliştirme aşamaları

Herhangi bir hastalığın K farklı aşamasında seyrinin N bağımsız birimde incelendiğini düşünelim. Bunu yapabilmek için K aşamalı olasılıksal sürecin (stochastic process) düzenlenmesi gerekmektedir.

Gözlemler serisinde her bir birim i ile simgelenerek aşağıdaki gibi ifade edilir,

$$D_{ij} = (X_{ij}, T_{ij})_{\substack{i \in \{1, \dots, N\} \\ j \in \{1, \dots, n_i\}}}$$

formüldeki

n_i : i bireyinin örnek (model) sayısıdır,

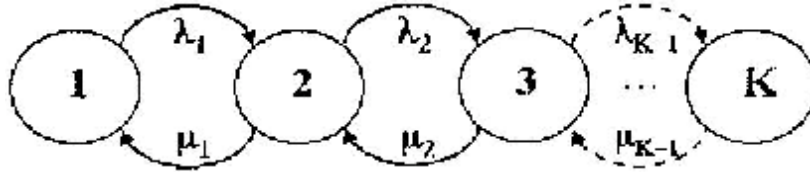
X_{ij} : j. örnekte (modelde) i bireyinin gözlemlenen durumunu

T_{ij} : j.vizitte X_{ij} aşamasında gözlemlenen i bireyi zamanını

ifade etmektedir.

$\theta = (\lambda, \mu)$ olarak ele alındığında, $\lambda(\lambda = \lambda_1, \dots, \lambda_{K-1})$ ve $\mu(\mu = \mu_1, \dots, \mu_{K-1})$

ardışık aşamalar arasındaki geçiş olasılıklarını ifade etmektedir. Homojen zaman dilimlerinde izlem yapılmış K geriye dönülebilir aşamalı uzunlamasına bir Markov modellemesinin geçiş grafiği Şekil-2'de gösterilmiştir (6).



Şekil 2. K aşamalı geriye dönülebilir Markov modelinde geçişlerin grafiksel gösterimi

Şekil-2'deki düğümler ilgilenilen olayın (hastalık, tedavi v.b.) K aşamasını, oklar ilgilenilen geçiş oranlarının olası geçiş yönünü göstermektedir. λ_k , k durumundan $k+1$ durumuna geçiş oranını ve μ_k $\{1, \dots, (K-1)\}$ 'deki tüm k değerleri için $k+1$ durumundan k durumuna onarım (iyileşme, şifa) oranını ifade etmektedir.

K durumdan oluşan $S = \{1, 2, \dots, K\}$ bir Markov modelde düğümler arası geçiş ve onarım oranlarının yer aldığı geçiş matrisi Λ 'nın açık şekilde ifade edilişi aşağıdaki gibidir (6),

$$\Lambda = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \dots & 0 \\ \mu_1 & -(\mu_1 + \lambda_2) & \lambda_2 & 0 & \vdots \\ 0 & \mu_2 & -(\mu_2 + \lambda_3) & \lambda_3 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 0 \\ \vdots & \vdots & \vdots & 0 & \mu_{k-2} & -(\mu_{k-2} + \lambda_{k-1}) & \lambda_{k-1} \\ 0 & \dots & \dots & 0 & 0 & \mu_{k-1} & -\mu_{k-1} \end{pmatrix}$$

Matrisin (l, m) elemanı aşama l 'den aşama m 'ye geçiş yoğunluğunu göstermektedir.

$P(t) = (P_{l,m}(t))_{l,m}$ geçiş olasılık matrisidir. Matristeki $P_{l,m}(t)$ ifadesi $\{1, \dots, K\}$ 'daki tüm l, m için t zamanında l durumundan m durumuna geçiş olasılıklarını gösterir. Geçiş olasılık matrisi $P(t)$ ve geçiş yoğunluk matrisi Λ arasındaki ilişki Kolmogorov ileri diferansiyel denklemiyle,

$$\frac{\partial P(t)}{\partial t} = P(t)\Lambda$$

şeklinde ifade edilir (6). Diferansiyel denkleminin tekil çözümü bu sistemde $P(t) = \exp(\Lambda t)$ olarak ifade edilebilir, sınır koşullara bağlı olarak $P(0) = I$ formüldeki I birim matristir. Homojenitenin ve sürecin tersinebilirliğinin benimsenmesiyle Kolmogorov denkleminin çözümü aşağıdaki denklemle verilir,

$$P(t) = P \text{diag}\{\exp(\rho_1 t), \dots, \exp(\rho_k t)\} P^{-1}$$

Geçiş matrisi Λ 'nın özdeğerleri $(\rho_k)_k$ ile gösterilir, P , k sütunlu kare matristir, ρ_k için sağ özvektör ve $\text{diag}\{\exp(\rho_k t)\}_k$ tüm k 'lar için $\{1, \dots, K\}$ 'da $(\exp(\rho_k t))_k$ katsayılarıyla diagonal matrise uygundur.

Bu nedenle geçiş yoğunluklarına göre geçiş olasılıklarının fonksiyonel şekli

$$P_{l,m}(t) = \frac{\sum_{k=1}^K A'_{lm}(\rho_k) \exp(\rho_k t)}{\prod_{n=1, n \neq k}^K (\rho_k - \rho_n)} \quad \forall k, l, m, n \in \{1, \dots, K\}$$

formülü ile ifade edilir. Formüldeki $A'_{lm}(\rho_k)$, $(\rho_k I - \Lambda)$ transpose matrisinin (l, m) elemanının kofaktörüdür (eşçarpan). Λ geçiş matrisinin özdeğerleri $-\Lambda$ karakteristik polinomunun kökleri- yaklaşık olarak ikiye bölünür, özdeğerlerin birinin sıfır olduğu bilindiğinden diğerleri gerçek sayı ve negatiftir. Bu olasılıklar t sonsuz eğilimde olduğunda bir eşik sınır değere doğru yakınsaktırlar (6).

Bulgular

Akut Lenfoblastik Lösemiye ilişkin Markov model kurgulaması

Markov modelin sağlık alanında uygulanabilirliğini göstermek amacıyla literatür bilgileri (13, 14) ve uzman görüşleri doğrultusunda St Jude Toplam Terapi XIII B protokolünün Akut Lenfoblastik Lösemi (Acute Lymphoblastic Leukemia - ALL) hastalığında tedavi etkinliğine ilişkin bir model öngörülmüştür.

Model oluştururken öncelikle ALL'ye ilişkin sonuç durumlar saptanmıştır (13).

- I- *Remisyon*
- II- *Relaps*
- III- *Dirençli seyir*
- IV- *Ex*

Literatür bilgileri (13) ve uzman görüşleri doğrultusunda hastalığın sonuç durumları arasında izin verilen durum geçişleri aşağıdaki şekilde olabilmektedir.

Remisyon	→	Remisyon / Relaps / Ex
Relaps	→	Remisyon / Ex
Dirençli seyir	→	Remisyon / Ex

Geçiş matrisi Δ 'yı belirleyebilmek için ALL hastası bir kohortun 5 yıllık izlem sonucu elde edilen bulguları kullanılmıştır. Geçiş matrisi aşağıda ifade edilmektedir.

	Remisyon	Relaps	Dirençli Seyir	Ex
Remisyon	0,65	0,27	0	0,08
Relaps	0,24	0	0	0,76
Dirençli Seyir	0	0	0	1
Ex	0	0	0	1

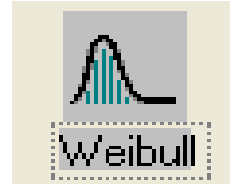
Kurguladığımız çözümlemde Markov modelin döngü uzunluğu olarak 5 yıl öngörülmüştür. Ayrıca modele girecek ALL hastalarının başlangıç durum dağılımı olarak Weibull dağılımı öngörülmüştür.

Weibull Dağılımı

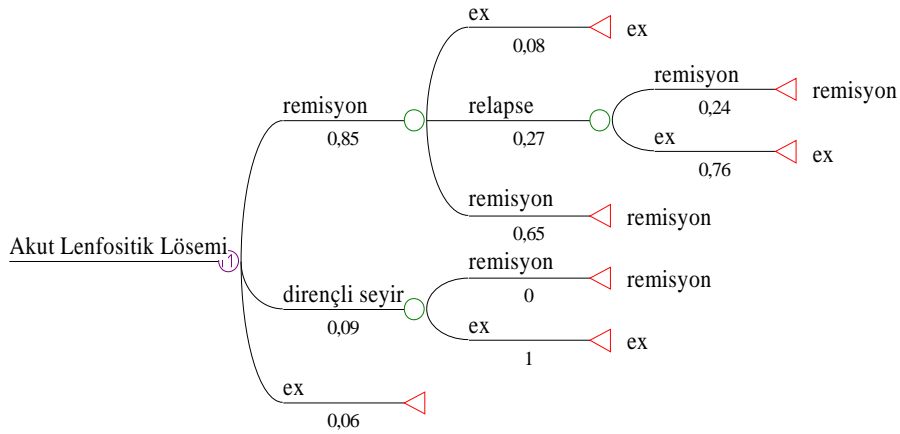
$$f(x; a, c) = 1 - e^{-(x/a)^c}$$

$$x > 0; a, c > 0$$

a, c; Weibull dağılımının parametreleri



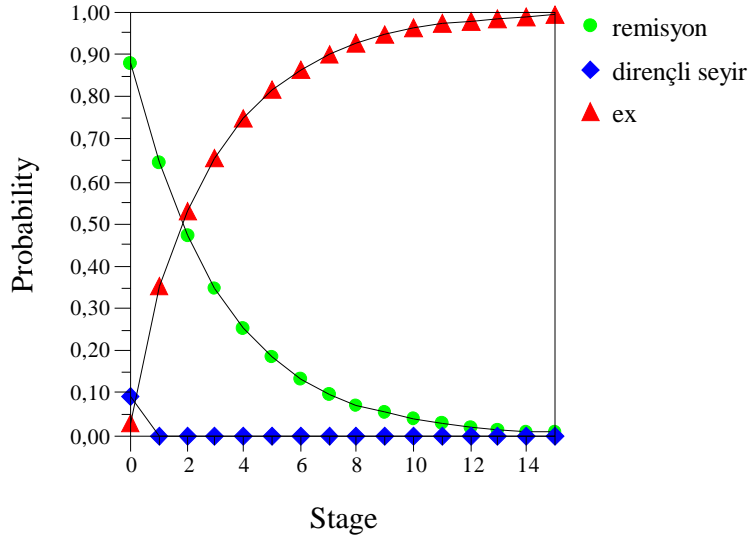
ALL'ye ilişkin Markov Döngü ağacı aşağıdaki şekil 3'te gösterilmektedir. Şekilde ALL kohortunun her bir döngüdeki dağılımı, sonuç durumlar arasındaki olası geçiş yolları ve geçiş olasılıkları yer almaktadır.



Şekil 3. Markov döngü ağacı (M, markov düğümü)

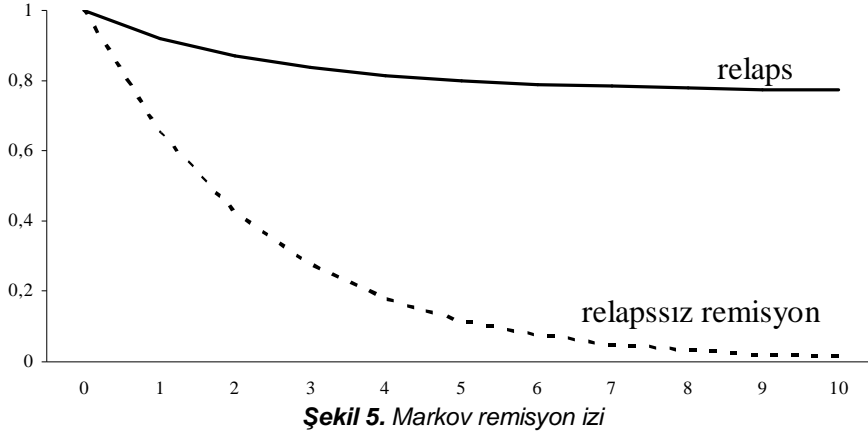
Şekil-4, Markov döngü ağacının (Şekil-3) çözümlemesi sonucu markov durumlarına ilişkin olarak elde edilen ilk 15 döngüye ilişkin kohortun her bir döngüdeki dağılımının grafiksel gösterimini ifade etmektedir. Şekle göre döngü uzunluğu arttıkça (>2) söz konusu kohortun ex olma olasılığının remisyonda kalma olasılığından daha yüksek olduğu gözlenmektedir. Dirençli seyreden olguların tümü izlem döneminde remisyona girmeden ex oldukları için bu markov durumunun diğer döngülerdeki olasılık değerleri 0 olarak yer almaktadır.

Markov Probability Analysis



Şekil 4. ALL Kohortunun Markov olasılık çözümlemesi

ALL hastalığında amaç olguyu relaps ya da ex olmadan remisyonda tutabilmektir. İzlenen kohortun verileri doğrultusunda başlangıç durumu remisyon olan 10000 denemelik bir kohort benzetim çözümlemesi düzenlenerek 10 döngü sürecinde olguların remisyonda kalma beklentisi hesaplanmıştır. Şekil 5'te söz konusu tedavi uygulaması ile döngülerde Relaps olmadan remisyonda kalma ve relaps olma olasılıkları yer almaktadır. Şekle göre 2. döngüde herhangi bir olgunun relapssız remisyonunda kalma olasılığı % 43'tür, bu döngüden sonra olasılık artan bir hızda düşmektedir.



Şekil 5. Markov remisyon izi

Tartışma

Tedavi yöntem/rejimlerine bağlı olarak hastalıkların gelecekteki seyri çok büyük önem taşımaktadır. Hastalık sonuç değerlerinin gelecekte ne olabileceği –incelenen olay ölümcül sonuç değer oluşturabiliyorsa yada toplumda nadir gözleniyorsa, AIDS gibi– yönünde öngörülerde bulunabilmek tıpsal karar vericiler ve hastalar için büyük önem taşımaktadır. Bu nedenle Markov modeller, zamana bağlı olarak tekrarlayan olayların ve bu olaylarla ilgili klinik durumların olasılık ve yararlarını ele almak için tıpsal karar vericilere yol gösteren bir araç olarak kullanılabilirler. Model, hastalıklar zaman sürecinde birbirini izleyerek gözlemleniyorsa –örneğin stabil anjinal hastaların akut MI geçirmesi yada lokal meme kanserli hastalarda kanserin yeniden nüks etmesi gibi– bu hastalıklara ilişkin kestirimlerde bulunmaya olanak tanır. Modelin döngüsel yapısı yardımıyla belli bir zamanda veya zaman sürecinde meydana gelecek olaylar tahmin edilebilir.

Sonuç olarak Markov modeller yardımıyla (alternatif) klinik stratejilerin yaşam beklentisi, kalite düzeltmeli yaşam beklentisi (QALE-quality-adjusted life expectancy) ve yaşam maliyetleri hesaplanarak, maliyetler, etkiler ve sağlıkla ilgili yaşam kaliteleri değerlendirilebilir. Ayrıca model Monte Carlo ve Kohort benzetimleri yardımıyla nadir gelişen olayları makul bir zaman süreci içerisinde ele almak için uygun bir yöntemdir.

Kaynaklar

1. Drummond M.F., McGuire A. Economic Evaluation in Health Care, Oxford University Press, 2001.

2. Buxton M.J., Drummond M.F., Van Hout B.A., Prince R.L., Sheldon T.A., Szucs T., Vray M. Modelling in Economic Evaluation: An Unavoidable Fact of Life Health Economics, 6: 217–227, 1997.
3. Sonnenberg F.A., Beck J.R. Markov Models in Medical Decision Making: A Practical Guide, Medical Decision Making, 13:322-338, 1993.
4. Taha H.A. Operation Research an Introduction, Prentice-Hall, NJ, 1997.
5. Hunink M., Glasziou P. Decision making in health and medicine, Cambridge University Pres, 2001.
6. Kousignian I., Autran B. , Chouquet C., Calvez V., Gomard E., Katlama C., Riviere Y., the IMMUNOCO Study group, Costagliola D. Markov modelling of changes in HIV-specific cytotoxic T-lymphocyte responses with time in untreated HIV-1 infected patients, Statistics in Medicine, 22:1675 –1690, 2003.
7. Karnon J. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation, Health Economics, 12: 837–848, 2003.
8. Briggs A., Sculpher M. An introduction to Markov modelling for economic evaluation, Pharmacoeconomics, 13(4):397-409, 1998.
9. Michaels JA, Galland RB. Management of asymptomatic popliteal aneurysms: the use of a Markov decision tree to determine the criteria for a conservative approach. Eur J Vasc Surg 7:136-43, 1993.
10. McLaughlin F.S., Pickhardt R.C. Quantitative Techniques for management decisions, Houghton Mifflin Company, Bostyon, 1997.
11. Plunkett BA, Grobman WA. Routine hepatitis C virus screening in pregnancy: A cost-effectiveness analysis, American Journal of Obstetrics and Gynecology, 192: 1153–1161, 2005.
12. Empanan C., Wolters H., Laukötter M., Dame C., Senninger N. Cost-Effectiveness Analysis of Basixilimab Induction and Calcineurin-Sparing Protocols in “Old to Old” Programs Using Markov Models, *Transplantation Proceedings*, 35:1324-1325, 2003.
13. Hoelzer D., Gökbüget N., Ottmann O., Pui CH., Relling MV., Appelbaum FR., van Dongen JJM. and Szczepański T. Acute Lymphoblastic Leukemia, Hematology, 162-192, 2002.
14. Pui CH, Sandlund JT, Pei D, Campana D, Rivera GK, Ribeiro RC, Rubnitz JE, Razzouk BI, Howard SC, Hudson MM, Cheng C, Kun LE, Raimondi SC, Behm FG, Downing JR, Relling MV, Evans WE. Improved outcome for children with acute lymphoblastic leukemia: results of Total Therapy Study XIIIB at St Jude Children's Research Hospital, Blood, 1;104(9): 2690-6, 2004.

