



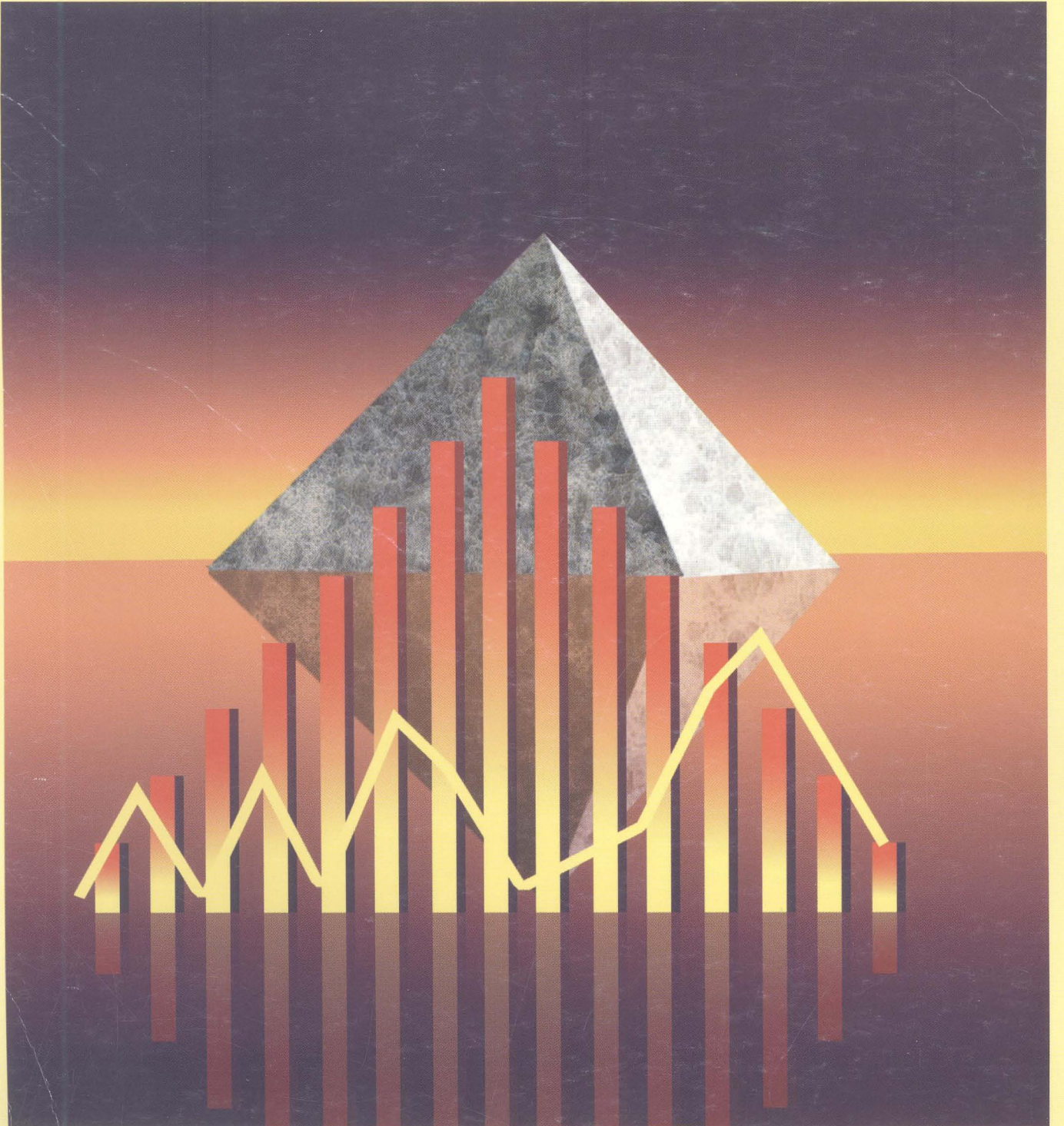
# *İstatistik Araştırma*

## *Journal of Statistical Research Dergisi*

Cilt 03 *Volume*

No 03 *Number*

Aralık 2004 *December*





# İSTATİSTİK ARAŞTIRMA DERGİSİ

**Sahibi**

**Devlet İstatistik Enstitüsü Adına**  
**Ömer DEMİR**  
Devlet İstatistik Enstitüsü Başkanı

**Genel Editör**  
**Fetih YILDIRIM**

## Editörler Kurulu

Alaattin ERKANLI, Duke Univ., USA  
Ali YAZICI, Atılım Üniv., Ankara  
Alptekin ESİN, Gazi Üniv., Ankara  
Ayдын ÖZTÜRK, Ege Üniv., İzmir  
Aykut TOROS, Hacettepe Üniv., Ankara  
Bedriye SARACOGLU, Gazi Üniv., Ankara  
Ceyhan İNAL, Hacettepe Üniv., Ankara  
Ergün KARAAGAOGLU, Hacettepe Üniv., Ankara  
Erkan TÜRE, Marmara Üniv., İstanbul

Fatin SEZGIN, Bilkent Üniv., Ankara  
Fikri AKDENİZ, Çukurova Üniv., Adana  
İmdat KARA, Başkent Üniv., Ankara  
Mithad GÖNEN, Mem. Sloan Kett. Cancer Center,  
USA  
Olca ARSLAN, Çukurova Üniv., Adana  
Refik SOYER, George Washington University, USA  
Soner GÖNEN, Gazi Üniv., Ankara  
Zehra MULUK, Başkent Üniv., Ankara

## Amaç ve Kapsam

İstatistik Araştırma Dergisi, İstatistik araştırmaların kalitesinin artırılmasını, istatistik metodolojisi ve uygulamasının geliştirilmesini, literatürde yer alan çalışmaların tartışılmasını, istatistik uygulamalarıyla ilgili araştırmaların ele alınmasını, teorik ve uygulamalı alanlardaki araştırmacılar arasındaki iletişimin ortak çalışmalar ve yayınlarla güçlendirilmesini amaçlayan bir yayındır.

İstatistik alanında aşağıdaki özellikleri taşıyan çalışmalar Dergi kapsamında değerlendirilir:

1. İstatistik Teorisi, Olasılık Teorisi ve Stokastik Süreçler, Örneklem ve Alan Araştırmaları, Uygulamalı İstatistik, İstatistiksel Kalite Kontrol, Biyoistatistik, Risk Aktüarya Analizi ve Sigortacılık, Ekonometri Yöneylem Araştırması Uygulamaları, Demografi, Bilgisayar Uygulamaları ve Bilgi Sistemleri, gibi istatistiğin her dalında yeni bilgi üretimine yönelik olan tüm araştırmalar.
2. Sosyal Bilimler, Fen Bilimleri, Sağlık Bilimleri ve benzeri alanlara ilişkin veri derleme, veri çözümlemesi ve veri sunumu ile ilgili metodolojilerin geliştirilmesine yönelik araştırmalar.
3. Türkiye’de ve Dünya’da Resmi İstatistiklerin geliştirilmesine yönelik araştırmalar.
4. Yayınlanan İstatistik verileri yeni bilimsel gelişmelerle analiz edip yorumlayan araştırmalar.