18 YAŞ ÜZERİ SAĞLIKLI BİREYLERDE KANTİTATİF ULTRASON ÖLÇÜMLERİNİN İNCELENMESİNDE DEĞİŞİM KATSAYISININ KULLANILMASI

Başak Gürtekin, Sevda Özel, Rian Dişçi

*İstanbul Üniversitesi İstanbul Tıp Fakültesi, Biyoistatistik ve Demografi Bilim
Dalı*

e-mail: bgurtek@istanbul.edu.tr

Özet

Çalışmamızda, Tıp'ta, Fizik Tedavi alanında kemik kalitesini belirlemede topuktan yapılan tekrarlı ölçümlerle elde edilen Kantitatif Ultrason (KUS) değerlerindeki değişkenliğin ve tekrarlı ölçüm sayısının, değişim katsayısı kullanılarak belirlenmesi amaçlanmıştır.

Çalışma 20.10.2004-27.12.2004 tarihleri arasında İstanbul Üniversitesi Fizik Tedavi ve Rehabilitasyon Yüksekokulu'nda gerçekleştirilmiştir. Topuk ölçümleri Fizik Tedavi ve Rehabilitasyon Yüksekokulu'na ait Sahara Klinik Kemik Sonometresi (Hologic, Waltham, MA, USA) ile 18 yaşın üzerinde 20 kadın ve 20 erkek bireye iki ayrı gözlemci tarafından yapılmıştır. Ölçümler, ilk 5'i 1. gözlemci, ikinci 5'i 2. gözlemci tarafından olmak üzere tek bir topuktan toplam 10 kez tekrarlanmıştır.

İki gözlemci tarafından gerçekleştirilen ölçümlerde KUS değişkenlerinden biri olan BUA için bireyin kısa dönem kesinliği ($SD_{kesinlik}$) ve bu kesinliğin yüzde olarak ifadesi (CV_{SD}) kadınlarda ve erkeklerde ayrı ayrı hesaplanmıştır.

Bu çalışmada iki gözlemci arasında fark olduğu ve tekrarlı ölçüm sayısının artırılmasının değişkenliği azaltmada etkili olmadığı, uygun tekrarlı ölçüm sayısının ise 2 olmasının yeterli olacağı kanısına varılmıştır.

Anahtar kelimeler: Değişim katsayısı, kantitatif ultrason, topuk.

Giriş

Bir serinin tüm değerlerini içeren dağılım ölçülerinden biri de değişim katsayısıdır. Standart sapmanın büyük olması popülasyonu meydana getiren değişkenlerin kendi ortalamalarından olan sapmalarının fazla olması anlamını taşımaktadır. Ancak bazı hallerde, standart sapma, dağılımı gerçek

anlamda yansıtmayabilir. Bu durum, özellikle iki popülasyon dağılımının birbiriyle karşılaştırılmasında karşımıza çıkmaktadır.

İki popülasyondan birini meydana getiren gözlemlerin değer bakımından diğer popülasyonu teşkil edenlerden daha büyük olması yanında standart sapması da diğerinden daha büyük olacaktır. Halbuki gerçekte belki de, o popülasyonu meydana getiren gözlemlerin kendi ortalamalarından olan sapmaları diğer popülasyondakinden daha büyük değildir. İşte standart sapmanın bu sakıncasını gidermek için, dağılımın diğer bir ölçüsü olarak σ/μ şeklinde ifade edilen “değişim katsayısı” kullanılmaktadır.

Değişim katsayısı kavramı, istatistikçi Karl Pearson (1857-1936) tarafından 1895’te ortaya atılmıştır

(1). Örnekten hesaplandığında ve % olarak ifade edilmek istendiğinde

$$D.K. = \frac{s}{\bar{x}} \times 100 \quad (1)$$

şeklinde formüle edilebilir (2).

s : standart sapma

\bar{x} : aritmetik ortalama

Kısaca, değişim katsayısı birim sayıları ve ölçü birimleri birbirlerinden farklı olan değişkenlerin ortalamaya göre yayılışlarını karşılaştırmak için yararlanılan bir dağılım ölçüsüdür. Ölçü birimi yoktur.

- DK % < %5 ise değişkenin; homojen, dar bir değer aralığında yayılma gösterdiği ve X’in türdeş dağılıma sahip olduğu varsayılır.

- % 5 ≤ DK % < %10 ise X’in normal genişlikte sayılabilecek bir dağılıma sahip olduğu varsayılır.

- % 10 ≤ DK % < %20 ise değişkenin türdeş dağılımının kaybolmaya başladığı varsayılır.

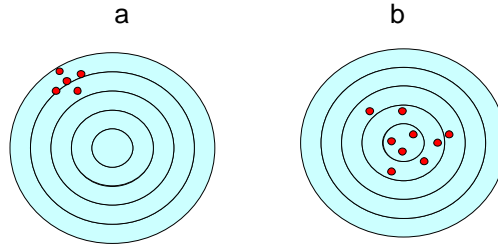
- DK% ≥ %20 ise değişkenin türdeş olmayan bir dağılıma sahip olduğu ya da verilerin hatalı toplandığı varsayılır (3).

Değişim katsayısı, Tıp’ta Fizik Tedavi alanında kemik kalitesini belirlemek amacıyla topuktan yapılan Kantitatif Ultrason ölçümlerindeki değişkenliği dolayısıyla ölçümün kesinliğini belirlemek için kullanılmaktadır.

Aynı bireyde topuktan yapılan ölçümler her seferinde farklı sonuçlar verdiği için ölçüm birkaç kez yinelenmektedir. Çalışmamızda da bu kesinliği belirlemek ve kaç ölçümde en iyi sonucun alınacağını bulmak amaçlandığı için değişim katsayısı kullanılmıştır.

Kesinlik (Presizyon) Hataları

Kesinlik hataları (precision errors), tanısal bir tekniğin tekrarlanabilirliğinin belirlenmesi için tanımlanır. Doğruluk hataları da (Accuracy errors), ölçülen sonuçların “gerçek” değerlerden ne derecede saptığını yansıtır. Kesinlik ve doğruluk arasındaki farkı anlatmak için sıklıkla bir okçunun performansı örneği verilir (Şekil 1a, b).



Şekil 1. a) Kesinlik

b) Doğruluk

Yukarıdaki şekilde de (Şekil 1b) görüldüğü gibi bir okçu, hedefe yakın dairelere ok atıyor ve dairelerin dışına çıkmıyorsa doğruluğu yüksek ancak kesinliği düşük atışlar yapıyordur. Okçu hedefe uzak bir yere ok atıyor ama atışlarında hep aynı noktayı tutturuyorsa doğruluğu düşük ancak kesinliği yüksek atışlar yapıyordur (Şekil 1a).

Bireyin Kısa Dönem Kesinliği

Bir bireydeki tekrarlanan ölçümlerin rassal değişimlerinin normal dağıldığı varsayıldığında kesinlik, Gaussian olasılık dağılımında σ parametresinin tahmini ile temsil edilir. Kısa dönem kesinlik (SD), j deneğindeki $i = 1, \dots, n_j$ 'ye kadar olan tekrarlı ölçümlerin standart sapması olarak tanımlanır(4).

$$SD_j = \sqrt{\frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{n_j - 1}} \quad (2)$$

n_j : ölçüm sayısı

x_{ij} : j deneğinde yapılan i. ölçümün sonucu

\bar{x}_j : j deneğindeki tüm x_{ij} 'lerin ortalaması

Ölçümün gerçek ortalaması bilinmediğinden ve n kere tekrarlanan ölçümlerin ortalamasından tahmin edileceğinden SD^2 'nin Gaussian olasılık dağılımında σ^2 parametresinin sapmasız tahmincisi olarak değerlendirilebilmesi için paydanın (n_j-1) olarak ifade edilmesi gerekir. Bu uyarlama SD 'nin ölçümlerinin tekrarlanma sayısından bağımsız olduğunu garanti eder. Paydadaki (n_j-1) yani (tekrarlanan ölçüm sayısı-1) değeri bu değerlendirme ile ilişkili serbestlik derecesi (s.d.)'dir.

Kesinlik hataları, tam sayılarla ya da tekrarlanan ölçümlerin tipik olarak yüzdelerle verilen değişim katsayısı (CV) formülü ile ifade edilir (4).

$$CV_{SD_j} = \frac{SD_j}{\bar{x}_j} \times 100\% \quad (3)$$

\bar{x}_j : Tüm x_{ij} 'lerin ortalaması

Genel uygulamanın aksine kesinlik hatasının doğru değerlendirmesi, deneklerdeki bireysel kesinlik hatalarının aritmetik ortalaması ile değil varyanslarının (SD^2) aritmetik ortalaması ile tanımlanır (4).

$$SD^2 = \left(\sum_{j=1}^m SD_j^2 \right) / m \quad (4)$$

İstatistiksel olarak bunun nedeni, standart sapmanın değil varyans (SD^2)'nin Gaussian normal dağılımının σ^2 parametresinin sapmasız tahmini olarak sayılabileceğidir. Bunun sonucunda her bir m deneği için (2) numaralı denklem ile hesaplanmış olan kesinlik hataları aşağıdaki ifade ile verilmektedir (4).

$$SD = \sqrt{\left(\sum_{j=1}^m SD_j^2 \right) / m} \quad (5a)$$

Her bir birey için iki ölçüm yapılırsa (birinci ve ikinci sonuç arasındaki fark d_j olarak gösterilirse) aşağıdaki formül uygulanır;

$$SD = \sqrt{\left(\sum_{j=1}^m d_j^2 \right) / 2m} \quad (5b)$$

Kesinliđi yüzde olarak göstermek istersek řu formülün kullanılması önerilir;

$$C V_{SD} = \left(SD / \left[\left(\sum_{j=1}^m \bar{x}_j \right) / m \right] \right) \times 100 \% \quad (6)$$

Gereç ve Yöntem

Bu çalışmada topuktan yapılan ölçümler (20.10.2004 – 27.12.2004) tarihleri arasında İstanbul Üniversitesi Fizik Tedavi ve Rehabilitasyon Yüksekokulu'nda gerçekleştirilmiştir. Çalışmaya 18 yaşın üzerinde 20 kadın ve 20 erkekten oluşan toplam 40 kişi katılmıştır. Ölçümlerin yapılmasını sağlayan topuk KUS aleti İstanbul Üniversitesi Fizik Tedavi ve Rehabilitasyon Yüksekokuluna ait Sahara Clinical Bone Sonometer (Hologic, Waltham, MA, USA) idi (Şekil 2).



Şekil 2. Sahara Clinical Bone Sonometer
(Sahara Klinik Kemik Sonometresi) resmi

Bu çalışmada bireylere konu hakkında bilgi vererek çalışmaya gönüllü olarak katılmaları sağlanmıştır. Ölçümlere geçmeden önce bireyde dominant olma durumuna bağlı olarak hangi ayaktan ölçümler yapılacaksa o ayak çıplak olarak KUS cihazında ayak platosuna yerleştirilmiştir (Şekil 3) ve

ayağın hareket etmesini önlemek için şekil 3'te de görüldüğü gibi ayak bir kayış ile bağlanmıştır.



Şekil 3. KUS ile yapılan ölçümün önden görünüşü

Kantitatif Ultrason aletindeki kauçuk pedler (problar) topuğun iki yanından ayağa temas etmektedir. Pedlere yağ bazlı ultrason temas jeli sürülür. Her bireyde boy uzunluğunun farklılık göstermesinden dolayı oturma pozisyonları KUS cihazına bağlı olarak ayarlanmıştır. Daha sonra ölçümler iki ayrı gözlemci tarafından, belirlenen tek bir topukta toplam 10 kez tekrarlanmış olup, ilk 5 ölçüm 1. gözlemci, ikinci 5 ölçüm ise 2. gözlemci tarafından yapılmıştır. Her bir birey için ölçüm işlemi ortalama 20 dakika sürmüştür ve çıkan sonuçlar kaydedilmiştir.

Bulgular

Ölçüm yapılan 40 deneğin 19'u (% 47,5) 18 – 20 yaş grubunda; 21'i (% 52,5) 21 – 27 yaş grubunda idi.

Tablo 1. Deneklerin yaş, boy, kilo ve beden kitle indeksi (BKİ) değerleri

Cinsiyet	n	Yaş (yıl) $\bar{x} \pm SD$	Boy (cm) $\bar{x} \pm SD$	Kilo (kg) $\bar{x} \pm SD$	BKİ (kg/m ²) $\bar{x} \pm SD$
Kadın	20	21,05 ± 2,46	164,35 ± 6,32	54,55 ± 4,54	20,20 ± 1,32
Erkek	20	20,95 ± 1,76	177,00 ± 6,21	68,90 ± 6,75	22,03 ± 2,33
Toplam	40	21,00 ± 2,11	170,68 ± 8,90	61,73 ± 9,22	21,11 ± 2,09

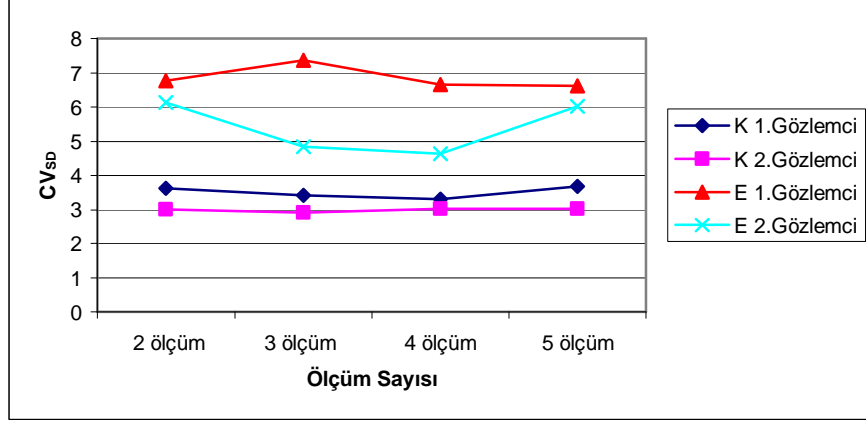
Tablo 2. BUA deęişkeninin cinsiyete, gözlemciye ve ölçüm sayılarına göre tanımlayıcı istatistikleri

	Ölçüm Sayısı	1. GÖZLEMÇİ			2.GÖZLEMÇİ		
		\bar{x}	\pm	SD	\bar{x}	\pm	SD
Kadın (n=20)	1	78,74	\pm	12,98	78,16	\pm	13,39
	2	78,75	\pm	12,86	77,91	\pm	13,20
	3	78,81	\pm	12,93	78,16	\pm	13,30
	4	78,82	\pm	12,53	78,17	\pm	13,16
	5	78,77	\pm	12,60	78,15	\pm	13,27
Erkek (n=20)	1	81,89	\pm	17,85	80,63	\pm	14,08
	2	81,71	\pm	15,22	81,55	\pm	15,61
	3	82,10	\pm	15,43	81,13	\pm	15,31
	4	82,23	\pm	15,80	80,90	\pm	15,01
	5	82,48	\pm	15,67	81,13	\pm	15,50

Tablo 3: Kadınlarda ve erkeklerde BUA deęişkeni için iki gözlemci tarafından gerçekleştirilen ölçümlerde bireyin kısa dönem kesinlięi ($SD_{kesinlik}$) ve bu kesinlięin yüzde olarak ifadesi (CV_{SD})

Cinsiyet	Ölçüm sayısı	1.Gözlemci		2.Gözlemci	
		$SD_{kesinlik}$	CV_{SD}	$SD_{kesinlik}$	CV_{SD}
Kadın (n=20)	2	2,848	3,617	2,345	3,010
	3	2,689	3,412	2,279	2,916
	4	2,608	3,309	2,363	3,023
	5	2,904	3,687	2,370	3,032
Erkek (n=20)	2	5,531	6,769	4,995	6,125
	3	6,047	7,365	3,932	4,847
	4	5,475	6,658	3,749	4,634
	5	5,457	6,616	4,878	6,013

Tablo 3 ve Grafik 1 de görüldüğü gibi 2. gözlemciye göre hem kadınlarda hem de erkeklerde daha düşük kal-maktadır. Ölçüm sayısının artması deęişkenlięi etkilememektedir. Ölçüm sayısının 2 olması yeterli olacaktır.