## Yayın İlkeleri

1. Bu dergiye alınacak araştırmaların, özgün, yaratıcı, bilimsel kuram ve metodolojiye uygun olmaları; mevcut uygulama ve kurama katkıda bulunmaları esastır. Yayın dili Türkçe’dir.
2. Dergi, istatistiğin alanına giren tüm konuları kapsayan araştırmalara açıktır.
3. Dergide, 3(üç) hakem tarafından incelenip “**Yayınlanabilir**” olurlarını almış araştırmalar yayımlanır. Yayınlanmayan yazılar sahibine geri verilmaz.
4. Makaleler basılı dört kopya ve manyetik ortamda (3.5’lik diskette) dergi sekreteryasına gönderilir. Yayınlanmak üzere kabul edilmiş araştırma son düzeltmesi için yazar(lar)a gönderilir. Bu aşamada metnin değiştirilmesi değil, metne son şeklinin verilmesi beklenir.
5. Bu yayının 5846 Sayılı Fikir ve Sanat Eserleri Kanunu’na göre her hakkı Başbakanlık Devlet İstatistik Enstitüsü Başkanlığı’na aittir, Gerçek veya tüzel kişiler tarafından izinsiz çoğaltılamaz ve dağıtılamaz.
6. Metin hazırlama kalıbına, telif haklarına uymayan herhangi bir yerde yayınlanmış ve yayınlanmak üzere kabul edilmiş çalışmalar genel editör tarafından yazarına iade edilir.
7. Tüm yazışmalar dergi sekreteryası ile yapılır. Abonelik, eski nüshalar, makale ayrı basımları, reklamlar ve ödemelerle ilgili talepler isteme ve abone adresine yapılır.

**ISSN: 1303 – 6319**

**Dergi Sekreteryası**

**Gönül ERDEM - Atalay BİÇYAP**

<b>İstatistik Araştırma Dergisi Sekreterya Adresi</b> Devlet İstatistik Enstitüsü Araştırma-Planlama ve Koordinasyon Dairesi Başkanlığı	Tel: +90 312 410 07 02 – 410 07 32 Faks: +90 312 425 35 85 E-mail: <a href="mailto:dergi@die.gov.tr">dergi@die.gov.tr</a> URL: <a href="http://www.die.gov.tr">http://www.die.gov.tr</a>
<b>İsteme ve Abone Adresi</b> Döner Sermaye İşletmesi	Tel: +90 312 410 03 23 – 410 03 19 Faks: +90 312 417 58 86
Necatibey Cd. No: 114 06100 Yücepete / ANKARA	

# JOURNAL OF STATISTICAL RESEARCH

Owner

On Behalf of The State Institute of Statistics  
Ömer DEMİR

President, The State Institute of Statistics

Editor in Chief  
Fetih YILDIRIM

## Editorial Board

Alaattin ERKANLI, Duke Univ., USA  
Ali YAZICI, Atılım Üniv., Ankara  
Alptekin ESİN, Gazi Üniv., Ankara  
Aydın ÖZTÜRK, Ege Üniv., İzmir  
Aykut TOROS, Hacettepe Üniv., Ankara  
Bedriye SARACOĞLU, Gazi Üniv., Ankara  
Ceyhan INAL, Hacettepe Üniv., Ankara  
Ergün KARAAGAOĞLU, Hacettepe Üniv., Ankara  
Erkan TÜRE, Marmara Üniv., İstanbul

Fatin SEZGIN, Bilkent Üniv., Ankara  
Fikri AKDENİZ, Çukurova Üniv., Adana  
Imdat KARA, Başkent Üniv., Ankara  
Mithad GÖNEN, Mem. Sloan Kett. Cancer Center,  
USA  
Olçay ARSLAN, Çukurova Üniv., Adana  
Refik SOYER, George Washington University, USA  
Soner GÖNEN, Gazi Üniv., Ankara  
Zehra MULUK, Başkent Üniv., Ankara

## Objective and Scope

Journal of Statistical Research is a publication that aims to improve the quality of statistical researches, to develop the statistical methodology and application, to discuss the researches which take place in the literature, to assess the researches on statistical applications, to strengthen the communication between the researchers in theoretical and applied fields by associated studies and publications.

Researches having the following qualities in the field of statistics, are taken into consideration in the scope of the Journal:

1. Researches dealing with the production of new information on statistical matters such as Statistics Theory, Probability Theory and Stochastic Processes, Sampling and Survey, Applied Statistics, Statistical Quality Control, Biostatistics, Risk Actuary Analysis and Insurance, Econometrics, Operational Research, Demography, Computer Applications and Information Systems.
2. Researches dealing with the development of methodologies on data collection, evaluation and presentation in the fields of Social Sciences, Applied Sciences, Medical Sciences, etc.,
3. Researches dealing with the development of Official Statistics of Turkey the world.
4. Researches, dealing with the interpretation and analyses of the statistical data published with new scientific developments.

## Principles of Publication

1. Researches are to be original, creative, fit in methodology and science and contribute to the existing application and theory. Publication language is Turkish.
2. The Journal is open to researches covering all the subjects in the field of statistics.
3. Researches approved by a three referee's mission are published. Unpublished articles are not given back to the author.
4. Articles are sent to the Secretariat of Journal in the forms of print out (4 copies) and magnetic (3,5'' diskette). Researches that are accepted to be published are re-sent to the author(s) for correction. It is expected that at this stage the article is to be given the final form and not to be changed any more.
5. According to the Law No. 5846, SIS holds the copyrights of this publication. The Journal is not duplicated or distributed without authorisation.
6. Researches, which are not in conformity with the form of text preparation, copyrights and previously published or accepted to be published are given back to the author by General Editor.
7. All of the correspondence is to be done with the Secretariat. Requests regarding to the subscription, preceding issues, offprint, advertisements and payments are submitted to the address of Request and Subscription.

ISSN: 1303 - 6319

Journal Secretary

*Gönül ERDEM - Atalay BİÇYAP*

**Journal of Statistics Research Secretary Address**  
State Institute of Statistics  
Department of  
Research-Planing and Coordinating

Tel: +90 312 410 07 02 – 410 07 32  
Faks: +90 312 425 35 85  
e-mail: [dergi@die.gov.tr](mailto:dergi@die.gov.tr)  
URL: <http://www.die.gov.tr>

**Request and Subscription Address**  
State Institute of Statistics  
Revolving Fund Management

Tel: +90 312 410 03 23 – 410 03 19  
Fax: +90 312 417 58 86

Necatibey Street No: 114  
06100 Yücece / ANKARA



## EDİTÖRDEN

Merhaba,

Çok Değerli Okuyucularımız:

Uzun bir süre sonra, nihayet dergimizin Aralık 2004 sayısını sizlere sunmuş bulunuyoruz. Gerçekten de uzun süre demek yanlış değil çünkü hakemlerimizin bir kısmı işlerinin çokluğundan veya mevcut adreslerinden belirli bir sürede de olsa ayrılma durumunda olmalarından dolayı, değerlendirme süreci uzun zaman almaktadır. Elde olmayan nedenlerle oluşan tüm aksamalara rağmen; Ülkemizde, üniversitelerin yayımladığı dergileri dışarıda tutarsak, nedeni veya nedenleri ne olursa olsun, ana içeriği yalnızca istatistik veya uygulamalı istatistik olan ve düzenli yayımlanan tek dergi ne yazık ki bizim dergimizdir. Burada “ne yazık ki” diyorum bu alana emek vermekte olan bir bilim insanı olarak, gönlümüzden geçen çok sayıda ve düzenli çıkan istatistik dergilerinin olmasıdır. Böylece oluşabilecek rekabetten, makalelerin bilimsel düzeyine ilişkin kalite katsayıları da hiç kuşkusuz o denli yüksek olacaktır. Umarım, yakın gelecekte istatistik biliminin Ülkemizde daha da gelişmesine katkıda bulunabilecek ve bu alanda olumlu yönde rekabeti artırabilecek çok sayıda kuramsal ve uygulama ağırlıklı istatistik dergilerinin yayın hayatına kavuşturulur. Bu noktada, tüm istatistik alanında araştırma yapan insanlarımızın bu dergiye sahip çıkmalarının ve yayımlanan makalelerin daha da kaliteli olması için emek vermelerinin bir yurtseverlik görevi olduğunu vurgularsam yanlış yapmamış olurum.