Grafik 1. BUA değişkenine ait CV_{SD} değerlerinin cinsiyete, gözlemciye ve ölçüm sayısına göre aldığı değerler

Tartışma ve Sonuç

KUS değişkenleri için yapılan ölçümlerin kesinliğini belirlemede değişim katsayısı sıklıkla kullanılmaktadır (5, 6, 7, 8, 9). Araştırmacılar, değişkenliği azaltabilmek yani kesinliği yüksek bir sonuç elde edebilmek için ölçüm sayılarını yüksek tutmaktadırlar. Gereksiz yere tekrar sayısının yüksek olması araştırmayı zaman yönünden olumsuz olarak etkiler. Bu nedenle çalışmamızda KUS değişkenleri için tekrar sayısının kaç olması gerektiğinin belirlenmesi amaçlandı. Çalışmada, iki gözlemci tarafından gerçekleştirilen 5'er tekrarlı ölçümün sonuçları incelendiğinde; gözlemcilerin elde ettiği sonuçların farklı olduğu ve ölçüm sayısını arttırmakla daha iyi sonuçların alınmadığı görüldü. Kantitatif Ultrason (KUS) değişkenlerinden BUA için 2. gözlemcinin elde ettiği değerlerin 1. gözlemciye göre daha iyi olduğu gözlemlendi. Söz konusu farkın kullanılan jel miktarından ve ayakların sabitlenme durumundan kaynaklandığı düşünüldü. Ölçümün fazla sayıda tekrarlanması gereksiz olduğu ve tekrar sayısının 2 olmasının yeterli olacağına karar verildi.

Kaynaklar

1. Zar JH. Biostatistical Analysis. 3rd edition., Prentice Hall, New Jersey; 1996.
2. Yurtsever N. Deneysel İstatistik Metotlar, 1. Baskı, Toprak ve Gübre Araştırma Enstitüsü Müdürlüğü Yayınları, Ankara; 1984: 40-41.

3. Özdamar K. SPSS ile BİYOİSTATİSTİK, 4. baskı, Kaan Kitabevi. Eskişehir; 2001.
4. Glüer CC, Blake G, Lu Y, Blunt BA, Jergas M, Genant HK. Accurate Assessment of Precision Errors: How to Measure the Reproducibility of Bone Densitometry Techniques. *Osteoporosis Int.* 1995; 5:262 – 270.
5. Drozdowska B. Quantitative ultrasound at the calcaneus in corticosteroid – treated male patients with bronchial asthma. *Ultrasound in Med. and Biol.* 2004; 30 (8): 1057 – 1061.
6. Sosa M, Saavedra P, Torres M, Alegne J, Gomez C, Macias J, Guanabens N, Hawkins F, Lozano C, Martinez M, Mosquera J, Cano R, Quesada M, Salas E and the GIUMO Study Group. Quantitative Ultrasound Calcaneus Measurements: Normative Data and Precision in the Spanish Population *Osteoporos Int* 2002; 13: 487 – 492.
7. Lee M, Czerwinski SA, Choh AC, Towne B, Demerath EW, Chumlea WC, Sun SS, Siervogel RM. Heritability of calcaneal quantitative ultrasound measures in healthy adults from the Fels Longitudinal Study. *Bone* 2004; 35: 1157 – 1163.
8. Frost ML, Blake GM, Fogelman I. Contact Quantitative Ultrasound: An Evaluation of Precision, Fracture Discrimination, Age – Related Bone Loss and Applicability of the WHO Criteria. *Osteoporos Int* 1999; 10: 441 – 449.
9. Hadji P, Hars O, Bock K, Albert US, Beckmann MW, Emons G, Schulz KD. Age Changes of Calcaneal Ultrasonometry in Healthy German Women. *Calcif Tissue Int* 1999; 65:117 – 120.

MEVSİMSEL İNSİDANS HIZI FARKLARININ ANLAMLILIK TESTİ VE BİR UYGULAMA

Ahmet Dirican, Ömer Uysal, Sevim Purisa, Ayşegül Bostan,
Mustafa Şenocak

İstanbul Üniversitesi, Cerrahpaşa Tıp Fakültesi, Biyoistatistik BD.

e-mail: adirican@istanbul.edu.tr

Özet

Epidemiyolojik çalışmalarda söz konusu hastalığın insidans hızının mevsimlik veya daha farklı periyodik dönemlere ait eğilimlerini anlamaya gerek duyulur. Çalışmanın verileri uygulama bireyleri üzerinden toplanırken, belli periyotlara ait özel kategoriler olarak bireyin ölüm, doğum, nüks veya hastaneye yatışları kaydedilir.

Çalışmanın başlıca amacı bir toplumda ele alınan hastalığın potansiyel eğilimlerini tanımlamaktır. Böylece hastalığın patofizyolojik mekanizmalarının altında yatan temel nedenler ortaya çıkarılır ve diğer önemli çevresel faktörler tanımlanarak, özellikle kronik hastalıkların nüks veya mortalite risklerinin hangi zamansal kesitte yükseleceği tahmin edilir.

Sonuçta hastalıkların mevsimler veya daha küçük periyotlardaki epidemiyolojik verilerin biyoistatistiksel anlamlılığı Roger's testi ile denetlenerek uygulama alanında hastalığın risk dönemlerine uygun ve yeterli önlemleri almamıza olanak sağlar.

Tıbbi önlemlerin geliştirilmesine ve bilimsel uygulayıcıların eğitim stratejilerinin belirlenmesine de temel oluşturan bu çalışma bir klinik uygulama örneği üzerinde ayrıntılı olarak irdelenmiştir.

Anahtar kelimeler: Mortalite; İnsidans hızı; Epidemiyolojik veriler; Mevsimler.

SAĞLIK PERSONELİNİN EMPATİK EĞİLİMLERİNİN VE MESLEKİ DOYUMLARININ ÇOK DEĞİŞKENLİ KOVARYANS ANALİZİ (MANCOVA) İLE DEĞERLENDİRİLMESİ

Semra Akgöz¹, İsmet Kan¹, Alis Özçakır², Yusuf Sivrioğlu³,
Kamuran Tombul⁴, Yıldız Altınsoy⁴

Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı¹

Uludağ Üniversitesi Tıp Fakültesi Aile Hekimliği Anabilim Dalı²

Uludağ Üniversitesi Tıp Fakültesi Psikiyatri Anabilim Dalı³

Uludağ Üniversitesi Tıp Fakültesi Sağlık Uygulama ve Araştırma Merkezi⁴

Özet

Kaliteli sağlık hizmetlerinin yürütülmesinde empatik iletişim çok önemlidir. Sağlık personelinin, hizmetlerin sunumunda daha etkin ve verimli çalışabilmeleri için ise, mesleki doyum önem taşımaktadır. Bu çalışma, Uludağ Üniversitesi Sağlık Uygulama ve Araştırma Merkezi'nde çalışan sağlık personelinin mesleki doyum ve empatik eğilim düzeylerinin demografik özellikleri, meslekleri ve kendileri ile ilgili görüşleri açısından farklılık gösterip göstermediğini saptamak amacıyla yapılmıştır.

Tanımlayıcı nitelikteki bu çalışmanın örneklemini, Uludağ Üniversitesi Sağlık, Uygulama ve Araştırma Merkezi'nde (SUAM) en az 6 aydır uzman veya araştırma görevlisi olarak çalışan 304 hekim ve 479 hemşire oluşturmuştur. Araştırmaya katılma oranı % 82,4'dir (N=645).

Araştırmada, Musal ve arkadaşlarının benzer konularda yapılan araştırmalardan ve Minnesota İş Doyum Ölçeğinden yararlanarak oluşturdukları 22 sorudan oluşan Mesleki Doyum Ölçeği ile Dökmen tarafından geliştirilmiş, kişilerin günlük yaşamlarında empati kurma potansiyellerini ölçmeyi amaçlayan 20 sorudan oluşan Empatik Eğilim Ölçeği kullanılmıştır. Beşli Likert tipindeki ölçeklerin güvenilirlik analizi sonucunda Cronbach'ın Alfa Katsayıları sırasıyla 0.9199 ve 0.8920 ($p < 0.0001$) olarak bulunmuştur.

Empatik eğilim ölçeği için ortalama skor 71.02 ± 0.36 , mesleki doyum için 4.22 ± 0.06 'dır. Empatik eğilim ve mesleki doyum arasında pozitif, anlamlı korelasyon ($r = 0.298$), yaş ile empatik eğilim ($r = 0.163$) ve mesleki doyum ($r = 0.283$) arasında da mevcuttur ($p < 0.001$).

Çok değişkenli kovaryans analizine dahil edilen bağımsız değişkenler temelinde, (mesleki bilgi ve becerisini kullandığını düşünme, geçmişe dönebilse mesleğini yeniden tercih edecek olma ve çevresindekilere duygularını rahatlıkla ifade edebilme) 2x2x2 faktöriyel dizaynda 8 grup oluşmuştur ve grupların her birisi önerilen minimum örneklem büyüklüğüne sahiptir ($n \geq 20$). Normallik analizinden sonra her bir bağımlı değişkenin hata varyansının gruplarda eşit olduğu görülmüştür (Levene test $p > 0.05$). Box's M test p-değeri 0.554'dür, böylece varyans-kovaryans matrislerinin homojenliği varsayımı karşılanmıştır. Bağımlı değişkenlerin korelasyonu için Bartlett küresellik testi anlamlı bir p-değerine (< 0.001) sahiptir. Pillai's Trace, Wilk's lambda ve Roy's largest Root istatistikleri etkileşimlerin etkisinin anlamlı olmadığını ($p > 0.05$) gösterirken, tüm bağımsız değişkenler ve kovaryet değişken "yaş" için anlamlı sonuçlar vermiştir ($p < 0.001$). "Mesleki bilgi ve becerisini kullandığını düşünme" empatik eğilim ve mesleki doyum için ($p < 0.01$), "çevresindekilere duygularını rahatlıkla ifade edebilme" empatik eğilim için ($p < 0.001$), "geçmişe dönebilme imkanı olsa yeniden mesleğini tercih edecek olma" mesleki doyum için ($p < 0.001$) gruplar arasında anlamlı farklılık göstermiştir. Kovaryet değişken "yaş" ile bağımlı değişkenler arasında anlamlı ilişki saptanmıştır ($p < 0.001$).

Çalışmamızda sağlık personelinin mesleki doyumunun ve empatik eğiliminin yaş ile birlikte arttığı görülmüştür. Geçmişe dönebilse mesleğini yeniden tercih edecek olanların empatik eğilimleri, mesleki bilgi ve becerilerini kullanabildiğini düşünenlerin ise hem mesleki doyumları hem de empatik eğilimleri anlamlı olarak daha yüksektir. Çevresindekilere duygularını güçlkle ifade edenler ise anlamlı olarak daha düşük empatik eğilim düzeylerine sahiptir.

DÖRT FARKLI NORMALLİK TESTİNİN TIP I HATA VE GÜÇ BAKIMINDAN DEĞİŞİK DAĞILIMLARDA İNCELENMESİ

Derya Öztuna, Atilla Halil Elhan, Ersöz Tüccar

Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı
[e-mail:dgokmen2001@yahoo.com](mailto:dgokmen2001@yahoo.com)

Özet

Sürekli bir verinin normal dağılıp dağılmadığını araştırmak amacıyla geliştirilmiş birçok yöntem vardır. Bunlardan bir bölümü **grafik**, bir bölümü de **test** yöntemleridir. Ancak, grafik yöntemleri dağılım yapısı hakkında bilgi sahibi olmamıza olanak sağlarken; dağılımın normal olduğunu garanti etmemektedir. Bu nedenle verinin dağılımının normal dağılıma (çan eğrisi) uygun olup olmadığına karar vermek için en iyi yol, grafikleri uygun test yöntemleri ile birlikte incelemektir.

Dağılımların normal dağılıp dağılmadığını test etmek amacıyla kullanılan en yaygın test yöntemleri; Ki-kare uyum iyiliği testi, Kolmogorov Smirnov (K-S) D testi, Lilliefors düzeltmeli Kolmogorov Smirnov (K-S) testi, Anderson-Darling testi, Cramer-von Mises testi, Shapiro-Wilk W testi, D'Agostino çarpıklık testi, Anscombe-Glynn basıklık testi, D'Agostino-Pearson K^2 testi ve Jarqua-Bera LM testidir. Çalışmamızda, Tip I hata ve güç bakımından Lilliefors düzeltmeli Kolmogorov Smirnov (K-S) testi, Shapiro-Wilk W testi, D'Agostino-Pearson K^2 testi ve Jarqua-Bera LM testinin performansları karşılaştırılmıştır. Sonuçlar genel olarak incelendiğinde, normal dağılıma sahip örneklerde Jarqua-Bera LM testinin, normal dağılıma sahip olmayan örneklerde ise Shapiro-Wilk W testinin diğer testlere göre daha iyi performansa sahip oldukları belirlenmiştir.

Anahtar Kelimeler: Lilliefors ve Kolmogorov-Smirnov testi, Shapiro-Wilk testi, D'Agostino-Pearson testi, Jarqua-Bera testi, Tip I hata, Güç.

Kaynaklar

1. D'Agostino R.B., Belanger A., D'Agostino Jr. (1990). *A Suggestion for Using Powerful and Informative Tests of Normality*. The American Statistician, 44 (4), 316-321.

2. Mendes M. ve Pala A. (2003). *Type I Error Rate and Power of Three Normality Tests*. Pakistan Journal of Information and Technology, 2 (2), 135-139.
3. Jarqua C.M. ve Bera A.K. (1987). *A test for normality of observations and regression residuals*. International Statistical Review, 55(5):163-172.

ÜLKEMİZDE AĞRI ve KAPI İĞNE ZAMANINI, ULUSLARARASI KLAVUZLARIN ÖNERDİĞİ SÜREYE UYGUN MUDUR?

Hatice Bulut¹, Gülşah Seydaoğlu², Fevziye Türemez²

¹ Çukurova Üniversitesi Tıp Fakültesi Kardiyoloji AD

² Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik AD

e-mail : turemezf@yahoo.com

Özet

Amaç: Uluslararası kılavuzlar akut miyokard infarktüsü (AMI) geçiren hastalarda, hastanın acile girişinden iğnenin yapıldığı ana kadar geçirilen süreyi 30 dakika, ağrının başladığı andan iğnenin yapıldığı ana kadar geçen süreyi de 60 dakika olarak belirlemiştir. Bu çalışmada bölgemizde, AMİ geçiren hastalarda trombolitik tedavi uygulamasına kadar geçen sürenin belirlenmesi, bu sürenin mortalite ile ilişkisi ve tedavi gecikmesini etkileyen faktörler araştırıldı.

Yöntem: Bu çalışmada Ocak 2003 ile Aralık 2003 arasında Çukurova Üniversitesi acil ünitesine başvuran ve AMİ olduğu saptanan 107 hasta bir ay süresince izlenmiştir. Hastanın acile girişinden iğnenin yapıldığı ana kadar geçirilen süre; kapı iğne zamanı-KİZ, ağrının başladığı andan iğnenin yapıldığı ana kadar geçen süre; ağrı iğne zamanı-AİZ olarak tanımlanmıştır.

Bulgular: Çalışmaya 83 erkek 24 kadın olmak üzere 107 hasta (yaş ortalaması 55.3±10.1) alınmıştır. Tüm hastalarda ağrının başlaması ile acil servisine başvuru arasında geçen sürenin ortalaması (başvuru gecikmesi) 223.3±175.1 dk (medyan 180 dk), acilde gecikme süresi ortalama 26.2±13.9 dk (medyan 25 dk), koroner yoğun bakım ünitesinde (KYBÜ) gecikme süresinin ortalaması ise 16.4±11.7 dk (medyan 10 dk) olarak belirlenmiştir. Kapı iğne zamanı 42.7±17.2 dk (medyan 42 dk), ağrı iğne zamanı ortalaması 266.1±177.6 dk (medyan 235 dk) olarak saptanmıştır. KİZ, 36 hastada (% 33.6) 30 dakikanın altında iken AİZ sadece 4 hastada (% 3.7) 60 dakikanın altındadır. Cinsiyet, yaş grubu, yaşadığı yer, ağrının başlama saati, daha önce bir sağlık kurumuna başvurma faktörlerinin AİZ için risk faktörü olduğu saptanmıştır. Tedavi sonrasında % 86.9'unda patolojik Q, % 57.0'ında % 50 ST çökmesi saptanan hastaların 3'ü 28 günden önce, 4'ü ise 28 gün

sonrasında kaybedilmiştir. Gecikme sürelerinin, ölen hastalarda sağ kalanlara oranla anlamlı olarak daha uzun olduğu belirlenmiştir.

Sonuç: Kapı-ağrı iğne zamanlarındaki gecikmenin AMİ hastalarının mortalitesini etkilediği gözlenmiştir. Bölgemizde elde edilen kapı-ağrı iğne zamanların uluslararası kılavuzların önerdiği (KİZ; 30 dk ve AİZ; 60 dk) sürelerinden uzun olduğu belirlenmiştir. Bunu etkileyen önemli faktörlerin hastaların geç başvurusu ve acil ünitesindeki gecikme olduğu düşünülmektedir.

GENELLEŞTİRİLMİŞ TAHMİN DENKLEMLERİ: GTD

Seval Kul, E. Arzu Kanık

Mersin Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı

e-mail: sevalkul@mersin.edu.tr

Özet

Bu çalışmada genelleştirilmiş tahmin denklemlerinin kullanım alanları, quasi-likelihood fonksiyonunun elde edilişi, bağlantı fonksiyonlarının ve üzerinde durulan ilişki matrislerinden en uygun olanın tercih edilmesi ve genelleştirilmiş tahmin denklemlerinde karşılaşılan over dispersion gibi bazı sorunlar tartışılmıştır. Örnek bir veri seti üzerinde GTD yönteminin uygulanması R paket programı üzerinde gösterilmiştir.

Genelleştirilmiş tahmin denklemlerinde anahtar karakteristik aynı deneklerden elde edilen ölçümler ilişkilidir ve dolayısıyla bağımsızlık varsayımı altında çalışılan birçok istatistiksel yöntemin kullanılması mümkün olmayacaktır. Bu ilişkinin görmemezlikten gelinmesi ve standart istatistik yöntemlerinin kullanılması yanlışlığı ortaya çıkararak yetersiz tahminlere neden olabilecektir. Standart tekrarlanan ölçümlü çalışmalarda ise bireylerin birbirinden bağımsız olduğu varsayılır. Fakat bu varsayım ölçümlerin aynı kümedeki deneklerden elde edildiği birçok klinik ve epidemiyolojik çalışmada bozulur. Kümeler; aileler, kardeşler, ikizler, eşler, sınıflar ve aynı hasta odasında kalan hastalar gibi çok çeşitli olabilir. Bu tür özelliklere sahip çalışmalarda GTD kullanılabilir alternatif bir yöntemidir.

Genelleştirilmiş tahmin denklemlerinde katsayıların tahminleri quasi-likelihood fonksiyonundan yararlanılarak elde edilir.

$$\sum_{i=1}^K \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{S}_i = \mathbf{0}$$

Quasi-likelihood fonksiyonunun bileşenlerine göz atacak olursak;

$\mathbf{S}_i = y_i - \mu_i$; Her bireyin tekrarlanan ölçümlerinin kendi ortalamalarından farkıdır.

\mathbf{D}_i : Bağlantı fonksiyonunun modelde tahmin edilecek her bir parametre için kısmi türevinin alınması ile elde edilir.

Varyans sandviç tahmini (sandwich estimation) adı verilen V_i ile tahmin edilir.

$$V_i = A_i^{1/2} R_i(\alpha) A_i^{1/2}$$

A_i , $n_i \times n_i$ boyutlu tekrarlanan ölçümlere ait Varyans kovaryans matrisinin köşegenlerinden elde edilen ve varyanslardan oluşan bir köşegen matrisi, $R_i(\alpha)$ üzerinde durulan ilişki matrisidir. Aynı küme içindeki gözlemler arasındaki ilişki $R_i(\alpha)$ $n_i \times n_i$ boyutlu simetrik bir matris ile ifade edilir. Ve bu matrise “Üzerinde Çalışılan İlişki Matrisi” adı verilir. Genelleştirilmiş tahmin denklemlerinde birçok ilişki matrisinden söz edilebilir. Bunlardan bazıları; Yapılandırılmamış (Unstructured) korelasyon, Bağımsızlık (Independence) korelasyon yapısı, Değişebilir (Exchangeable) korelasyon yapısı, Otoregresif (Auto-regressive) korelasyon yapısı, Sabit (Fixed) korelasyon yapısı, M-bağımlı (M-dependent) korelasyon yapısıdır.

BİR KENTTE 1995-2005 YILLARI ARASINDA TUTULAN 63.831 ÖLÜM SERTİFİKASI BİZE NE ANLATIYOR?

Gülşah Seydaoğlu, Fevziye Türemez, Nilgün Arpacı, Nazan Alparslan

Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik AD

ÖZET

Bu çalışmanın amacı Adana ilinde toplanan ölüm sertifikalarının değerlendirilmesi ve kalite kontrolünün yapılmasıdır.

Yöntem

Çalışmada 1995-2002 yılları arasında toplanan ve bilgisayar ortamına ölüm sebebi kodlanmamış olarak girilmiş olan 47.665 ölüm sertifikası 150 başlıklı hastalık kodlamasına göre kodlandı. 2003-2005 (ilk 6 ay) yıllarında kodlanarak girilmiş 16.241 ölüm sertifikası ile birleştirilerek değerlendirildi.

Bulgular

Tüm ölümlerin % 58.1'i erkek, % 41.9'u kadındır, cinsiyet bilgisi olmayan form sayısı sadece 48'dir. Yaş bilgisi olmayan form sayısı 151 (% 0.2), medeni durum bilgisi olmayan 3733 (% 5.8), öğrenim durumu bilgisi olmayan 3919 (% 6.1) meslek bilgisi olmayan 4192'dir (% 6.6). Yerleşim yeri % 68,8'i il merkezi, % 19.8'i ilçe merkezi, % 10.8'i köy olan ölümlerin 407'sinin (% 0.6) ikametgah bilgisi eksiktir. Vakaların sadece 3853'üne (% 6.0) otopsi yapılmıştır. Ölüm sebeplerine göre yapılan sınıflamada 13483 (% 21.1) formda uygun olmayan ölüm sebebi (semptom, son veya ara neden gibi) veya okunaksız açıklamalar yazıldığı saptanmıştır. Bu uygunsuz formların % 24.4'ünde solunum yetmezliği, % 16.2'sinde kardiyak arrest, % 13.9'unda sepsis, % 11.1'inde beyin kanaması, % 7.9'unda böbrek yetmezliği gibi temel neden olamayacak açıklamalar yazmaktadır. Bunların dışında kalan nedenler içinde en sık ölüm nedeni kalp hastalıklarına bağlı ölüm (% 30.6) olarak görülmektedir. Fakat bu sınıflamada da şüpheler vardır, ölüm nedeni olarak yalnızca "kalp yetmezliği" yazılmış % 11.4 formda temel neden kalp hastalığı mıdır yoksa son neden midir belli değildir. Buna rağmen ölüm verileri yıllara göre nüfusa oranlanarak çeşitli ölüm hızları

saptanmaya çalışılmış (yaşa, cinsiyete, hastalıklara göre) ve DİE ve sağlık bakanlığı verileri ile karşılaştırılmıştır.

Sonuç

Bilgisayar teknolojisinin yaygınlaştığı, hekim sayısının arttığı, öğretim elemanları sayısının yeterli olduğu ülkemizde hala ölümler ve hastalıklar ile ilgili tutulan kayıtlarda önemli eksiklik ve yanlışlıklar olduğu görülmektedir. Bu konuda ülke çapında hekim eğitimi seferberliği oluşturulmasında biyoistatistik uzmanlarının görev ve sorumluluğunun fazla olduğu inancındayız.

FAKTÖR ANALİZİ MODELİNDE ORTAK VARYANSLARI BELİRLEMEDE GENELLEŞTİRİLMİŞ TERSLERİN KULLANIMI

Seval Süzölmüş¹, Sadullah Sakallıođlu²

*Çukurova Üniversitesi, Osmaniye Meslek Yüksekokulu OSMANİYE¹
Çukurova Üniversitesi Fen-Ed. Fak. İstatistik Bölümü Balcalı-ADANA²
e-mail:suzulmus@mail.cu.edu.tr*

Özet

Bu çalışmada, faktör analizi modelinde ortak varyansların tahmini üzerinde çalışıldı. Daha önceki çalışmalarda varyans-kovaryans matrisi Σ 'nın pozitif tanımlı olduđu kabul edilmiştir. Σ 'nın non-negatif tanımlı olduđu kabul edilerek genelleştirilmiş ters kullanımı içeren az sayıda çalışma bulunmaktadır. Bu çalışmada Σ 'nın non-negatif tanımlı olduđu kabul edilerek genelleştirilmiş ters kullanımı ele alınıp faktör analizi modelindeki ortak varyansların tahmini incelenmiştir.

Anahtar Kelimeler: Faktör Analizi, Ortak Varyans, Genelleştirilmiş Ters.

FUZEL YAĞINDAN İZOAMİL ASETATIN SÜREKLİ ÜRETİMİNDE OPTİMUM KOŞULLARIN BELİRLENMESİ İÇİN İKİ YÖNTEMİN BİR ARADA KULLANILMASI: YANIT YÜZEYLERİ VE YAPAY SINIR AĞLARI

Ömer Utku Erzengin¹, Havva Elemzade², Gamze (Yalçın) Erzengin², Afife Güvenç²

*Marmara Üniversitesi, Tıp Fakültesi, Biyoistatistik ABD¹
Ankara Üniversitesi, Mühendislik Fakültesi, Kimya Mühendisliği Bölümü²*

Biyoteknoloji ve eczacılıkta kimyasal üretim aşamalarında kullanılan optimizasyon yöntemlerinden biri yanıt yüzeyleridir. Kimyasal deneyler pahalı ve zaman alıcı olduğundan, önceden belirlenmiş değişkenler ve noktalarda veri elde edip modellemelerin yapılmasıyla en uygun koşulların belirlenmesi amaçlanmaktadır. Elde edilen modeller ışığında, deneyleri yapılmayan nokta-lar ve aralıklar için üretim aşamasında tahminlerde bulunmaktadır.

Fuzel yağı etil alkol üretiminde elde edilen bir yan üründür. Dünyada özellikle etil alkolün büyük ölçeklerde üretildiği ülkelerde yan ürün probleminin önemli bir kısmını oluşturmaktadır. Bazı ülkelerde endüstrinin enerji ihtiyacının bir kısmını karşılamak için yakılmaktadır. Türkiye’de ise, etkin bir şekilde değerlendirilememektedir.

Fuzel yağının değerlendirilmesi kapsamında yapılan kaynak araştırması sonucu, bu maddeden gıda, farmasötik ve kozmetik endüstrilerinde önemli tat ve koku bileşenlerinin üretildiği görülmüştür. İlaçlarda (şurup, tatlandırıcı vb) kullanılan çilek ve muz aromaları olan etil bütirat ve izoamil asetat sözü edilen tat ve koku bileşenlerine örnek olarak verilebilir.

Fuzel yağı başlıca izoamil alkol, propil alkol, izobütil alkol, etil alkol ve n-bütil alkol gibi düşük molekül ağırlıklı alkoller ve az miktarda sudan oluşmaktadır. Fuzel yağından izoamil asetatın (y) üretimi için sıcaklık (x_1), basınç (x_2), substrat / CO₂ miktarı (x_3), besleme akış hızı (x_4) değişkenleri altında denemeler yapılmıştır. Elde edilen veri yanıt yüzeyi (response surface methodology) yöntemiyle modellenmektedir.

Yanıt yüzeyleri metodolojisi (Response Surface Methodology, RSM) iki ve/veya daha fazla değişken yardımıyla tanımlanan yüzeylerin matematiksel ve istatistiksel tekniklerin yardımı sayesinde modellenmesidir.

Yapay sinir ağırları birçok modellemeye alternatif olarak ortaya atılmış bir modelleme tekniğidir. Yapay sinir ağırları (Artificial Neural Network, ANN) model olarak beyindeki sinir topluğunun bilgiyi iletme prensibini esas alır.

Yanıt yüzeylerine karşılık gelen yapay sinir ağırlarının kullanımı literatürde de yenidir. Vaithiyalingam ve arkadaşları 2002 (kasım ayı) yılında selüloz asetat bütirat yayılımı üzerine, Dutta ve arkadaşları 2004 yılında Pseudomonas sp izolasyonu üzerine, Sun ve arkadaşları 2003 yılında kontrollü ilaç salınım sistemleri üzerine, Takayama ve arkadaşları 2004 yılında ilaç formüllerinin optimizasyonu üzerine yanıt yüzeyleri ve yapay sinir ağırları ile karşılaştırmalı makaleler yayınlamışlardır.

Yapılacak sunumda fuzel yağından izoamil asetatın sürekli üretiminde optimum koşulların belirlenmesi için yanıt yüzeyleri ve yapay sinir ağırları bir arada kullanılacak ve iki yöntem birbiri ile karşılaştırılacaktır.

Kaynaklar

1. Anderson M ve arkadaşları. Design-Expert. Software Version 6 User's Guide. Minneapolis: Stat-Ease, Inc; 2003.
2. Christopher D, Himmel and Gary S. Advantages of Plasma Etch Modeling Using Neural Networks Over Statistical Techniques. Ieee Transactions On Semiconductor Manufacturing 1993; 6: 2.
3. Cundy AL. Use of Response Surface Metamodels in Damage Identification of Dynamic Structures (Yüksek Lisans Tezi). Virginia: Virginia Polytechnic Institute and State University 2003.
4. Dutta JR, Dutta PK, Banerjee R. Optimization of culture parameters for extracellular protease production from a newly isolated Pseudomonas sp. using response surface and artificial neural network models. Process Biochemistry 2004; 39:2193-2198.
5. Demuth H, Beale M. Neural Network Toolbox For Use with MATLAB. Eighth printing. Apple Hill Drive: The MathWorks, Inc; 2004.
6. Papila NU, Neural Network And Polynomial-Based Response Surface Techniques For Supersonic Turbine Design Optimization (Doktora Tezi). Florida: University Of Florida 2001.
7. Sarle WS. Neural Networks and Statistical Models. Proceedings of the Nineteenth Annual SAS Users Group International Conference. 1994.
8. Seung S. Introduction to Neural Networks (Massachusetts Institute of Technology Elektronik ortamda ders notları) <http://hebb.mit.edu/>; 2002.
9. Subramanian N, Yajnik A, Murthy RSRamachandra. Artificial Neural Network as an Alternative to Multiple Regression Analysis in Optimizing

Formulation Parameters of Cytarabine Liposomes. AAPS Pharm Sci Tech 2004; 5.

10. Sun Y, Peng Y, Chen Y, Shukla AJ. Application of artificial neural networks in the design controlled release drug delivery systems. *Advanced Drug Delivery Reviews* 2003; 55: 1201–1215.
11. Takayama K, Fujikawa M, Obata Y, Morishita M. Neural network based optimization of drug formulations. *Advanced Drug Delivery Reviews* 2003; 55:1217-1231.
12. Vaidyanathan R, Papila Ni, Shyy W, Tucker PK, Griffin(2) LW. Neural Network And Response Surface Methodology For Rocket Engine Component Optimization. *AIAA* 2000; 4880.
13. Vaithiyalingam S, Khan MA. Optimization and characterization of controlled release multi-particulate beads formulated with a customized cellulose acetate butyrate dispersion. *International Journal of Pharmaceutics* 2002; 234:179-193.

TÜRKİYE'DE İL VE İLÇE MERKEZLERİ ÖLÜM İSTATİSTİKLERİYLE MORTALİTENİN YAŞA, CİNSİYETE VE ZAMANA GÖRE DEĞİŞİMİ (1990-2003)

R. Erol Sezer¹, Hafize Sezer²

Cumhuriyet Üniversitesi Tıp Fakültesi Aile Hekimliği Anabilim Dalı¹

Cumhuriyet Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı²

Giriş

1963'den beri düzenli aralıklarla yapılan nüfus ve sağlık araştırmaları, bebek, 1-4 yaş ve beş yaş altı ölüm hızları ile bu hızların değişim özelliklerinin bilinmesine olanak sağlamıştır (1, 2). Türkiye'nin 1963'de binde 208 olarak tahmin edilen bebek ölüm hızı, 1966-67 döneminde binde 170, 1973 de binde 100-150 olarak tahmin edilmiştir (1). Beş yaş altı ölüm hızı, 1973 araştırmasında(1) binde 218 olarak bildirilmiştir. Her iki hız da zaman içinde sürekli düşmüştür. 1988 araştırmasında(3) bebek ölüm hızı binde 82, beş yaş altı ölüm hızı binde 98 olarak bildirilirken, 2003 araştırmasında bu hızlar sırayla, binde 29 ve binde 37 olarak bulunmuştur (2). Bu araştırmalar, bebek, 1-4 yaş ve beş yaş altı ölüm hızlarının bölgelere, kent-kır olma durumuna ve bazı demografik değişkenlere göre değişimini inceleme olanağı vermektedir. Özetle Türkiye'de beş yaş altı dönemle ilgili bebek ve çocuk ölümlülüğü konusunda güvenilir bilgiler son kırk yıl için vardır. Fakat 5 yaş üstü nüfusun ölümlülüğü konusundaki bilgiler çok azdır. Nüfus Etütleri Enstitüsü tarafından beş yılda bir yapılan nüfus ve sağlık araştırmaları beş yaş üstü nüfusun ölümlülüğünü inceleme kapsamı dışında tutmaktadır. Devlet İstatistik Enstitüsü'nün (DİE) 1989 Türkiye Nüfus Araştırması (1989 TNA) kent-kır ayırımı olacak şekilde bölgelere göre kaba ölüm hızlarıyla son bir yıl için yaşa ve cinsiyete özgü ölüm hızlarını tahmin etmiştir (4). İzleyen dönemde Türkiye'nin beş yaş üstü nüfusunun ölümlülüğüyle ilgili yayınlanmış bir araştırmaya ulaşamamıştır. Türkiye'de ölüm istatistikleri, il ve ilçe merkezlerini kapsayacak şekilde Devlet İstatistik Enstitüsü tarafından yayınlanmaktadır. Bu istatistiklerin gerek nitelik gerekse nicelik yönünden sorunlu olduğu belirtilmiştir (5, 6).