Dergimizin bu sayısında değerlendirilmek üzere- özel sayılar için gönderilenler hariç, gönderilmiş makale sayısı 96 olup; bunlardan yayınlanmaya uygun bulununların sayısı 32, uygun bulunmayanların sayısı 25, çekilen makale sayısı 2, Dergimizin Aralık 2004 sayısında yayımlanacak makale sayısı 8 olup, halen hakem süreci aşamasında veya yazarların hakemlerin eleştirilerine cevap verme sürecinde olan makale sayısı ise 23 adettir. Dergimizin yayımlanan makaleler ile ilgili olarak hertürlü eleştirilerin yer verildiği yazılara açık tartışma bölümünün de olduğunu hatırlatmak isterim. Bu bölümün canlandırılmasında kuşkusuz çok büyük yararlar vardır.

Dergimizin bu sayısının basım aşamasına gelene dek hiç bir karşılık beklemeden ve özverili olarak emekleri geçen tüm hakemlerimize, dergi sekreteryasında görevli sayın Gönül ERDEM ve Atalay BİÇYAP'a ve başta desteklerini hiç eksiltmeden sürdüren Devlet İstatistik Enstitüsü Başkanı Sayın Doç. Dr. Ömer DEMİR'e ve emeği geçen DİE'nin tüm personeline sizler adına teşekkür etmeyi bir borç biliyorum. Diğer sayılarda buluşmak ümidi ile sağlıklı ve bol ürünli günler dilerim.

Saygılarımla,

Prof. Dr. Fetih YILDIRIM  
Dergi Editörü



**2004 YILI NİSAN,ARALIK SAYILARINDA**  
**EMEĐİ GEÇEN HAKEMLER**

**Prof. Dr. Soner GÖNEN – Gazi Üniversitesi**

**Prof. Dr. Hamza GAMGAM – Gazi Üniversitesi**

**Prof. Dr. Reşat KASAP – Gazi Üniversitesi**

**Prof. Dr. Hülya BAYRAK – Gazi Üniversitesi**

**Prof. Dr. Erkan TÜRE – Marmara Üniversitesi**

**Prof. Dr. Şevkinaz GÜMÜŞOĐLU – Dokuz Eylül Üniversitesi**

**Prof. Dr. Fatin SEZGİN – Bilkent Üniversitesi**

**Prof. Dr. Kazım ÖZDAMAR - Osman Gazi Üniversitesi**

**Prof. Dr. İrini DİMİTRİYADİS – Bahçeşehir Üniversitesi**

**Prof. Dr. Hülya ÇINGİ – Hacettepe Üniversitesi**

**Prof.Dr. Hülya BAYRAK – Gazi Üniversitesi**

**Doç.Dr. Salih ÇELEBİOĐLU – Gazi Üniversitesi**

**Dr. Murat Caner TESTİK – Çukurova Üniversitesi**

**Dr. Nezir KÖSE – Gazi Üniversitesi**

**Dr. Özlem TÜRKER – Çankaya Üniversitesi**

**Dr. Özlem İLK – ODTÜ**

**Dr. Nezir KÖSE – Gazi Üniversitesi**



## Andrew Fonksiyon Çizim Tekniği Yaklaşımı ile Çokdeğişkenli Kalite Kontrol Eğrileri ve Bir Uygulama

Yavuz UYAR\*

### ÖZET

*Bu çalışmada, çokdeğişkenli kalite kontrol yöntemlerinden biri incelenmiştir. Ayrıca; kalite kontrol işlemlerinin seri üretim yapan işletmeler için taşıdığı önem de ele alınmıştır. İşletmelerin pazar payını arttırabilmek ve üretim esnasında kaynak israfını önlemek, üretim maliyetini azaltmak sanayi kuruluşlarının temel hedeflerinden biridir. Bunu gerçekleştirmenin bir yolu da üretim sürecini kontrol altında tutmaktır. Çalışmanın uygulama bölümünde bir sanayi kuruluşunda üretilen ve çimento fabrikalarında kalsiyum oksitlerin parçalanması ve termik santrallerde öğütülmesi için kullanılan ve Rotor Çekici olarak adlandırılan ürüne ilişkin veriler kullanılmıştır. Veriler belli aralıklarla ve 5'erli gruplar halinde 50 adet ürüne ilişkin 3 kalite özelliğinden elde edilmiştir. Bu verilere Andrew's fonksiyon çizim tekniği yaklaşımı uygulanmıştır. Bu metoda göre üretim sürecinin kontrol altında olduğu görülmüştür. Eğer üretim süreci kontrol altında olmasaydı, bu durum karar alıcılara iletilerek hatanın sebeplerinin araştırılması gerekecekti. Bu çalışma ile seri üretim yapan işletmelerde yöntemin etkin olarak kullanılabileceği görülmüştür.*

**Anahtar Kelimeler :** Çokdeğişkenli Kalite Kontrol, Üretim Süreci, Kalite

### 1. GİRİŞ

Seri üretime dayalı işletmeler, özellikle sanayi ürünlerini hem iç hem de dış pazarlarda satabilir ve rakipleri ile yarışabilir düzeyde üretebilmek için yoğun çaba harcamak durumundadır (Doganaksoy, vd., 1991). Bu çaba içinde, kaynak israfını önlemek, maliyeti düşürmek ve ürün kalitesini yükseltmek başta gelmektedir. Bu kapsamda, işletme için kalite kontrol tekniklerini kullanmak zorunlu hale gelmiştir. İstatistik kalite kontrol; Bir ürünün en yüksek derecede yararlı aynı zamanda da bir pazara sahip olacak tarzda imalini sağlamak üzere, istatistik prensip ve tekniklerinin, üretimin bütün safhalarında ki uygulamalarından ibarettir (Burnak, 1988). Kalite kontrol işlemi ile kaynak ve dolayısıyla üretim maliyeti düşürülebilir. Seri ve kaliteli üretim için bu gereklidir. Yaygın olarak kullanılan kalite kontrol grafikleri, tek değişken üzerinden çeşitli kontrol grafikleri çizilerek oluşturulmaktadır. Ancak, tüketiciler

\* Dr., Devlet İstatistik Enstitüsü, ANKARA/TÜRKİYE



artık bir ürünün sadece tek bir özellik bakımından kaliteli olmasını yeterli bulmamaktadır (Jacson, 1995).

Çokdeğişkenli kalite kontrol, bir ürün veya sürecin en az iki ya da daha fazla özelliğinin aynı anda birlikte değerlendirilmesidir (Uyar, 1993). Bu çalışma ile çokdeğişkenli kalite kontrol işleminin bir fonksiyon aracılığı ile nasıl elde edileceği gösterilmeye çalışılmıştır.

## 2. METOD

Bir ürünün çeşitli değerlerinden oluşan  $X$  vektörü  $x = (x_1, x_2, \dots, x_p)$  ile tanımlansın.  $X$  vektörü  $p$  boyutlu uzayda bir noktayı ifade etmektedir. Andrew's metodunun temelini  $f(t)$  fonksiyonu oluşturur ve bu fonksiyon,

$$f(t) = \begin{vmatrix} x_1 & x_2 & \dots & x_p \\ t_1 \\ t_2 \\ \cdot \\ \cdot \\ t_p \end{vmatrix} = x_1 t_1 + x_2 t_2 + \dots + x_p t_p, \quad -\pi \leq \theta \leq \pi$$

ile tanımlanır (KULKARNI, S.P., PARANJAPE, S.P., 1984). Bu fonksiyondaki  $t(\theta) = (t_1, t_2, \dots, t_p)$ 'dir. Andrew's fonksiyon çizimi  $f(t)$ 'nin  $-\pi \leq \theta \leq \pi$  üzerindeki çizimidir. Uygulamada  $t$ 'nin seçimi aşağıdaki kümeden yapılabilir.

$$t(\theta) = (\sin \theta \quad \cos \theta \quad \sin 2\theta \quad \cos 2\theta \quad \dots)$$

Bu vektörden oluşturulacak  $t(\theta)$  ile fonksiyon  $[-\pi, \pi]$  aralığında çizilir ve çokdeğişkenli nokta iki boyutta bir eğri olarak elde edilir. Yığın  $\mu$  ve  $\Sigma$  parametrelili bir dağılım olsun.  $p$  değişkenli  $X$  noktası  $X \in E$ 'yi sağlarsa sürecin kontrol altında olduğu kabul edilir.  $E$  kümesi,

$E = \{X : (X - \mu)' \Sigma^{-1} (X - \mu) \leq c_\alpha\}$ 'dir.  $c_\alpha$ ,  $\chi_{p, \alpha}^2$  dağılımının  $100(1 - \alpha)$  yüzde noktasındaki değeridir. Çalışmada  $\alpha = 0,05$  olarak alınmıştır. Eğer  $X \notin E$  ise sürecin kontrol dışına çıktığı belirtilir. Yığına ilişkin  $\mu$  ve  $\Sigma$  parametreleri bilinmiyorsa, örnekten elde edilen  $\bar{x}$  ortalama ve  $S^2$  tahminleri kullanılır ve  $c_\alpha$ 'da  $T_\alpha^2$  ile yer değiştirir ve  $T^2$  değeri,

$$T_\alpha^2 = \frac{p(n-1)}{(n-p)} F_{\alpha; p, N-p}, \text{ eşitliğinden elde edilebilir.}$$

**Önerme:** Süreç sadece ve sadece  $f(t)$ ,  $R$  bölgesine düşüyorsa kontrol altındadır.  $R$  bölgesi:



$$R = \{X: AKE < x_t < ÜKE\}, \quad -\pi \leq \theta \leq \pi$$

$$\text{Üst Kontrol Eğrisi (ÜKE)} = \mu t + (c_\alpha t \Sigma t)^{1/2}$$

$$\text{Alt Kontrol Eğrisi (AKE)} = \mu t - (c_\alpha t \Sigma t)^{1/2}$$

(R, kontrol bölgesi olarak kabul edilecek.)

Andrew's fonksiyon çizim tekniğinin daha kolay anlaşılması için uygulamada aşağıdaki basamaklardan yararlanır.

- I. Basamak:  $\mu t \pm (c_\alpha t \Sigma t)^{1/2}$  ile alt ve üst sınır eğrileri elde edilir. Daha öncede belirtildiği gibi  $\mu$  ve  $\Sigma$  bilinmiyorsa örnekten elde edilen  $\bar{x}$  ortalama ve  $S^2$  değerleri kullanılır. Daha sonra  $-\pi \leq \theta \leq \pi$  aralığındaki bazı  $\theta$ 'lar için çizimi yapılır.
- II.  $f_i(t)$  hesaplanır,  $f_i(t) = \sum_i^n x_t$  hesaplanır. Burada  $i$  örnek çapına eşittir.
- III. Eğer herhangi bir safhada, herhangi bir  $\theta$  için  $f_i(t)$  ÜKE veya AKE'yi kesiyorsa ona karşılık gelen gözlem kontrol dışı olarak kabul edilerek bu durumun nedenleri tespit edilmeye çalışılır.

### 3. UYGULAMA

#### 3.1. Ele Alınan Parça

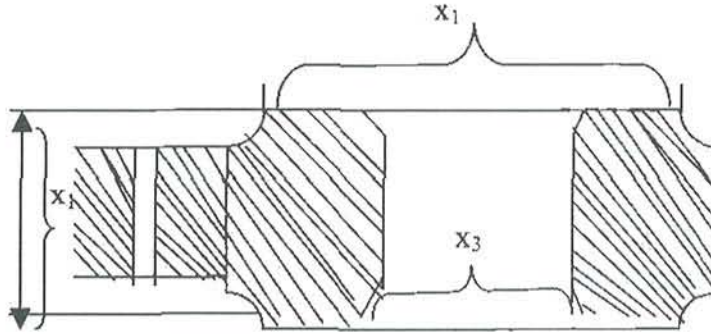
Andrew's metodunun uygulaması için teknik resimde verilen Rotor Çekici'nin üretimi sırasında elde edilen veriler kullanılmıştır. Veriler, ele alınan parçanın üretimi için geçen safhaların tamamını kapsayacak şekilde derlenmiştir. Ürüne ait bir çok özellik önemli sayılmasına rağmen şekilde işaretlenen boyutların daha hassas olduğu vurgulanmış ve buradaki ölçümlerin kalite için temel teşkil edeceği belirtilmiştir.

$x_1$ : Parça kalınlığı (mm)

$x_2$ : Bağlantı deliği (mm)

$x_3$ : Delik havşası (mm)





### 3.2. Veriler

Veriler, üretim sürecinin bütün safhalarını kapsayacak şekilde belli aralıklarla alınarak ölçümleri yapılmıştır. Her bir örnek 5 gruptan oluşmak üzere toplam 10 örnek alınmıştır. Dolayısıyla üç değişken için 150 adet ölçüm yapılarak aşağıdaki tablo elde edilmiştir.