Koroner kalp hastalığı, serebrovasküler hastalık ve kanser, dünyada ve Türkiye'de 35 yaş üstü nüfusun en önemli üç ölüm nedenidir. Türkiye'de 1990'lı yıllarda yapılan bir izleme çalışmasına göre 20 yaş üzeri ölümlerin, % 42.5'i koroner kalp hastalığı, % 25'i kanser, % 16.2'i serebrovasküler hastalık nedeniyledir (7). Yirminci yüzyılın ikinci yarısı, bu hastalıkların önlenmesi ve tedavisi konusunda önemli keşiflere sahne olmuştur. Gün-

müzde bu hastalıklar artık önlenabilir hastalıklar arasındadır. Kardiyovasküler hastalığın tedavisi konusunda yakın geçmişte ölüm oranlarını önemli ölçüde düşürücü gelişmeler yaşanmıştır. Koroner kalp hastalığı ve serebrovasküler hastalık mortalite hızları gelişmiş ülkelerde hızla düşmektedir (8). Sigarayla mücadele programları etkili olan ülkelerde akciğer kanseri ölüm hızları da azalmaya başlamıştır (9).

Bu çalışma şu sorulara yanıt aramak amacıyla yapılmıştır: Türkiye’de mortalitenin durumu, değişim yönü ve hızı nedir? Türkiye’nin il ve ilçe merkezleri ölüm istatistikleri bu sorunun yanıtına katkı sağlayabilir mi?

Yöntem

İnceleme dönemi olarak 1990-2003 yılları seçilmiştir. Bunun nedenlerinden birisi ölümlerin yaşa ve cinsiyete göre dağılımını içeren en son yapılmış ulusal araştırmanın 1989 TNA olması, diğeri ise DİE’nin en yeni ölüm verilerinin 2003 yılına ait olmasıdır.

İnceleme döneminin ölümle ilgili verileri, DİE tarafından her yıl yayınlanmakta olan ve il ve ilçe merkezlerinde olan ölümleri kapsayan ölüm istatistiklerinden sağlanmıştır (10). DİE bu bilgileri “Ölü gömme izin kağıdının” bir parçası olan ve sağlık müdürlükleri tarafından DİE’ye gönderilen “Ölüm istatistik formları’ndan” sağlamaktadır.

İnceleme dönemi ölümlüğünün incelenmesinde yaşa ve cinsiyete özgü hızlar ile yaşa ve cinsiyete göre direkt standardize ölüm hızları kullanılmıştır. Direkt standardizasyonda standart nüfus olarak, kurgu bir nüfus olan Avrupa standart nüfusu alınmıştır. Yaşa ve cinsiyete özgü ölüm hızlarının hesabında yıl ortası nüfus olarak ilgili yılların tahmini yıl ortası nüfusları kullanılmıştır. 1990 nüfus sayımı sonuçlarına göre nüfusun yaş ve cinsiyet dağılımı (11), 1985-90 dönemi için il ve ilçe merkezleri nüfus artış oranı, 2000 yılı nüfus sayımına göre il ve ilçe merkezleri nüfusu, 1990-2000 nüfus sayımları arasında il ve ilçe merkezleri nüfusunun artış oranı, 2003 Türkiye Nüfus ve Sağlık Araştırması’nın nüfusun cinsiyete ve yaş gruplarına göre yüzde dağılımı, yıl ortası nüfusların hesaplanmasında kullanılmıştır. Yıl ortası nüfus hesabı, $P = P_0 * e^{hz}$ formülü ile yapılmıştır. İl ve ilçe merkezleri nüfus artış oranları 1985- 1990 dönemi için binde 33.83, 1990-2000 dönemi için binde 26,81 olarak DİE tarafından hesaplanmıştır. Türkiye’nin 2003 yılı hızları, gelişmiş ülke örneği olarak İngiltere’nin hızlarıyla karşılaştırılmıştır. Çünkü bu ülkenin ölüm istatistikleri güvenilir ve internet ile ulaşılabilir özelliktedir.

Bulgular

1-4 no’lu tablolar, il ve ilçe merkezlerinin 1990 ve 2003 yıllarına ait, yaşa ve cinsiyete özgü ölüm hızları ile, direkt standardize ölüm hızlarını göstermektedir. Bu tablolar, yıl ortası nüfusun yaş ve cinsiyete göre dağılımıyla, ölümlerin yaş ve cinsiyete göre dağılımını da içermektedir.

Bildirimlerin yeterliliği: Tablo 1 ve Tablo 2'deki veriler birlikte kullanıldığında, 1990 yılı için il ve ilçe merkezlerinin kaba ölüm hızı binde 4.56 olarak bulunur. 1989 Türkiye Nüfus Araştırması'na (1989 TNA) göre Türkiye'nin Ekim 1989'dan önceki bir yıla ait kaba ölüm hızı binde 7.79, kentlerde (Nüfusu 20 bin ve üzerinde olan yerleşim birimlerinde) binde 7.07, nüfusu daha az olan yerleşim birimlerinde (kır) binde 8.48 olarak bulunmuştur. Bu çalışmada bulunan binde 4.56'lık kaba ölüm hızı, 1989 TNA kaba ölüm hızı olan binde 7.07'den % 35 daha düşüktür.

Yaşa, cinsiyete ve zamana göre ölüm hızlarının değişimi: Ölüm hızları her iki cinsiyette de yaşla önemli düzeyde değişmektedir. Ölümlülüğün en yüksek olduğu dönemler, sırayla; 75 yaş ve üstü dönem, 65-74 yaş dönemi, bebeklik dönemi ve 45-64 yaş dönemidir. Ölümlülüğün en düşük olduğu yaş grupları da sırasıyla 5-14 ve 15-24 yaş gruplarıdır. Gerek 1990, gerekse 2003 yılı yaşa özgü ölüm hızları, her yaş grubu için kadınların ölüm riskinin erkeklerden düşük olduğunu göstermektedir. Tüm yaş gruplarını kapsayacak şekilde yaşa göre direkt standardize ölüm hızı 1990'da erkekler için binde 11.62, kadınlar için binde 8.70'dir. Aynı hızlar 2003'de erkekler için binde 7.97, kadınlar için binde 6.07'dir. Tüm yaş gruplarını kapsayan yaşa göre standardize hızlar için 1990 ile 2003 arasında gerçekleşen düşme düzeyi erkeklerde % 31.4, kadınlarda % 30.2'dir.

Ölümlülük, 35-64 yaş grubu için incelendiğinde; yaşa göre direkt standardize ölüm hızı, 1990'da erkekler için binde 8.70, kadınlar için binde 4.31; 2003'de erkekler için binde 6.07, kadınlar için binde 3.08'dir. Bu yaş aralığı için standardize ölüm hızlarının 1990 ve 2003 arasında gerçekleşen düşme düzeyleri, erkeklerde % 30.2, kadınlarda % 28.5'dir.

Tablo 1. Türkiyede il ve ilçe merkezlerinde erkekler için ölüm hızları, 1990

Yaş	Ölüm sayısı 1	Yıl ortası nüfus 2	Ölüm hızı (milyonda) 3 (1:2)	Avrupa stan- dard nüfusu 4	Beklenen ölüm sayısı 3*4
0	12154	321396	37816	1600	60.51
1-4	1895	1362571	1391	6400	8.90
5-14	1594	4082604	390	14000	5.47
15-24	2272	3664295	620	14000	8.68
25-34	3000	2908068	1032	14000	14.44
35-44	4637	2041983	2271	14000	31.79
45-54	8612	1208941	7124	14000	99.73
55-64	17221	911560	18892	11000	207.81
65-74	15402	354950	43392	7000	303.74
75-79	18497	96074	105233	2000	420.93
80-84		48359		1000	
85+		31339		1000	
Toplam	85284 ^a	17032140	Binde 5.01	100000	1162 ^b

^aYaşı belli olmayan 305 kişi, grup büyüklüğüyle orantılı biçimde, yaş gruplarına dağıtılmıştır.
^bKurgu Avrupa Standart Nüfusu, standart nüfus alındığında 1990'da Türkiye'de erkekler için il ve ilçe merkezleri direkt standardize ölüm oranı binde 11.62, 35-64 yaş için direkt standardize ölüm hızı: Binde 8.7'dir.

Tablo 2. Türkiyede il ve ilçe merkezlerinde kadınlar için ölüm hızları, 1990

Yaş	Ölüm sayısı 1	Yıl ortası nüfus 2	Ölüm hızı (milyonda) 3 (1:2)	Avrupa stan- dart nüfusu 4	Beklenen ölüm sayısı 3*4
0	9789	306182	31971	1600	51.15
1-4	1596	1293779	1234	6400	7.90
5-14	1111	3760619	295	14000	4.12
15-24	1195	3213959	372	14000	5.21
25-34	1529	2719440	562	14000	7.87
35-44	2269	1871539	1212	14000	16.97
45-54	4007	1174815	3411	14000	47.75
55-64	8741	929391	9405	11000	103.46
65-74	10790	423393	25485	7000	178.30
75-79	23981	120693	94460	2000	377.84
80-84		78165		1000	
85+		55017		1000	
Toplam	65008 ^a	15946992	Binde 4.08	100000	800.57 ^b

^aYaşı belli olmayan 196 kişi, grup büyüklüğüyle orantılı biçimde, yaş gruplarına dağıtılmıştır.
^bKurgu Avrupa Standart Nüfusu, standart nüfus alındığında 1990'da Türkiye'de kadınlar için, il ve ilçe merkezleri direkt standardize ölüm oranı binde 8, 35-64 yaş için direkt standardize ölüm hızı: Binde 4.31'dir.

Tablo 3. Türkiye'de ölüm kayıtlarına göre il ve ilçe merkezlerinde erkekler için ölüm hızları, 2003

Yaş	Ölüm sayısı 1	Yıl ortası nüfus 2	Ölüm hızı (milyonda) 3 (1:2)	Avrupa stan- dart nüfusu 4	Beklenen ölüm sayısı 3*4
0	7503	455524	16471	1600	26.35
1-4	1387	1822095	761	6400	4.87
5-14	1160	4485514	259	14000	3.63
15-24	2344	4415792	531	14000	7.43
25-34	2803	3904490	718	14000	10.05
35-44	5018	3160778	1588	14000	22.23
45-54	11751	2393822	4909	14000	68.73
55-64	17860	1347979	13249	11000	145.74
65-74	27021	836674	32296	7000	226.07
75-79	26235	232409	70552	2000	281.81
80-84		139446		1000	
85+				1000	
Toplam	103082	23194523	Binde 4.4	100000	796.91

Kurgu Avrupa Standart Nüfusu, standart nüfus alındığında;
35-64 yaş erkekler için direkt standardize ölüm hızı: Binde 6.069

Tablo 4. Türkiye'de il ve ilçe merkezlerinde kadınlar için ölüm hızları, 2003

Yaş	Ölüm sayısı 1	Yıl ortası nüfus 2	Ölüm hızı (milyonda) 3 (1:2)	Avrupa standart nüfusu 4	Beklenen ölüm sayısı 3*4
0	5375	410286	13101	1600	20.96
1-4	1041	1641145	634	6400	4.06
5-14	840	4561416	184	14000	2.58
15-24	1412	4633820	305	14000	4.27
25-34	1660	4199399	395	14000	5.53
35-44	2738	3354692	816	14000	11.42
45-54	5639	2485851	2268	14000	31.75
55-64	9630	1375665	7000	11000	77.00
65-74	19402	989514	19608	7000	137.26
75-79	33511	265478	69426	2000	277.70
80-84		217210		1000	
85+				1000	
Toplam	81248	24134476	Binde 3.4	100000	572.53

Kurgu Avrupa Standart Nüfusu, standart nüfus alındığında Türkiye'nin 35-64 yaş kadınlar için direkt standardize ölüm hızı: Binde 3,081

Tartışma

Bu çalışmada il ve ilçe merkezleri için binde 4.56 olarak bulunan 1990 yılı kaba ölüm hızı, 1989 TNA'sının, binde %7.09 olarak bulunduğu kentlere ait kaba ölüm hızından % 35 daha düşüktür. 1989 TNA, nüfusu 20 binden az olan yerleşim yerlerini kır olarak kabul etmiştir. Oysa nüfusu 20 binden az olan birçok ilçe merkezi vardır ve 1989 TNA kırsal kesim için kaba ölüm hızını binde 8.48 olarak bulmuştur. Bu nedenle bu çalışmada bulunan hızların en az % 35 düşük saptanmış olduğu düşünülebilir. 1989 TNA, bebek ölüm hızını kentler için binde 51.6 olarak bulmuştur. Bu çalışmanın il ve ilçe merkezleri için, her iki cinsiyet birlikte alındığında binde 34.96 olarak bulunduğu bebek ölüm hızı da, 1989 TNA değerinden % 32.2 oranında daha düşüktür. Bu çalışmada hızların düşük çıkmasında, ölüm istatistik formlarının eksik toplanması önemli bir etmen olmalıdır. Nüfus sayımlarında yerleşme yeri nüfusunu fazla gösterme gayretiyle yapılabilecek hileler ile diğer başka sorunlu durumlar da bu eksikliğe katkıda bulunmuş olabilir.

Türkiye il ve ilçe merkezleri için 2003 yılına ait erkeklerde binde 7,97 olarak bulunan yaşa göre standardize ölüm hızı, İngiltere'nin Avrupa standart nüfusu kullanılarak erkekler için hesaplanmış 2003 yılı yaşa göre standardize ölüm hızına (binde 7.99) benzer bir değerdir (12). Kadınlar için bu çalışmada bulunan 2003 yılı yaşa göre standardize ölüm hızı (binde 5,73) da İngiliz kadınlarının 2003 yılı yaşa göre standardize ölüm hızına (binde 5.58) benzemektedir. İngiltere'de, kalite kontrol mekanizmaları olan, tüm

ülkeyi kapsayan, mevzuatı ve örgütü olan, iyi işleyen bir ölüm kayıt sistemi vardır. Bu sistemin ayrıntılarına İngiltere'nin ulusal istatistikler internet sitesinin ilgili sayfalarından ulaşılabilir. İngiltere istatistiklerinin tam olduğu, Türkiye ölüm istatistikleriyle bulunan hızların ise gerçeği en az % 35 eksik yansıttığı düşünülürse Türkiye'de ölümlülüğün genel olarak İngiltere'den en az % 35 daha yüksek olduğu söylenebilir.

35-64 yaş aralığındaki ölümlülük bu çalışmanın odaklandığı konulardan biridir. Türkiye il ve ilçe merkezlerinin 35-64 yaş aralığındaki erkeklere ait yaşa göre standardize hızı bu çalışmada binde 6.07 olarak bulunmuştur. Oysa bu değer İngiltere'de 2003 yılında binde 4.63'tür. Türkiye'nin bu değeri, İngiltere'nin ilgili değerinden % 31 daha fazladır. Türkiye değerini binde 6.07 olarak değil de bunun % 35 fazlası olarak düşünürsek, Türkiye değeri İngiltere değerinden % 77 daha yüksek olur. Bu durum, 35-64 yaş aralığındaki Türk erkeklerinde ölümlülüğün yüksek olduğunu göstermekte, kronik hastalıkların önlenmesi, erken teşhisi ve tedavisine yönelik ulusal kontrol programlarına önem verilmesini gerektirmektedir.

Türkiye il ve ilçe merkezlerinin 35-64 yaş aralığındaki kadınlara ait yaşa göre standardize ölüm hızı bu çalışmada binde 3.08 olarak bulunmuştur. Bu değer İngiltere'de 2003 yılında binde 2.91'dir. Türkiye'nin bu değeri, İngiltere'nin bu değerinden % 6 kadar daha yüksektir. Türkiye'de kadınların sigara içme oranlarında geçen on yıllık dönemde ortaya çıkan hızlı artış(2), gerekli müdahale yapılmazsa, ileride Türk kadınların durumunu, Türk erkeklerinin yukarıda tanımlanan olumsuz durumuna doğru çekebilir.

1990 hızları ile 2003 hızlarının karşılaştırılması: 1990 ve 2003 yılları, yaşa ve cinsiyete özgü ölüm hızları ile karşılaştırıldığında, hem kadınlarda hem erkeklerde her yaş grubunda ölüm hızlarının düştüğü görülmektedir. En fazla düşme her iki cinsiyette de bebeklik dönemindedir. 1-4 yaş ve 55-64 yaş grupları da her iki cinsiyette ölüm hızlarının bebeklik döneminden sonra en çok düştüğü gruplardır. Ölüm hızlarının en az düştüğü yaş grubu, her iki cinsiyette de 15-24 yaş grubudur. Yaşa göre standardize hızlarla, 2003 yılı ölüm hızı, 1990 ölüm hızından erkeklerde % 31, kadınlarda % 28 daha düşüktür. 35-64 yaş aralığı için yaşa göre standardize hızlarla 2003 yılı ölüm hızı, 1990 yılı ölüm hızından, erkeklerde % 30, kadınlarda % 29 daha düşüktür. Bu bulgular, 1990 ve 2003 yılları ölüm kayıtlarının aynı eksiklik düzeyinde olması sayılıyla, Türkiye'de ölümlülüğün son 13 yılda yaklaşık % 30 düzeyinde düştüğünü düşündürmektedir.