Tablo 1: Ele alınan parçanın ölçüm değerleri (mm)

Kalite Kontrol Değişkenleri			Kalite Kontrol Değişkenleri				
	$X_1$	$X_2$	$X_3$		$X_1$	$X_2$	$X_3$
1. Örnek	37,000	210,000	142,000	6. Örnek	37,000	209,000	142,700
	37,000	210,500	142,000		37,500	210,900	142,000
	37,500	210,500	142,000		37,800	210,000	142,600
	37,500	210,000	142,000		37,000	210,400	142,500
	37,500	210,500	142,000		37,600	210,100	142,000
2. Örnek	37,400	209,500	141,900	7. Örnek	37,300	210,000	142,000
	37,200	210,800	142,500		37,800	210,500	143,800
	37,000	210,000	142,600		37,700	210,100	142,000
	37,600	210,900	142,000		37,000	210,700	142,800
	37,500	210,100	142,700		37,500	210,300	142,700
3. Örnek	37,000	210,000	142,000	8. Örnek	37,300	210,500	141,000
	36,500	211,000	142,000		37,200	210,100	142,000
	37,800	210,600	142,200		37,400	210,900	142,800
	37,500	210,600	142,000		37,600	210,400	143,000
	37,100	210,700	142,800		37,450	210,900	142,600
	37,000	210,100	142,500		37,700	210,500	142,000

	37,300	210,900	142,400		37,400	210,000	142,800
4. Örnek	37,600	210,000	142,100	9. Örnek	37,600	209,000	142,500
	37,000	210,500	143,000		37,200	210,100	142,600
	37,800	210,300	142,600		36,000	211,000	142,700
	37,600	210,000	142,000		37,700	210,500	142,000
	37,700	210,400	142,400		37,400	210,000	142,800
5. Örnek	37,000	210,300	142,100	10. Örnek	37,600	209,000	142,500
	37,900	210,600	142,700		37,200	210,100	142,600
	37,400	210,000	142,100		36,000	211,000	142,700

Bu verilere bağlı olarak  $S^2$  matrisi ve kontrol grafikleri için gerekli olan diğer değişkenlere ilişkin değerler aşağıda verilmiştir.

$$S^2 = \begin{vmatrix} 0,180 & -0,046 & -0,011 \\ -0,046 & 0,250 & 0,018 \\ -0,011 & 0,018 & 0,200 \end{vmatrix}$$

Tablo 2 . t vektörü ve değerleri

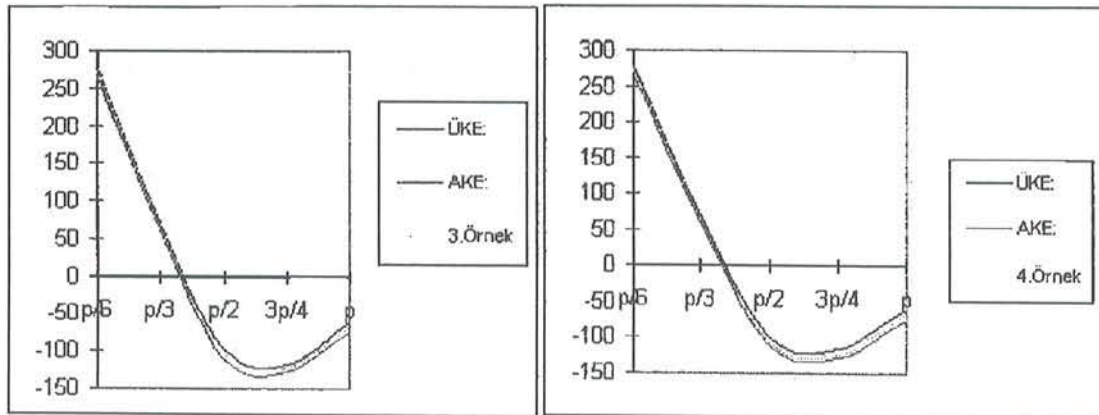
$\theta$	$\sin\theta$	$\cos\theta$	$\sin 2\theta$
$\pi/6$	0,50	0,87	0,50
$\pi/3$	0,87	0,50	-0,50
$\pi/2$	1,00	0,00	-1,00
$3\pi/4$	0,71	-0,71	0,00
$\pi$	0,00	-1,00	1,00

$$\alpha = 0,05$$

$$\bar{x} = \begin{vmatrix} 37,307 & 210,306 & 142,376 \end{vmatrix}$$

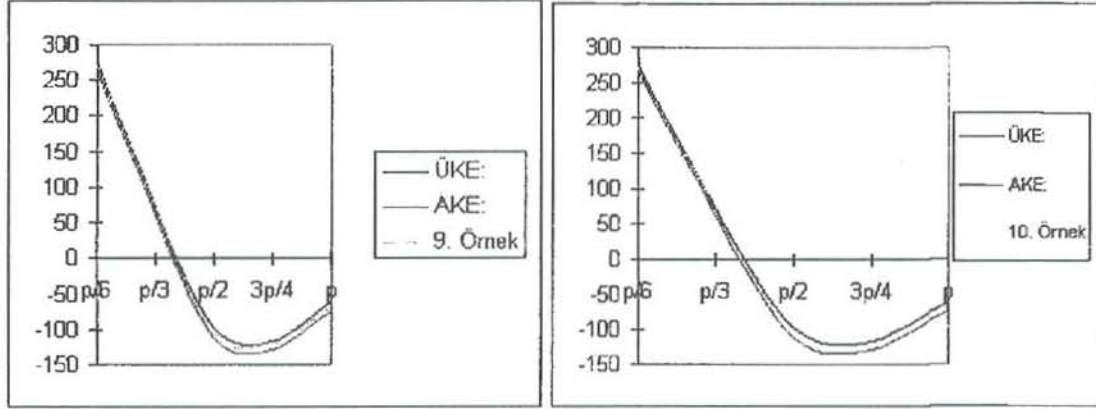
$$T^2 = 114,960$$

Bu değerlere göre, Andrew's fonksiyon çizim tekniğinin yukardaki 3 basamağı sırasıyla hesaplanarak her bir örnek için kontrol grafikleri





oluşturulabilir. Bu çalışmada her bir örnek için grafik çizimi fazla yer tutacağından metodun işleyişini göstermek için kısıtlı sayıda örnek için kontrol grafikleri oluşturulmuştur. Bunlar; 3, 4, 9 ve 10. örneklerdir. Grafiklerde,  $f(t) = x_i t'$  eğrileri oluşturulurken  $x_i$  vektörleri  $i$ . örnek ortalamalarından oluşturulmuştur.



Seçilmiş örnek gruplarına ait yukarıdaki grafikler elde edilmiştir. Grafikler incelendiğinde değerlerin çok küçük ve birbirine yakın olmandan dolayı AKE, ÜKE ve örnek eğrileri üst üste gelmiş gibi görülebilir ancak; örnek grafiklerinde örneklere ilişkin eğriler, alt ve üst kontrol eğrilerini keserek diğer tarafa geçmemiştir. Buradan da ele alınan parçaya ilişkin üretim sürecinin kontrol altında olduğu söylenebilir.

#### 4. SONUÇ

Bu çalışma ile çokdeğişkenli kalite kontrol grafiklerinin Andrew's fonksiyon çizim tekniği yaklaşımı ile oluşturulabileceği gösterilmiştir. Üretim süreçlerinin özellikle seri üretim yapan işletmelerde kontrol altında tutulması çok önemlidir. Üretim maliyeti yüksek ürünlerde kusurlu üretime ilişkin ıskarta maliyeti işletmelere büyük zarar vereceğinden dolayı üretim sürecinin kontrol altında tutulması işletmelerce istenen temel faktördür. Ayrıca; bu çalışma ile kalite kontrol bakımından artık tek değişkenin yeterli olacağı ilave değişkenlerle kaliteyi açıklamanın daha gerçekçi ve yararlı olacağı vurgulanmaya çalışılmıştır. Çalışmada ele alınan parçanın, üretim esnasında, belirlenen kalite kontrol değişkenleri bakımından kontrol altında olduğu söylenebilir. Seri üretim yapan işletmelere çalışmada sunulan bilimsel ve uygulanabilir bu yöntemin kullanılması önerilebilir.