Sonuç

Türkiye'de ölümlülüğü değerlendirme ve izleme yönünden önemli bir eksiklik vardır. Ölüm kayıt sistemi yetersizdir, mevcut ölüm kayıtları yaklaşık üçte bir oranında eksiktir. Düzenli aralıklarla yapılan ve ölümlülüğü ulusal düzeyde değerlendiren ve izleyen araştırmalar yoktur. Bu durum sağlık hiz-

metlerinin etkili sunumunu engelleyen temel eksikliklerden biridir. Türkiye’de bebek ve çocuk ölümlülüğünün yanı sıra 35 yaş sonrası dönem ölümlülüğü de yüksek düzeydedir. Bu nedenle bu yaş grubunun ölüm nedenleri olan koroner kalp hastalığı, serebrovasküler hastalık ve kansere yönelik ulusal müdahale programlarının başlatılmasına ve güçlendirilmesine gerek vardır.

Kaynaklar

1. Hacettepe Üniversitesi Nüfus Etüdüleri Enstitüsü. Türkiye’de Nüfus Yapısı ve Nüfus Sorunları 1973 Araştırması. Hacettepe Üniversitesi Yayınları D-25, Ayyıldız Matbaası, Ankara, 1978.
2. Hacettepe Üniversitesi Nüfus Etüdüleri Enstitüsü. Türkiye Nüfus ve Sağlık Araştırması 2003. Hacettepe Üniversitesi Nüfus Etüdüleri Enstitüsü, Ankara, 2005.
3. Hacettepe Üniversitesi Nüfus Etüdüleri Enstitüsü. Türkiye Nüfus ve Sağlık Araştırması 1988. Hacettepe Üniversitesi Nüfus Etüdüleri Enstitüsü, Ankara, 1989.
4. T.C. Başbakanlık Devlet İstatistik Enstitüsü. 1989 Türkiye Nüfus Araştırması. Devlet İstatistik Enstitüsü Matbaası, Ankara, 1991.
5. Sümbüloğlu V, Sümbüloğlu K. Sağlık Hizmetlerinde Veri Toplama ve Değerlendirme Yöntemleri. Aydoğdu Ofset, Ankara, 1995.
6. Fişek HN. Halk Sağlığına Giriş. Çağ Matbaası, Ankara, 1983.
7. ONAT A. Erişkinlerimizde kalp hastalıkları prevalansı, yeni koroner olaylar ve kalpten ölüm sıklığı. ONAT A (editör):TEKHARF Türk erişkinlerinde kalp sağlığı, risk profili ve kalp sağlığı. Yenilenmiş ikinci baskı, Ohan Matbaacılık, s:16-23, 2000.
8. U.S. DHHS, National Institutes of Health, National Heart, Lung, Blood Institutes. Morbidity and Mortality: 2004 Chart Book on Cardiovascular, Lung and Blood Diseases. May 2004.
9. United States Department of Health and Human Services. The health consequences of smoking: A report of the Surgeon General. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2004.
10. T.C. Başbakanlık Devlet İstatistik Enstitüsü. Ölüm İstatistikleri İl ve İlçe Merkezlerinde 1990. Devlet İstatistik Enstitüsü Matbaası, Ankara, 1993.
11. T.C. Başbakanlık Devlet İstatistik Enstitüsü. 1990 Genel Nüfus Sayımı: Nüfusun Sosyal ve Ekonomik Nitelikleri. Devlet İstatistik Enstitüsü Matbaası, Ankara, 1993.
12. Office for National Statistics, UK. Mortality Statistics: Review of the Registrar General on Deaths by Cause, Sex, and Age in England and Wales, 2003. Series DH2 no.30, <http://www.statistics.gov.uk>

DİSKİRMINANT (AYIRMA) ANALİZİ İLE EL KULLANIMI TAYİNİ

Bülent Ediz¹, Şengül Cangür¹, Gökhan Ocakoğlu¹, İlker Ercan¹,
N.Şimşek Cankur², İhsaniye Coşkun², İsmet Kan¹

Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik AD.¹

Uludağ Üniversitesi Tıp Fakültesi Anatomi AD.²

Bilimin temel yöntem bilimlerinden birisi de karmaşık durumları daha makul duruma indirgeyerek sınıflandırmaktır. Nesnelerin sınıflandırılması belki de tüm bilimsel çabaların temelini oluşturan bir çalışmadır. Sınıflandırma problemi, araştırmacının bir birey üzerinde bireyin çeşitli özellikleri bakımından ölçüm yapması ve bu bireyi elde ettiği ölçümlere dayanarak sayısı önceden bilinen gruplardan birisine yerleştirmek istemesiyle ortaya çıkmıştır. Botanik ve antropolojideki sınıflandırma problemleri “**irksal benzerlik katsayısı**” ve “**ayırma fonksiyonu**” ‘nun, dolayısıyla çok değişkenli istatistikte “**ayırma analizi(AA)**”nin doğuşuna neden olmuştur.

En çok kullanılan elin saptanması amacıyla, Uludağ Üniversitesi Tıp Fakültesi Anatomi Anabilim Dalı koleksiyonunda bulunan, 120 humerus'ta (62 sol, 58 sağ) sulcus intertubercularis ölçümleri yapıldı. Ölçüm parametreleri olarak sulcus intertubercularis'e ait genişlik, derinlik, iç duvar açıları ve dış duvar açıları alındı. Elde edilen ölçümlere ayırma (diskriminant) analizi uygulandı. Ayırma analizi ile iç ve dış duvar açılarına ait değerlerin, sağ ve sol el ayırımında anlamlı etkisinin olduğu saptandı ($P<0,001$). Ancak, genişlik ve derinlik değerleri el ayırımında anlamlı etkiye sahip değildi ($P>0,05$).

Anlamlı bulunan değişkenlerle elde edilen ayırma fonksiyonunun doğru sınıflama oranı % 85,8 olarak bulundu. Bu sonuçlar iç ve dış duvar açılarının ölçümlerinin, sağ el kullanımının baskınlığını belirlemede değerli kriterler olduğunu düşündürdü.

Anahtar Kelimeler: Diskirminant (ayırma) analizi, El tercihi, sulcus intertubercularis.

İKİ YANIT DEĞİŞKENLİ DEĞİŞİK VERİ SETLERİNDE LOJİSTİK REGRESYON VE DİSKİRMİNANT ANALİZİ'NİN DOĞRU SINIFLANDIRMA ORANLARI:

Bir Simülasyon Çalışması

**Bülent Ediz, Güven Özkaya, Deniz Sığırlı, Şengül Cangür,
Gökhan Ocakoğlu, İlker Ercan, İsmet Kan**

Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik AD

e-mail: ediz@uludag.edu.tr

Gözlenen verilerin analiziyle ilgilenildiğinde, kuramsal istatistik bir modelin matematik fonksiyonlarla ifade edilmesi gereği ortaya çıkar. Bu fonksiyonlar, gözlenen verilerden hareketle gelecekteki olaylar hakkında tahmin yapılmasına ve olaylara etki eden faktörlerin belirlenmesine olanak sağlarlar. Olaylara çözüm getirmede, aynı amaç için farklı modeller gerekebilir. Ancak, aynı amaç için farklı modellerin kullanılması bazı risklere yol açabilir. Tercih belirleyecek etken, riski en az olan yol seçilerek bu riskleri minimum yapan daha etkin modelin kullanılması olmalıdır. Gözlemleri, verilerin yapısında bulunan olası gruplara atamak için kullanılan yöntemlerden ikisi, ayırma (discriminant analysis) analizi (**AA**) ve lojistik regresyon analizleridir (Logistic Regression Analysis)(**LRA**).

Bu amaçla biri iki yanıt düzeyli bağımlı (binary), 10'da bağımsız değişkenden oluşan 11 değişkenli veri setleri oluşturuldu. Verilerin üretiminde GENSTAT 8.0 for Windows programı kullanıldı. Veri setleri; büyüklükleri (n) 10, 20, 30, 50, 100, 300, 500, 1000, 2000, 5000 ve 10000, bağımlı değişkende ilgilenilen özelliğin ortaya çıkma olasılığı (p) 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 ve 0.9, bağımsız değişkenlerin tümü sürekli, tümü kesikli ve 5'i sürekli 5'i kesikli (karışık) değişkenlerden oluşacak şekilde veri setleri oluşturuldu. 11 farklı büyüklük, 9 farklı bağımlı değişken oranı ve 3 farklı veri tipinde toplam 297 veri seti incelendi.

Yapılan çalışmalar sonucunda LRA ve AA'nin $n < 20$ için kullanılması durumunda sınıflandırma sorunu yaşanmaktadır. Kesikli-Sürekli-Karışık bağımsız değişkenlerden oluşan veri setlerinde; $n=20$ olduğunda LRA ve AA tüm veri setlerinde çok yakın doğru sınıflandırma oranlarıyla gözlenmektedir ($P > 0.05$). $n=30$ için; kesikli veri setlerinde bağımlı değişkenin gözlenme olasılığı 0.1, 0.2, 0.4, 0.7, 0.8 ve 0.9 için LRA daha yüksek sınıflandırma

oranına sahipti ($P < 0.01$), sürekli veri setlerinde bağımlı değişkenin gözlenme olasılığı 0.6 iken LRA daha yüksek doğru sınıflandırma oranı bulundu ($P < 0.001$), karışık veri setinde ise 0.4-0.8 aralığında LRA AA'nden anlamlı derecede daha yüksek sınıflandırma oranına sahipti ($P < 0.01$). $n \geq 50$ iken tüm veri setlerinde bağımlı değişkenin gözlenme olasılığı $p = 0.5$ olması durumunda LRA ve AA doğru sınıflandırma oranları aynı ya da çok yakın çıkmaktadır ($P > 0.05$). Bağımlı değişkenin gözlenme olasılığı 0.5'ten uzaklaştıkça LRA'nin doğru sınıflandırma oranı AA'dan anlamlı derecede yüksektir ($P < 0.001$).

Anahtar Kelimeler: Lojistik Regresyon, Diskirminant (ayırma) analizi, Doğru Sınıflandırma Oranı, Simülasyon.

ARAŐTIRMA ETİK KURULLARINDA BİYOİSTATİSTİK UZMANININ ÖNEMİ

Elif Atıcı, Sezer Erer

Uludağ Üniversitesi, Tıp Fakültesi, Deontoloji Anabilim Dalı, 16059 Görükle-BURSA

e-mail: elifatici@uludag.edu.tr

Özet

Araştırmacı ya da hekimin, denek ya da hastanın haklarının gözetilmesi, yapılan tıbbi uygulama ve araştırmaların bilimsel standartları taşınması, yasalara ve ahlaki değerlere uygunluğunun denetlenmesi gereklidir. Araştırma etik kurulları, insan üzerinde yapılacak araştırmalarda etik standartların ve bilimsel değerin sağlanmasında önemli bir rol üstlenirler. Bilimsel değeri olmayan bir araştırma aynı zamanda etik ilkelere de aykırıdır. Biyomedikal araştırmalarda geçerli ve güvenilir sonuçların elde edilmesinde istatistik biliminden yararlanma bilimsel ve etik yükümlülüktür. Bu nedenle, biyoistatistiğin yoğun olarak kullanıldığı araştırmaları etik ve bilimsel açıdan değerlendiren araştırma etik kurullarında biyoistatistik uzmanının bulunması gereklidir.

Bu çalışmanın amacı, araştırma etik kurullarında bir biyoistatistik uzmanının bulunmasının, araştırmaların bilimsel değerlendirilmesinin yanı sıra etik değerlendirilmesinde de gerekli olduğunu vurgulamaktır.

Anahtar Kelimeler: Biyoistatistik, Etik, Araştırma Etik Kurulları.

Giriş

Araştırma etik kurullarının amacı, insanlar üzerinde yapılan biyomedikal araştırmaların (farmasötik araştırmalara, cerrahi yöntemler, medikal araçlar, tıbbi radyasyon, tıbbi kayıtlar, biyolojik örnekler, epidemiyolojik, sosyal ve psikolojik araştırmalar gibi) etik ve bilimsel standartlara uygunluğunu değerlendirmektir. Araştırma etik kurullarına sunulan araştırma projeleri, etik açıdan ve bilimsel açıdan değerlendirilirken istatistik yöntemlerinin doğru kullanılıp kullanılmadığı da göz önüne alınmalıdır. Bunun değerlendirilmesi amacıyla araştırma etik kurullarında biyoistatistik uzmanının bulunması gereklidir, çünkü istatistiğin doğru kullanımı, araştırmacının bilimsel niteliğini artırır. Araştırmacının bilimsel kurallara uygun olarak planlanması, yürütülmesi ve değerlendirilmesi, aynı zamanda araştırmacının etik açıdan da kabul edilebilmesinin koşullarından birini oluşturur. Unutulmamalı ki, etik değerler ve bilimsel doğrular bilimin ayrılmaz parçalarıdır.

Bu çalışmayla amaçlanan, araştırma etik kurullarında bir biyoistatistik uzmanının bulunmasının, araştırmaların bilimsel değerlendirilmesinin yanı sıra etik değerlendirilmesinde de gerekli olduğunu vurgulamaktır.

Araştırma Etik Kurullarının Önemi ve Biyoistatistik Uzmanının Gerekliliği

Ahlaki ilkelerin, teknolojik ve bilimsel gelişmenin gerisinde kalması, yeni etik sorunların ortaya çıkmasına neden olur. Bu durumu engellemek için oluşturulan mekanizmalar öncelikle etik ilkelere sahip çıkılmasını sağlayacak etik eğitiminin verilmesi, etik yaklaşımın yasalar ile güvence altına alınması ve denetlemenin yapılmasıdır. Geçmişte insanlar üzerinde yapılan etik dışı tıbbi denemelerin, bilim dünyasında ve toplumda yarattığı rahatsızlık ve tartışmalar denetleme mekanizmalarının oluşturulmasını gerekli kılmıştır. Tarihsel sürece bakıldığında etik değerlerin belirlenmiş olmasına karşın, uygulamada sorunlar yaşandığı görülmektedir. Araştırmacı ya da hekimin, denek ya da hastanın haklarının gözetilmesi, yapılan tıbbi uygulama ve araştırmaların bilimsel standartları taşıması, yasalara ve ahlaki değerlere uygunluğunun denetlenmesi gerekli hale gelmiştir. Bu nedenle klinik uygulama ve ilaç araştırmalarının bilimsel niteliğinin değerlendirilmesinde danışmanlık hizmeti veren etik kurullar oluşturulmuştur (1). "Araştırma"; insanın bilgi edinmesinde, teknolojiye ilerlemenin sağlanmasında çevreyi tanıyarak ondan en iyi biçimde yararlanmasında ve sorunlarını çözmesinde başvurduğu önemli bir etkinliktir. Bilinmeyen olay ve etmenleri ortaya çıkarabilmek, bilinenleri geliştirmek, herhangi bir konuyu aydınlatmak, bir soruna çözüm getirmek, belirli kavram, kuram ya da yasalara ulaşabilmek amacıyla yapılan bilimsel bir çalışma olarak da tanımlanabilir. "Araştırma süreci" ise, çağın bilim çevrelerince uygunluğu kabul edilmiş bilimsel yöntemlerin kullanılmasıyla, yeni bulgular elde etmesine yönelik bir sorun çözme etkinliği olarak kabul edilebilir. Bilimsel bir araştırmada en önemli konu araştırmanın bilimsel yöntemlerle planlanması, yürütülmesi ve sonuçlandırılmasıdır. Ancak bilimsel yöntemlerden yararlanılarak yapılan araştırmanın sonucu geçerli ve güvenilir olabilir (2).

Araştırma etiği yalnızca bilimle sınırlı soyut bir kavram olarak görülmemeli, genel etik değerlerin özel bir alandaki yansımaları oluşturduğu da göz önüne alınmalıdır. Bilginin kendisi yansızdır. Ancak, toplumsal uygulamada ve bilginin teknolojiye dönüşümünde bilgi yansız olmaktan çıkar. Bu nedenle bilimsel bilgiyi ortaya çıkaracak olan araştırmacının etik olarak toplumca onaylanabilirliğinin de değerlendirilmesi gerekmektedir (3). Araştırmaların sonucunda elde edilen bilginin yansız olarak sunulabilmesi için araştırmanın planlanmasından değerlendirilmesine kadar geçen her aşamasında biyoistatistik ilke ve yöntemlerinin doğru bir şekilde kullanılması büyük önem

taşımaktadır. Bunun değerlendirmesini yapabilecek olan ise biyoistatistik uzmanıdır.

Araştırma etik kurulları, insan üzerinde yapılacak araştırmalarda etik standartların ve bilimsel değerin sağlanmasında önemli bir rol üstlenirler. İnsan üzerinde yapılan deneylerle ilgili etik sorunların kurumsallaşmasına karşın hasta ve denek için, araştırmacı ve araştırmacı kurum için ortaklaşa etik ve hukuksal denetim oluştururlar. Asıl amaç deneğin korunmasıdır. Koruma işleminden birinci derecede sorumlu olan araştırmacı, koruma işlemini denetleme ve yönlendirme durumunda olan yapılanmaların başında ise araştırma etik kurulları gelmektedir. Temel olarak araştırma etik kurullarının görevi, araştırma protokollerinin etik standartlara uygunluğunu kontrol etmektir (4).

Araştırma etik kurulları, üç önemli sorumluluğa sahiptir. İliki ve en önemlisi, kurulun denek haklarının korunduğundan emin olmasıdır. Bunun sağlanmasında deneklerin hiçbir baskı altında kalmadan, özgür iradeleri ile, araştırmacının yarar ve zararları hakkında anlaşılır ve tam olarak bilgilendirilmelerinin ve onamlarının alınması önem taşır. İkincisi, kurulun araştırmaya kaynak sağlayarak sonuçlardan etkilenecek olan topluma karşı olan sorumluluğudur. Üçüncü olarak kurul araştırmacıya karşı sorumludur. Bu sorumlulukların yerine getirilmesi ile araştırmacının etik ve bilimsel standartlarının yüksek tutulması sağlanır (4, 5).

İnsan denekler üzerinde biyomedikal araştırma yapmanın etik ve bilimsel standartlarının saptanması amacıyla Helsinki Bildirgesi, İnsan Deneklerle İlgili Biyomedikal Araştırmaların Uluslararası Etik Yol Gösterici Kuralları (CIOMS), Dünya Sağlık Örgütü (WHO) ve İyi Klinik Uygulamalar Konusunda Yol Gösterici Kurallar (ICH) gibi uluslararası yol gösterici kurallar oluşturulmuştur. Bu kuralların amacı, öncelikle deneğin haklarının korunmasını sağlamak, araştırmacının bilimsel ve etik açıdan değerlendirilmesinde yardımcı olmaktır.

Araştırma etik kurulları deneklerin güvenliği, sağlığı ve insan haklarının korunmasının, bilimin ve toplumun çıkarlarının hiçbir zaman deneğin sağlığı ile ilgili kaygıların üzerine çıkarılmamasının ve deneklerin olası bir kullanımının önlenmesinin temel güvencesidirler. Bu güvencenin sağlanması, Helsinki Bildirgesi ve ulusal düzenlemelerde tanımlandığı biçimde araştırma protokollerinin araştırma öncesi değerlendirilmesi ile olur. Araştırma etik kurulları, araştırma protokollerinin sadece etik ölçülere uygunluğunu denetlemekle kalmaz, aynı zamanda araştırmacının amacını, gerekçesini, yaklaşım ve metodolojisini göz önünde tutarak bilimsel açıdan da inceler. Bilimsel değeri olmayan bir araştırma etik ilkelere de aykırı olacağından araştırma projesinin etik yönleri ve bilimsel yönleri birbirinden ayrılamaz. Araştırmacının planlama, yürütülme ve değerlendirme aşamalarında yapılacak hataların, araştırmacının bilimselliğine gölge düşürebileceği göz önüne

alındığında araştırma projelerinin değerlendirilmesinde biyoistatistik uzmanına danışılması araştırmanın bilimsel düzeyinin değerlendirilmesinde önemli bir rol üstlenebilir. Bu nedenle araştırma etik kurullarında biyoistatistik uzmanına yer verilmelidir (1).

Türkiye’de Araştırma Etik Kurulları

Türkiye’de ilk araştırma etik kurulu 1986 yılında kurulmuştur. Araştırmaların etik kurul onayını aldıktan sonra başlamaları ve yine etik kurulun gözetiminde sürdürülmeleri gerekliliği ise yasal düzeyde 1993 ten beri bulunmaktadır. T.C. Sağlık Bakanlığı tarafından geliştirilen ve 23 Ocak 1993 tarih, 21480 sayılı resmi gazetede yayımlanan “İlaç Araştırmaları Hakkında Yönetmelik”, Türkiye’de etik kurulların oluşturulması sürecinde dönüm noktasını oluşturur. Adı geçen yönetmelikle birlikte 25 Mayıs 1994 tarih ve 21942 sayılı resmi gazetede yayınlanan “Farmasötik Müstahzarların Biyoyararlanım ve Biyoeşdeğerliliğinin Değerlendirmesi Hakkında Yönetmelik”, klinik araştırmaları düzenleyen esasları belirler. Sağlık Bakanlığınca 29 Aralık 1995 tarih ve 51748 sayılı resmi gazetede yayınlanan “İyi Klinik Uygulamaları Kılavuzu” ve “İyi Laboratuvar Uygulamaları Kılavuzu” ile araştırmaların uluslararası bilimsel ve etik standartlarda yapılmasını sağlamak üzere laboratuvar çalışmaları ve/veya klinik araştırmalarda uyulması gereken kurallar ve izlenecek yol belirlenmiştir.

İlaç Araştırmaları Hakkında Yönetmelik’te, ilaçla yapılacak klinik araştırma isteğinin incelenebilmesi, klinik araştırma ve çalışmaların etik açıdan değerlendirilmesi ve kontrolü amacıyla, Sağlık Bakanlığının merkezinde Bakanlıkça oluşturulan kurul *merkezi etik kurul*, araştırmayı yapacak kurum ya da kuruluşlarda oluşturulan ve bu yönetmelikte belirlenen görevleri yürüten kurullar ise *yerel etik kurullar* olarak tanımlanır. Herhangi bir ilaç konusunda bilimsel amaçla klinik araştırma yapmak isteyen kişi, kurum ya da kuruluşlar, bu yönetmelik hükümlerine uygun olarak yerel etik kurul aracılığı ile Sağlık Bakanlığına başvuruda bulunmak ve izin almak zorundadırlar. İnsan üzerinde araştırma yapılabilmesi için etik kurul onayının alınması yasal bir zorunluluktur (6).

Sağlık Bakanlığı bünyesinde kurulan merkezi etik kurulunun ve araştırmanın yapılacağı eğitim hastanelerinde kurulan yerel etik kurulların hangi üyelerden oluşacağı ve üyelerin özellikleri ilgili yönetmelikte belirtilmiştir. Bu üyeler klinisyen, farmakolog, eczacı, biyokimya uzmanı, patolog, yapılacak çalışmanın özelliğine uygun bir uzman ve mümkünse bir tıbbi deontoloji uzmanıdır. Üyelerin arasında biyoistatistik uzmanı yer almamaktadır.

Yasal düzenlemeler bulunmakla birlikte etik kurulların işleyişi ve yapılanmaları ile ilgili bazı sorunlar bulunmaktadır. Bunlar :

1. İlgili yönetmeliklerin sadece ilaçlarla ilgili insan üzerinde yapılacak arařtırmaları ele almıř olması,
2. Yerel etik kurulların eđitim hastanelerinde yer almalarında ele alınacak kriterlerin belirlenmemiř olması,
3. Karar verme sũrecinin iřleyiři aısından yerel etik kurullarca gũrũř birliđine varılmıř bir yũntemin olmaması,
4. Etik kurul kararlarında gecikmenin olabilmesi,
5. Bazı yerel etik kurullarda arařtırmalarda bilimsel niteliklerin dikkate alınmaması,
6. Etik kurullarda biyoistatistik uzmanının ȳngȳrũlmemesi.

Etik kurullarda ortak bir yapılanma ve iřleyiř biimi oluřturulması ve bunun tũm ũlkelerdeki arařtırma etik kurul yȳnergeleri ile uyum iinde olması kuřkusuz yararlı olacaktır. Son iki maddede yer alan sorunlar, ele aldığımız konu ile iliřkilidir. Arařtırmaların bilimsel niteliklerinin deđerlendirilmesinde planlamadan deđerlendirmeye kadar arařtırmanın her ařamasında biyoistatistik ilke ve yũntemlerinin dođru kullanımı ȳnem tařımaktadır. Bunun deđerlendirmesini yapabilecek bilgi ve yeteneđe sahip olan ise biyoistatistik uzmandır. Arařtırmaların bilimsel niteliklerinin deđerlendirilmesi aynı zamanda etik deđerlendirmenin kořullarından biridir ve bu nedenle etik kurullarda biyoistatistik uzmanı bulunmalıdır.

24 Temmuz 2003 tarih ve 25178 sayılı resmi gazetede yayınlanan *Avrupa Birliđi Mũktesebatının ũstlenilmesine İliřkin Tũrkiye Ulusal Programı'nda İla arařtırmaları hakkında Yȳnetmeliđin ve Farmasȳtik Mũstahzarların Biyoyararlanım ve Biyoeřdeđerliliđinin Deđerlendirmesi Hakkında Yȳnetmeliđin yũrũrlũkten kaldırılması ve 2001/20/EC sayılı Avrupa Parlamentosu ve Konsey Direktifi'ne uyumlu yeni yȳnetmeliđin yayınlanması ȳngȳrũlmũřtũr. Bu amala Beřeri Tıbbi ũrũnlerin Klinik Arařtırmaları Hakkında Yȳnetmelik Taslađı hazırlanmıřtır. Bu taslak sadece ila arařtırmalarını kapsamamaktadır. Bȳylece yukarıda sıraladıklarımız arasındaki ilk sorun ȳzũmlenmektedir. Aynı zamanda bu taslađa gȳre, merkezi etik kurulda biostatistik uzmanı yer almaktadır. Ancak yerel etik kurullarda ne biyostatistiki ne de etik uzmanı yer almamaktadır. Etik uzmanı ve biyostatistiki arařtırmaların etik ve bilimsel deđerlendirilmesinde ȳnemli bir yere sahiptir. Bunların yerel etik kurullarda yer almamaları taslađın ȳnemli bir eksikliđidir.*

Arařtırma Etik Kurullarında Etik ve Bilimsel Deđerlendirmenin Yapılması

Arařtırmaya dayanan tıbbi geliřme, insan denekler ũzerinde deneme yapılmasını gerekli kılmaktadır. Bununla birlikte, insan denekler ũzerindeki biyomedikal arařtırmalar tanı koyucu, iyileřtirici ve koruyucu iřlemleri geliřtirme, hastalıđın etyopatogenezini anlama amalı olmalıdır. İnsan ũzerindeki arařtırmaların etik boyutunun erevesini izen Helsinki Bildirgesi'nde; arařtırmaların genel bilimsel ilkelere uyması gerektiđi, insan ũzerindeki

araştırmalarda bilimin ve toplumun yararının, hiçbir zaman deneğin sağlığına verilen önemden daha öncelikli olamayacağı belirtilmektedir. Helsinki Bildirgesi'nde son olarak 2000 yılında değişiklikler yapılmıştır. Bildirgenin 13. maddesinde "İnsan üzerinde yapılacak olan her araştırma niteliğindeki yöntemde, araştırmanın düzenlenmesi ve uygulanmasının bir araştırma protokolüne bağlı olması ve bu protokolün değerlendirme, yorumlama ve yol göstermesi amacı ile bu konuda özelleşmiş bir etik kurula sunulması ve bu kurulun araştırmacı, destekleyici ve her tür etkiden uzak, bağımsız bir kurul olması" öngörülmektedir. Yapılan bu son değişiklikler ile araştırmanın yöntemine de müdahale edilmesi söz konusu olmaktadır (7). Bu açıklamadan yola çıkarak biyomedikal araştırmaların etik kurullar tarafından iki yönden değerlendirilmesi gerektiğini söyleyebiliriz:

- 1- Etik değerlendirme
- 2- Bilimsel değerlendirme

Etik Değerlendirme

İnsan deneklerin yer aldığı biyomedikal bir araştırmada göz önüne alınması gereken etik ilkeler; özerkliğe saygı (buna bağlı olarak aydınlatılmış onam), yararlı olma (buna bağlı olarak zarar vermeme), adalet ve gizlilik ilkeleridir. Bu ilkeler insan üzerinde yapılan denemelerin ahlaki yönüne şekil verirler.

Özerkliğe saygı ilkesi, araştırmada yer alan deneklerin aydınlatılmış onamlarının alınmasını ve araştırmanın risklerinden denekleri korumayı amaçlar. Yararlılık ilkesi, araştırmanın amacına uygun kar-zarar oranının gösterilmesine dayanır. Adalet ilkesi araştırmanın zarar ve yararının eşit olarak dağıtılmasını gerektirir. Gizlilik ilkesinin gereği ise, deneklerin kişilik haklarına saygı göstererek, kimlik bilgilerinin gizli tutulmasıdır (8).

Önceden denenmemiş olan ve insan üzerindeki etkileri kesinlikle bilinmeyen uygulamaların tedavi ya da bilimsel amaçla sınırsız olarak uygulanması hasta-deneğin hayatına, dokunulmazlığına büyük ölçüde zarar verebilir. Tıbbi denemeleri uygulayan araştırmacının sorumsuzluğu geri dönülmez olumsuz sonuçlara neden olabilir. Kişinin beden bütünlüğü üzerindeki girişimlerin etik açıdan olduğu kadar yasalarca da geçerli olabilmesi için bazı temel koşulların sağlanması gerekir. Bu koşulların ilki, hasta-deneğin geçerli onamının olmasıdır. Hasta-deneğin hiçbir kuşkuya yer vermeden uygulama ve bunun doğuracağı sonuçlar ve riskler hakkında bilgilendirilmesi, onamı verirken herhangi bir baskı altında kalmaması, kişinin içinde bulunduğu maddi, manevi ve fiziksel koşullardan faydalanılmaması verilen onamı geçerli kılar. İkinci önemli temel koşul, tıbbi denemelerden hasta-deneğin zarar görmemesi ve yaşamının tehlikeye atılmamasıdır. Hastalığın tedavisi için bilinen, bilimsel tıp tarafından onaylanmış olan tedavi yöntemlerinin sonuç vermediğinden emin olmak gerekir. Tedavi etme amacıyla ya da

bilimsel bilgi sağlama amacıyla girişimde bulunacak araştırmacı-hekimin bu konuda yeterli bilgi, yetenek ve beceriye sahip olması beklenir (9).

Adalet ilkesi gereğince; yaş, cinsiyet, ekonomik statü, kültür ve etnik değerlendirmeler de dikkate alınarak araştırmanın yarar ve sıkıntıları toplumdaki tüm grup ve sınıflar arasında eşit dağıtılmalıdır (10).

Araştırmanın planlanmasından sunumuna kadar geçen her aşamada karşılaşılabilecek etik sorunları üç ana başlık altında toplamak olanaklıdır:

1. *Araştırmanın planlanması ve uygulanması sırasında gözlenen etik sorunlar:*

1.a. *Araştırmanın amacı:* Araştırmanın toplumsal ve bilimsel gereksinimlere ne derece hizmet ettiği karşılaşılan önemli etik sorunlardan biridir. İnsan üzerinde yapılan deneyler, kişiler hakkında bilgi toplamak ya da kişilerin araştırmaya fiziksel katılımları ile olanaklıdır. Birinci durumda, kişilerin özel yaşamlarına, kültürel yapılarına etki söz konusu iken ikinci durumda az ya da çok beden bütünlüğüne müdahale ön plandadır.

Bu aşamada ortaya çıkabilecek etik sorunlar, özerkliğe saygı ile bağlantılı olarak aydınlatılmış onam ve yararlı olma ilkeleri açısından ele alınabilir. Araştırmanın amacı öncelikle deneğin yararı gözetilerek, topluma ve bilime katkı sağlaması olmalıdır. Klinik denemenin planlanması ve yürütülmesindeki asıl zorluk, deneye katılan hastalara en iyi ve en uygun tedaviyi sağlarken aynı zamanda oluşturulan veri ve geliştirilen tedaviden değerli sonuçlar alma çabasıdır.

Araştırmalar iki nedenle yapılır: 1-Doğrudan deneğin tedavisi, 2-Geliştirilmek istenen tıbbi uygulama ya da tedavinin sağlıklı denekler üzerinde denenmesi. Her iki durumda da öncelikle dikkat edilmesi gereken nokta deneklerin aydınlatılmış onamlarının alınması olacaktır. Denekler uygulamanın sonuçları ve riskleri hakkında açıkça bilgilendirilmeli, hiçbir maddi, manevi baskı altında olmadan onamları alınmalıdır. Bilgilendirmenin yapılması tek başına yeterli değildir. Bunun denek tarafından doğru şekilde anlaşıldığından da emin olmak gerekir. Deneğin tedavisine yönelik bir araştırmada A tedavisi ile iyi sonuç alındığı bilindiği halde, araştırmanın konusu olan ancak kar-zarar oranı tam olarak bilinmeyen B tedavisinin uygulanması etik açıdan yararlı olma ve buna bağlı olarak zarar vermeme ilkelerine uygun olmayacaktır.

1.b. *Araştırma yönteminin yer, zaman denekler ve bilimsellik açısından etik ilkelere uygunluğu:* Araştırmanın amacı uygulanacak yöntemle uyum içinde olmalıdır. Araştırmaya katılacak deneklerin çalışmaya alınma ve alınmama kriterleri açık olarak belirtilmeli, seçilen kontrol grubunun gerçek bir kıyaslamaya uygun olması sağlanmalı, deneklere ait özellikler çalışma ve kontrol grupları arasında eşit olarak dağıtılmalı, gözlemler gerekli

nesnellik içinde yapılmalıdır. Araştırmalarda karşılaşılan önemli sorunlardan birisi de, araştırmacının isteyerek ya da istemeden yanlı davranmasıdır. Yanlı davranma araştırma yönteminin seçiminde, değerlendirme ve yayın aşamalarında görülebilir.

Bu aşamada yararlı olma ve adalet ilkelerine uygun olarak hareket etmeye dikkat etmek gerekir. Araştırma sonuçlarının istenilen şekilde çıkması amacıyla araştırmacının denek seçiminde yanlı davranması ile araştırma sonuçları bilimsellik ve gerçeklikten uzak olacağı gibi etik açısından da onaylanamaz.