## KAYNAKLAR

- BURNAK, N., 1988, Çokdeđişkenli Kalite Kontrolünde Maliyet Analizi, Anadolu Ü., Eskişehir, 259.
- KULKARNI, S.P., PARANJAPE, S.P., 1984, Use of Andrew's' Function Plot Technique to Construct Control Curve For Multivariate Process, Statist. Meth., 13(20), 2511-2533.
- RICHARD J., HARRIS, A., A Primer of Multivariate Statistic, Academic Press.
- UYAR, Y., Çokdeđişkenli Kalite Kontrolü, Yüksel Lisans Tezi, Gazi Ü., 1993.
- JACSON, J.K., 1995, Multivariate quality control. Comm. Stat—Theory and Methods 14, 2657-2688.
- DOGANAKSOY, N., FALTIN F.B., TUCKER, W.T., 1991, Identification of out of control quality characteristics in a multivariate manufacturing environment. Comm. Stat—Theory and Methods 20(9), 2775-2790.

### **Construct Multivariate Quality Control Curve by the Approach of Andrew's Function Plot Technique and An Application**

#### **ABSTRACT**

*In this study, importance of multivariate quality control in serial production firms is considered. Increasing market share, preventing waste of resource and reducing the cost are main goal of the industrial enterprises. One of the ways for achieving this goal is the inspection of production process under control. In the application part of the study, data for the ROTOR ÇEKİCİ , used in the cement factory for the decomposition of the calcium oxide and in the thermic powerhouse for grinding, are used. Data are obtained from relating to the 3 quality attributes of 50 units product which are got with definite intervals and each group consist of 5. Andrew's function plot technique approach is applied to these data. It is seen that production process is under control using this method. If the process is not under the control, this situation will be informed to the decision makers for the investigation of error reasons. It is shown that method is an efficient technique for the*



*serial production systems. This method should be recommended for the control of a multivariate process.*

**Key Words:** *Multivariate Quality Control, Production Process, Quality*

## Leslie Modeli Kullanılarak Türkiye'deki Kadın Nüfusu İle İlgili Bazı Parametrelerin Tahmini

Gürol ZIRHLIOĞLU\*

### ÖZET

*Leslie matris modeli, yaş sınıfları içerisinde parçalanmış bir nüfusa ait yaşam oranı ve doğurganlık değerlerini kullanarak yaş yapısına ait çeşitli tahminlerde bulunmak için kullanılan bir yöntemdir.*

*Bu çalışmada, Devlet İstatistik Enstitüsü verileri kullanılarak 1955-2000 yılları arasında Türkiye'deki kadın nüfusuna ait bazı parametrelerin tahmini yapılmıştır. Onbeş yıllık yaş gruplarına ayrılan kadın nüfusunu 5 sınıf içinde toplanmış ve 0-14 ile 50+ yaş grupları doğurgan olmayan yaş sınıfı olarak dikkate alınmıştır. Çalışmada, büyüme oranı, üretkenlik değeri, doğurganlık, duyarlılık, esneklik ve t+1 zamanına ait nüfus miktarını tahminlemek amacıyla Leslie matris modeli kullanılmıştır.*

*Anahtar Kelimeler: Leslie matris modeli, nüfus büyüme oranı, üretkenlik değeri, yaşam oranı, doğurganlık.*

### GİRİŞ

Temel nüfus büyüme modellerinden biri olan üssel modelde iki temel demografik parametre kullanılır. Bu parametreler, nüfusun doğal büyümesini belirten yaşam ve doğurganlık parametreleridir. Bu değerler, nüfustaki tüm bireylerin ortalama oranı olarak tanımlanırlar. Populasyon ekolojisi çalışmalarında bir çok tür için önemli özelliklerden biri olan organizmanın yaşı, organizmaların yaşam ve çoğalma şansları üzerinde kuvvetli bir etkiye sahiptir (Akçakaya ve ark., 1999).

Leslie modeli, gelişme (yaşam evresi boyunca gelişme), yaşa özgü doğum ve yaşa özgü ölüm olarak belirtilen 3 çeşit ekolojik işlemi tanımlayan ve birçok biyolojik materyale uygulanabilen bir modeldir. Modelin uygulanmasında genel olarak pek çok populasyonda, nüfus büyümesi söz konusu olduğu zaman sadece dişi bireyler dikkate alınmaktadır. (Sharov, 1996; Akçakaya ve ark., 1999).

Leslie matrisi, yaşa özgü doğurganlık ve yaşam oranlarını içeren bir nüfus tahmin matrisidir. Yaşam ve doğurganlık dönemlerinde bireyler arasındaki farklılıklar gelecekteki nüfusun ne olacağı hakkındaki tahminler için önemli olabilir. Bu farklılıkları hesaplayabilmek için detaylandırılmış modellere ihtiyaç duyulur. Bu modelleri oluşturmanın bir yolu da nüfusu yaş sınırları içerisinde parçalamaktır. Bu model *yaş-yapı modeli* olarak adlandırılmaktadır. Yaş-yapı modellerinin temel

\* Yüzüncü Yıl Üniversitesi, Van Meslek Yüksekokulu, 65080, VAN.



varsayımı nüfusun kapalı olduğudur, yani hiçbir şekilde göç olayı yoktur ve bu modeller nüfusta bulunan bütün bireylerin yaşlarının belirtildiğini ifade ederler (Rickfles, 1993; Krebs, 1994; Akçakaya ve ark., 1999).

Yaşam oranı, t süresinde (x, x+1) yaş grubundaki bir organizmanın, t+1 süresinde (x+1, x+2) yaş grubunda hayatta kalma olasılığı olup, demografik tahminden veya çevresel varyasyondan dolayı yıldan yıla ya da sayımdan sayıma değişim gösterir. Ancak, yaşa özgü doğum ve ölüm oranları uygun bir periyot içerisinde değişmeden kaldığı zaman nüfusun kararlı bir yaş dağılımında olduğu varsayılır. Belirtilen bu şartlar altında nüfustaki her yaş sınıfı ve dolayısıyla nüfusun toplam büyüklüğü aynı oranda artar veya azalır. Kararlı yaş dağılımı ve büyüme oranı doğum ve yaşam değerlerine bağlıdır. Yaşam tablosundaki herhangi bir değişim nüfus büyüme oranı ve kararlı yaş dağılımındaki sonuçların da değişimine neden olacaktır (Ricklefs, 1993; Hayward ve McDonald, 1997).

Bu çalışmada, 1955-2000 yılları arasında Türkiye'deki kadın nüfusu dikkate alınarak bir Leslie modeli oluşturulmuş ve belirtilen nüfusa ait çeşitli parametrelerin tahmini yapılmıştır.

## 2. MATERYAL ve YÖNTEM

Çalışmada kullanılan veriler 1955 – 2000 yılları arasında genel nüfus sayımı sonucunda elde edilmiş olan kadın nüfusu olup, Devlet İstatistik Enstitüsü kaynaklarından elde edilmiştir (DİE, 2001; DİE, 2003).