1.c. *Araştırmanın mali desteği ile ilgili etik sorunlar:* Araştırma yapılırken öncelikle toplum çıkarları göz önüne alınmalıdır. Araştırmalara mali kaynak sağlayan şirketlerin yaklaşımları kar amacı doğrultusunda olduğunda, kendi çıkarlarıyla ya da beklentileriyle uyuşmayan araştırmalara destek olmama, istenmeyen sonuçlar çıktığında yayınlanmasını engellerek ya da geciktirerek yanlı davranma ve toplumu yanıltma gibi yaklaşımlarda bulunabilirler. Bu durumun engellenmesi için mali desteğin bir araştırmaya özgü değil genel bir araştırma fonuna verilmesi sağlanabilir. Bu konuda karşılaşılan diğer bir sorun ise tedavi hizmetlerine yönelik araştırmaların halk sağlığını korumaya yönelik araştırmalardan daha kolay mali kaynak bulabilmesidir. Bu durum, araştırmaların desteklenmesinde kar amacının ağırlığı olduğunun kanıtlarından biri olarak görülebilir.

Böyle bir durumda yararlı olma ilkesinin ihlali söz konusu olacaktır. İlaç geliştirilmesi ve pazarlanması trilyon dolarlık bir endüstri haline gelmiştir. Bu durum finansal taleplerin klinik denemelerin tasarlanmasının ve planlanmasının etkilenmesine neden olabilir. Şirketlerin desteklediği araştırmalarda sonuçların yanlı yansıtılması ve negatif sonuçların açıklanmaması araştırma konusunun yanlı yorumlanmasına, diğer araştırmacıların zaman kaybetmesine ve hastaların zarar görmesine neden olabilir.

2. *Araştırma sonuçlarının değerlendirilmesinde gözlenen etik sorunlar:* Sonuçların değerlendirilmesinde sık yaşanan sorunlardan biri çalışmaya alınmayan, çalışmadan çıkarılan ya da çalışma sırasında kayıtları kaybolan deneklere ait sonuçları hiç değerlendirmemek ya da bunlardan hiç söz etmemektir. Sonuçlarda bilerek değişiklik yapmak, sonuçları uydurmak, sonuçların bir kısmını çıkartmakla yanıltma ve sahtekârlık yapılması bilimsel açıdan olduğu kadar etik açısından da kabul edilemez.

Bilimsel yöntemlerin uygun olarak kullanılmaması, bilerek ya da bilmeyerek araştırma sonuçlarını etkileyebilecek hataların yapılması ile elde edilen sonuçlar yanıltıcı olabilir. Klinik araştırmaların değerlendirilmesinde iki önemli sorunun yanıtlanması gerekir: 1-Verilere güvenebilir miyim? 2-Elde edilen sonuçları hastalarımda kullanabilir miyim? Klinik araştırmalarda hatadan kaçınmak önemlidir. Yirminci yüzyılda istatistik yöntemlerinin kullanımında görülen artış ve bu yöntemlerin hassasiyeti hataya yol açan

sorunların anlaşılmasını kolaylaştırmıştır. Araştırmalarda istatistiksel yöntemlerin yersiz ya da yanlış kullanımı, sonuçların % 50 oranında farklı çıkmasına neden olabilmektedir. Böyle bir durumda, örneğin araştırmalara göre etkisi iyi bulunan bir ilacın kullanılması ile, beklenen yarar sağlanamayacaktır.

3. *Araştırmanın yayınlanması sırasında görülen etik sorunlar:* Araştırmanın birden çok yerde yayınlanması, tekrar yayınlanması, yayınlanmasının engellenmesi ya da geciktirilmesi ve yazarlık ile ilgili sorunlarla karşılaşabilmektedir (2-3, 11-15).

Bilimsel değerlendirmenin ve etik değerlendirmenin bir bütünün ayrılmaz parçaları olduğu görülmektedir. İnsan deneklerin yer aldığı araştırmaların etik açıdan değerlendirilmesi yapılırken aynı zamanda bilimsel açıdan değerlendirilmesi de yapılmaktadır. Bunun tersi de söylenebilir.

Bilimsel Değerlendirme

Biyomedikal araştırmalarda geçerli ve güvenilir sonuçlar elde edebilmek için istatistik biliminden doğru bir şekilde yararlanılması bilimsel ve etik bir zorunluluktur. Bilimsel değere sahip bir araştırma yüksek kalitede istatistiksel planlama ve veri analizini gerektirir. İstatistik, bilimsel yöntemin temelidir. Biyomedikal araştırmaların istatistiksel olarak uygun biçimde planlanması, yürütülmesi ve değerlendirilmesiyle bilimsel temele oturtulması araştırmaların etik açıdan kabul edilebilmesinin koşullarından birini oluşturur. Biyomedikal araştırmada tıp etiği ilkelerine uyulmaması araştırma sonuçlarını geçersiz kılmamakla birlikte biyoistatistik etiğine uyulmamasının araştırma sonuçlarını geçersiz kılacağı özellikle vurgulanması gereken noktadır (16).

Biyoistatistik ilke ve yöntemlerinin araştırmalarda doğru şekilde kullanılmasının önemini kavrayan araştırmacı: 1- çok yönlü algılama ve çok yönlü düşünme yeteneği geliştirir. Bu görüşü kazanan araştırmacı yanlış davranma hatasına düşmeyecektir. 2- elde edilen bulguların geçerli ve güvenilir olması için tek yolun bilimsel yaklaşım olduğu ve bunun da ancak biyoistatistik ilke ve yöntemlerinin yerinde ve doğru kullanımı ile sağlanabileceğinin farkına varır (2).

Bir araştırmanın üç aşaması vardır; planlama, yürütme ve değerlendirme. Yapılan araştırmanın bilimsel temele oturtulabilmesi için her üç aşama da biyoistatistik ilke ve yöntemleri kullanılarak düzenlenmelidir. Ne yazık ki yaygın olarak yapılan yanlış, araştırmacıların biyoistatistik yöntemlerini araştırmaya başlamadan önce değil, verileri topladıktan sonra kullanma alışkanlığının olmasıdır. Bunun nedenleri ise, araştırmacıların biyoistatistik ilke ve yöntemlerinin bilimsel araştırmadaki önemini yeterince kavrayamamaları ve doğru zamanda kullanmamaları, bilimsel dergi editörlerinin bu konuya gerekli duyarlılığı göstermemeleri, araştırmacıların akademik ünün çekiciliği ile verileri daha ilginç yapma ya da araştırmanın değerini artırma amaçları olarak sayılabilir (17). Bilimsel değer taşımayan, geçerli ve

güvenirliliği kuşkulu olan bu tür arařtırmalardan elde edilen bulgular yarar sağlayamayacağı gibi, kiřilere zarar da verebilir.

Biyomedikal arařtırmanın sonuçlarını etkileyebilecek önemli hataları řu řekilde sıralayabiliriz:

1. Uygun düzenlenmemiř bir arařtırmadan elde edilen bulgular bir tedavinin veya bir ilacın hatalı olarak yararlı olduđu sonucunu (yanlıř pozitif) veya tersini (yanlıř negatif) dođurabilir. Uygun olmayan alıřma düzeninden yanlılıđa açık ve beklenmeyen sonuçların ıkması kaçınılmazdır. Uygun olmayan düzenle yapılan deney sonuçları, daha sonra uygun düzenle yapıldıklarında deđişiklik gösterebilir.
2. Uygun örnekleme yönteminin kullanılmaması ve testin gücü dikkate alınmadan belirlenen küçük örneklem büyüklüđu, hipotezler hakkında yanlıř kararlar verilmesine neden olabilir. Örneklem büyüklüđünün yetersiz olması nedeniyle, tedavi yöntemleri arasındaki önemli farkları yakalama řansı kaybolabilir.
3. Verilerin ölçülmesinde kullanılan ölçü biriminin dođru seilmemesi yanlıř sonuçlara neden olabilmektedir. Örneđin, bir ilacın etkinliđini ölçmek için hangi ölçü birimi kullanılacaktır? Hastanın iyileřiip iyileřiemediđi mi yoksa iyileřieme süresi mi? Hangi ölçü biriminin kullanılacağına arařtırmanın bařında arařtırmacı tarafından karar verilmelidir. Bu ölçülerden hangisi arařtırma sorusunu karřılıyorsa o ölçü alınmalıdır.
4. Kontrol kullanılarak yapılan deney alıřmalarında, seilen deney grubunun kendisi aynı zamanda kontrol grubu olarak da kullanılabilir (eřileştirilmiř grup). Böyle bir yolun tercih edilmesinin gerekli olduđu durumlarda arařtırmacının bilerek kendi beklentilerine cevap vermesinden dolayı eřiştirilmiř grup yerine ayrı bir bađımsız kontrol grubu alıp hipotez testi yapması yanlıř sonuç dođurur.
5. Arařtırmalarda yapılabilecek en büyük hata yan tutmadır. Bir ilacın deney-kontrol gruplarına uygulanmasında hastaların hangi grupta olduklarını hissettirecek istenilen sonucun alınması, anket sorularının hazırlanması sırasında soruların istenilen cevabın alınmasına yönelik hazırlanması, ölçümleri bilerek yanlı olarak ölçmek, hipotez testlerinin seiminde uygun olanın deđil kendisine yarayacak olanın seilmesi gibi bilerek yapılan hatalar buna örnek olarak gösterilebilir (2, 17).

Arařtırma sonuçlarını etkileyen bu hatalar, arařtırmanın deđerinin ya da güvenirliliđinin azalmasına neden olan uygulamalardır ve bilimsel yanılma olarak adlandırılır. Arařtırmacının bilgi ve eđitim eksikliđinden kaynaklanan, farkında olmadan yanlıř sonuç alınmasına neden olan disiplinsiz

araştırma, bilimsel yanılmadan ayrı olarak değerlendirilmelidir. Burada söz konusu olan araştırmacının bilerek verileri uydurması, gerçekleri çarpıtmasıdır (18).

Araştırmalarda biyoistatistik yöntemlerinin bilinçsiz kullanımı, bulguların yanlış yorumlanması, sonuçların yanlış ve çarpıtılarak sunulması gibi bilim etiğine aykırı davranışların neden olacağı sonuçlar, tıp etiğinin yararlı olma, zarar vermeme ve adalet ilkelerinin ihlal edilmesini de beraberinde getirmektedir.

Tartışma ve Sonuç

Biyomedikal araştırmalarda yer alan uygulamalar sayısız istatistiksel ve etik güçlüklerle doludur. Araştırmaların bilimsel değerinin korunmasında araştırma etik kurulları, etik değerlendirmenin gerekli ve önemli bir parçasıdır. Deneklerin katıldığı araştırmalarda etik olarak kabul edilebilir amaçların ve bu amaçlara ulaşmak için uygun araçların seçilmesi gerekmektedir.

Biyomedikal araştırmalarda bilimsel yöntemlerin doğru şekilde uygulanması etik dışı olarak nitelendirilen olaylarda azalmaya neden olacaktır. Bu nedenle doğru uygulanan istatistiksel yöntemler ile araştırmanın en iyi şekilde planlanması ve veri analizinin yapılması etik olarak gereklidir.

Araştırmaların planlanması, yürütülmesi ve değerlendirilmesi aşamalarında karşılaşılabilen sorunlar :

1. Çalışmanın amacını iyi belirlenmemiş olması ya da amacı yöntemle uygunluk göstermemesi,
2. Deneklerin çalışmaya alınma ve alınmama kriterlerinin açık olarak belirtilmemesi,
3. Kontrol grubunun gerçek bir kıyaslama için uygun olmaması,
4. Klinik olarak önemli özelliklerin çalışma ve kontrol grupları arasında eşit olarak dağıtılmaması,
5. Gözlemlerin gerekli nesnellikle yapılmaması,
6. Çalışmadan çıkan ya da çalışma sırasında kayıtları kaybolan deneklere ait sonuçların hiç değerlendirilmemesi ya da araştırmada bundan söz edilmemesi olarak sıralanabilir.

Araştırmanın bilimsel kurallara uygun olarak planlanması, yürütülmesi ve değerlendirilmesi, aynı zamanda araştırmanın etik açıdan da kabul edilebilmesinin koşullarından birini oluşturur. Bu nedenle biyoistatistik uzmanının kasıtlı olmayan ya da yan tutularak yapılan hataları önlemede, ortaya çıkarmada ve araştırmaların sonuçlarının etkilerini ölçmede mutlaka yer almaları gereklidir.

Araştırma etik kurullarının multidisipliner bir yapıya sahip olması, araştırmaların tüm açılardan ve her aşamasında yeterli olarak değerlendirilmesine katkı sağlayacaktır. Her araştırma ekibinde bir biyoistatistik

uzmanının mutlaka bulunması gerektiği gibi araştırma projelerinin etik ve bilimsel açıdan değerlendirilmesi aşamalarında araştırma etik kurullarında da biyoistatistik uzmanının katılımı sağlanmalıdır.

Kaynaklar

1. Atıcı, E.: Etik Kurullar. In: Erdemir AD, Öncel Ö, Aksoy Ş (eds). Çağdaş Tıp Etiği. İstanbul: Nobel Kitab.; 2003: 409-36.
2. Sümbüloğlu V, Sümbüloğlu K. Sağlık Bilimlerinde Araştırma Yöntemleri. Ankara: Hatiboğlu Yayınevi; 1988.
3. Bilimsel Araştırmada Etik ve Sorunları. Ankara: Türkiye Bilimler Akademisi Yayınları; 2002.
4. Gelling L. Role of the Research Ethics Committee. Nurse Education Today 1999; 19(7): 564-569.
5. Haris J. Research Ethics Committees-the future. Notizie di Politeia 2002; 18(67): 123-138.
6. İlaç Araştırmaları Hakkında Yönetmelik. 29.01.1993 Tarih ve 21480 Sayılı Resmi Gazete.
7. World Medical Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects. JAMA. 2003; 284(23): 3043-3045.
8. Aaron D. Research Ethics. West Indian Med J. 1995; 44(4): 115-158.
9. Erdemir AD. The Place of Medical Experiments in Medical Ethics and Its Importance. Lectures on Medical History and Medical Ethics. İstanbul: Nobel Tıp Kitabevleri; 1995.
10. Operational Guidelines for Ethics Committees That Review Biomedical Research. WHO. Cenevre. 2000.
11. Macbeth F, Stephens R. Marketing Clinical Trials. Lancet 1996; 348: 111-112.
12. Chalmers I. Underreporting Research Is Scientific Misconduct. JAMA 1990; 263(10): 1405-1408.
13. Feinstein AR. Tıp Araştırmalarının Biyoistatistik Değerlendirmesinde Sık Yapılan Yanlışlar. In: Yazıcı H (çev.). Tıpta Bilimsel Yazım, Editörlük ve Denetleme. Ankara: TÜBİTAK; 1994: 61-7.
14. Gluud C. Trials and Errors in Clinical Research. Lancet 2000; 354:59.
15. Garattini S, Bertele V, Bassi LL. How Can Research Ethics Committees Protect Patients Better? BMJ 2003; 326(7400): 1119-1121.
16. Thall PF. Ethical Issues in Oncology Biostatistics. Stat Methods Med Res. 2002; 11: 429-448.
17. Marco CA, Larkin GL. Research Ethics:ethical issues of data reporting and the quest for authenticity. Acad Emerg Med 2000; 7(6): 691-694.
18. Buyse M, George SL, Evans S, et al. The Role of Biostatistics in the Prevention, Detection and Treatment of Fraud in Clinical Trials. Stat Med. 1999; 18(24):3435-3451.

OVER KANSERLERİNDE 5 YILLIK SAĞ KALIM ANALİZİ

E. Aslaner¹, M.A. Vardar², R. Burgut¹, A. Altıntaş²

*Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı¹
Çukurova Üniversitesi Tıp Fakültesi Kadın Hastalıkları ve Doğum Anabilim
Dalı²*

Amaç: Bu çalışmada over kanserinde, yaş, evre, histolojik tanı ve çeşitli tıbbi tedavilerin sağ kalım süresi üzerindeki etkilerini araştırmayı amaçladık.

Çalışmanın Yapıldığı Yer: ÇÜTF Balcalı Hastanesi Kadın Doğum Kliniği

Yöntem: Çalışma hastanemizde 1999-2004 yılları arasında yatarak tedavi gören 235 hastada yapılmıştır. Cerrahi evreler I ve IV arasında; % 29'u cerrahi, % 70'i cerrahi+KT ve % 1.3'ü diğer tıbbi tedavileri almıştır. Ca125 değerleri her evrede ölçüldü.

Bulgular: Yüksek Ca125 değerlerine Evre III ve IV'de rastlandı. Evre I'deki hastalar halen yaşamaktadırlar. Farklı cerrahi evreler istatistiksel olarak anlamlı derecede sağ kalım süresini değiştirmektedir.

Sonuç: Bu çalışmada Evre I'deki hastaların Evre III ve IV'e göre daha fazla yaşadıklarını gösterdi. Erken tanı, Over kanserinin tedavisi ve sağ kalım süresi üzerinde tartışılmayacak derecede önemlidir.

Anahtar Kelimeler: Over kanseri, Sağ kalım süresi, Erken tanı.