Kadın nüfusuna ait yaş aralıkları 15 yıl alınarak belirtilen yıllardaki toplam kadın nüfusu 5 yaş sınıfı içinde (0-14, 15-29, 30-44, 45-59, 60+) incelenmiştir. Doğurganlık değerleri hesaplanırken, doğurganlık yaşının 15-49 yaş sınıfları arasında olduğu düşünülerek, doğurgan özellikteki bireylerin toplamı için 15-49 yaş sınıfları arasında bulunan birey sayıları dikkate alınmıştır. Verilerin değerlendirilmesi için RAMAS EcoLab paket programı kullanılmıştır (Akçakaya ve Root, 1998).

### Nüfus Büyüme Oranı

Doğum ve ölüm olayları birey sayısı olarak belirtilmekten ziyade oran olarak belirtilirler. Artan nüfus oranı R sembolü ile gösterilir ve bir nüfusa ait sınırlı büyüme oranı olarak adlandırılır. Eğer  $R > 1$  ise, nüfusta bir artış meydana geldiği,  $R < 1$  ise, nüfusta azalma olduğu sonucu ortaya çıkar. Doğum ve ölüm olayları birbirleri ile dengede ise,  $R = 1$  olarak elde edilecektir.

Herhangi bir zamandan başlayarak gelecek zaman adımlarındaki nüfus büyüklüğünü tahminlemek için,

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

eşitliği kullanılır. Nüfus büyüme oranının belirlenmesi için ise,

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

ifadesi ile kullanılır. Burada  $N(t)$ ,  $t$  süresi içerisinde meydana gelen nüfus büyüklüğü,  $N(0)$  önceki döneme ait nüfus büyüklüğünü ifade etmektedir (Krebs, 1994; Akçakaya ve ark., 1999).

### Yaşam Oranı

Yaşam oranı demografik tahminden ya da çevresel varyasyondan dolayı yıldan yıla değişim gösterir. Leslie matris modelinin oluşturulmasındaki temel unsurlardan biri olan ve  $x+1$  yaşına kadar yaşayan bireylerin sayısı ( $N_1$ ) ile  $x$  yaşındaki bireylerin sayısı ( $N_0$ ) için yaşam oranı,

$$S_x(t) = \frac{N_{x+1}(t+1)}{N_x(t)} \quad (2)$$

şeklinde hesaplanır (Akçakaya ve ark., 1999). Burada,  $S_x(t)$ ,  $t$  zamanında  $x$  yaşındaki bireylerin yaşam oranı,  $N_{x+1}(t+1)$ ,  $t+1$  zamanında  $x+1$  yaşında olan bireylerin sayısı ve  $N_x(t)$  ise,  $t$  zamanında  $x$  yaşında olan bireylerin sayısıdır. Yaş sınıflarına ait yaşam oranları Şekil 1'deki gibi gösterilebilir.

Yaş	Zaman		
	t		t+1
0	$N_0$	$S_0$	$N_1$
1	$N_1$	$S_1$	$N_2$
2	$N_2$		$N_3$
:	:	:	:
8	$N_8$	$S_8$	$N_9$
9	$N_9$		

Şekil 1. Yaş sınıflarına ait yaşam oranları

Şekil 1 incelendiğinde,  $t+1$  zamanındaki  $N_1$  değeri  $t$  zamanında 0 yaş grubundan gelen canlı birey sayısıdır. Şekil 1 dikkate alınarak çalışmada kullanılan zamanlara ait verilerin dağılışı ve yaşam oranı değerlerinin hesaplanmasında bireyler arasındaki ilişki aşağıdaki Şekil 2'de gösterildiği gibidir.



Yaş Grupları	Zaman			
	1955	1970	1985	2000
1.grup (0-14)	$N_0$	$N_0$	$N_0$	$N_0$
2.grup (15-29)	$N_1$	$N_1$	$N_1$	$N_1$
3.grup (30-44)	$N_2$	$N_2$	$N_2$	$N_2$
4.grup (45-59)	$N_3$	$N_3$	$N_3$	$N_3$
5.grup (60+)	$N_4$	$N_4$	$N_4$	$N_4$

Şekil 2. Yıllara göre yaşam oranı değerlerinin hesaplanmasında bireyler arasındaki ilişki

Çalışmada kullanılan veriler  $N$  ile gösterilirse, 1.yaş grubuna ait kadın birey sayısı  $N_0$  ve 5.yaş grubuna ait kadın birey sayısı  $N_5$  olarak belirtilebilir. Buna göre 1. yaş grubuna ait  $S_0$  değerini elde etmek için öncelikle 1970 yılında 15-29 yaş grubunda olan kadın bireylerin ( $N_1(1970)$ ), 1955 yılındaki 0-14 yaş grubunda yer alan kadın bireylere ( $N_0(1955)$ ) oranı, 1985 yılında 15-29 yaş grubundaki bireylerin ( $N_1(1985)$ ), 1970 yılında 0-14 yaş grubundaki bireylere ( $N_0(1970)$ ) oranı ve 2000 yılında 15-29 yaş grubunda bulunan bireylerin ( $N_1(2000)$ ), 1985 yılında 0-14 yaş grubunda bulunan bireylere ( $N_0(1985)$ ) oranı hesaplanmış ve daha sonra bu oranların ortalaması alınmıştır. Aynı işlemler diğer yaş grupları için de yapılarak  $S_1$ ,  $S_2$  ve  $S_3$  değerleri elde edilmiştir.

### Doğurganlık

Herhangi bir  $t$  zamanındaki doğurganlık,  $t+1$  zamanında yaşayan dişi yavruların bir önceki zaman adımında yer alan ergin dişilere oranıdır. Yaş sınıflarına ait doğurganlık Şekil 3'deki gibi gösterilebilir.

Yaş	Zaman	
	$t$	$t+1$
0	$N_0$	$N_0$
1	$N_1$	$N_1$
2	$N_2$	$N_2$
:	:	:
:	:	:
8	$N_8$	$N_8$
9	$N_9$	$N_9$

Şekil 3. Yaş sınıfına ait doğurganlık

Şekil 3'de gösterilen sürekli oklar yaş sınıflarının doğurganlığını, kesikli oklar ise bireylerin  $t$  süresinden  $t+1$  süresine kadar olan yaşam oranlarını gösterir. Yaşa özgü doğurganlığa ait verilen bir kümede  $t+1$  zamanındaki sıfır yaş grubunda yer alan birey sayısı,

$$N_0(t+1) = F_0(t)N_0(t) + F_1(t)N_1(t) + F_2(t)N_2(t) + \dots + F_\omega(t)N_\omega(t) \quad (3)$$

şeklinde tahminlenir (Akçakaya ve ark., 1999).

Yapılan çalışmada 1955 yılında doğurgan olarak kabul edilen yaş gruplarındaki birey sayılarının toplamı ile 1970 yılındaki doğurgan olarak kabul edilmeyen yaş grubundaki (0-14) birey sayısı oranlanarak 1955 yılına ait doğurganlık değeri hesaplanmıştır. Aynı işlemler diğer yıllar için de yapılarak bu yıllara ait doğurganlık değerleri elde edilmiştir.

### Leslie Modeli

Göç olaylarının olmadığı varsayılarak, bir nüfusun içinde bulunduğu zamandan sonraki zaman adımına ait yaş yapısı;

$$N_{x+1}(t+1) = N_x(t)S_x \quad (4)$$

$$N_0(t+1) = \sum_{x=0}^{\omega} F_x(t) N_x(t) \quad (5)$$

eşitlikleri ile belirlenir (Krebs, 1994; Akçakaya ve ark., 1999). Dört numaralı eşitlik, 5 nolu eşitlikte gösterilen üretkenliğe dayanarak gelişme ve mortaliteyi gösterir. Beş nolu eşitlik ilk yaş sınıfındaki bireylerin sayısını belirtir. Dört nolu eşitlikte,  $t+1$  zamanında  $x+1$  yaşında olan bireylerin sayısı, bir önceki yaş aralığında yer alan bireylerin sayısı ile aynı yaş aralığındaki yaşam oranının çarpımına eşittir.

Eğer parametreler (yaşam oranı ve doğurganlık) bir matris formu üzerinde düzenlenirse, bu matrisin,  $t$  zamanındaki yaş dağılımına ait bir vektör ile çarpımı  $t+1$  zamanındaki yaş dağılımını verir (Sharov, 1996).

Yaşam oranı ve doğurganlık için matris formu,

$$\begin{bmatrix} N_0(t+1) \\ N_1(t+1) \\ N_2(t+1) \\ \vdots \\ N_\omega(t+1) \end{bmatrix} = \begin{bmatrix} F_0 & F_1 & F_2 & F_3 \\ S_0 & 0 & 0 & 0 \\ 0 & S_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & S_{\omega-1} & 0 \end{bmatrix} \begin{bmatrix} N_0(t) \\ N_1(t) \\ N_2(t) \\ \vdots \\ N_\omega(t) \end{bmatrix} \quad (6)$$

şeklinde yazılabilir. Eşitliğin sol tarafındaki vektör sonuç vektörü olup,  $t+1$  zaman adımındaki tüm yaş sınıflarında yer alan bireylerin sayısıdır. Eşitliğin sağ tarafında yer alan vektör,  $t$  zaman adımında tüm yaş sınıflarında yaşamını sürdüren dişi bireylerin sayısıdır. Matris ise, yaş sınıflarındaki doğurganlık ve yaşam oranlarına ait Leslie matrisidir. Belirtilen bu matris çarpım işlemi,

$$N(t+1) = L \cdot N(t)$$



matris notasyonu şeklinde gösterilir. Burada  $L$ , Leslie matrisidir.

### Üretkenlik Değeri

Leslie matrisi esas alınarak hesaplanan önemli parametrelerden biri de üretkenlik değeridir. Bu değer, yaşam oranı ve doğurganlık değerlerine bağlı olup, gelecek generasyonlar için herhangi bir yaş sınıfına ait etkinin yaşa özgü ölçümünü ifade eder (Akçakaya ve ark., 1999).

Üretkenlik değerinin hesaplanabilmesi için yaşam oranı ve doğurganlık değerlerinin yer aldığı Leslie matrisine ait öz vektörün hesaplanması gerekmektedir. Leslie matrisinin sağ ve sol olmak üzere 2 öz vektörü vardır. Leslie matrisi ile sağ öz vektörün (kolon vektörü) çarpımı sonucunda yaş sınıflarına ait kararlı yaş dağılımları, sol öz vektörün (sıra vektörü) çarpımı sonucunda ise yaş sınıflarına ait üretkenlik değerleri elde edilir. Ekolojik çalışmalarda öz vektörlerin elde edilmesinde kullanılan öz değer, nüfus büyüme oranı değeridir.

$A$   $k \times k$  boyutlu bir kare matris ve  $\lambda$ ,  $A$  matrisinin öz değeri olarak düşünülün. Eğer  $x$ , sıfır değeri içermeyen ( $x \neq 0$ ) bir vektör olarak kabul edilirse

$$xA = \lambda x$$

eşitliğinde  $x$ ,  $A$  matrisinin  $\lambda$  öz değeri ile birleştirilmiş bir öz vektörü olarak adlandırılır (Johnson ve Wichern, 1988). Burada  $A$  matrisi bir Leslie matrisi olarak dikkate alınırsa,

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 0 & F_1 & F_2 \\ S_0 & 0 & 0 \\ 0 & S_1 & 0 \end{bmatrix} = \lambda \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$$

eşitliği sonucunda elde edilecek olan  $x_1$ ,  $x_2$  ve  $x_3$  değerleri Leslie matrisine ait öz vektör değerleri olup yaşa özgü üretkenlik değerlerini belirtmektedir.

### Duyarlılık ve Esneklik Analizi

Duyarlılık, Leslie matrisinde bulunan yaşam oranı ve doğurganlık değerlerinin herhangi birinin mutlak büyüklüğünde meydana gelen değişimin büyüme oranı ( $\lambda$ ) üzerindeki etkisini ölçer. Örneğin, Leslie matrisinin  $a_{ij}$  elemanı için duyarlılık,

$$s_{ij} = \frac{\partial \lambda}{\partial a_{ij}}$$

eşitliği ile belirtilebilir (Holsinger, 2003). Burada,  $\partial \lambda$  Leslie matrisindeki yaşam oranı veya doğurganlık değerlerinin herhangi birinin mutlak büyüklüğünde meydana gelen değişimden sonraki büyüme değeri ve  $\partial a_{ij}$  Leslie matrisinin  $i$ . sıra  $j$ . sütunundaki değiştirilmiş yaşam oranı veya doğurganlık değeridir. Yapılan çalışmada yaşam oranı ve doğurganlık değerlerinin mutlak büyüklüklerinde 0.1 değerinde değişiklik yapılarak yaş sınıflarına ait duyarlılık değerleri elde edilmiştir.

Esneklik ise, yaşam oranı ve doğurganlık değerlerinde meydana gelen değişimin oransal etkisine ait bir ölçümdür. Esneklik değeri,

$$e_{ij} = \left( \frac{a_{ij}}{\lambda} \right) \left( \frac{\partial \lambda}{\partial a_{ij}} \right)$$

ifadesi ile belirtilir (Holsinger, 2003). Burada,  $a_{ij}$  Leslie matrisindeki i. satır j. sütun elemanı,  $\lambda$  büyüme oranı değeridir.

### 3. BULGULAR

Devlet İstatistik Enstitüsü verilerinden elde edilen 1955-2000 yılları arasındaki kadın nüfusuna ait yaşa özgü yaşam oranları Tablo 1'de belirtildiği gibidir.

**Tablo 1.** Yaşa özgü yaşam oranları

Yaş (x)	Yaşam Oranı ( $S_x$ )			
	1955	1970	1985	Ortalama
0-14	0.946	0.966	0.971	0.961
15-29	0.983	0.966	0.978	0.976
30-44	0.811	0.899	0.983	0.898
45-59	0.852	0.851	0.868	0.857

Tablo 1'de belirtildiği gibi 1955 yılında 0-14 yaş grubunda yer alan bireylere ait yaşam oranı değeri 0.946 olarak elde edilmiştir. Bu değer 1955 (t) yılında bu yaş grubunda bulunan bireylerin 1970 (t+1) yılında bir üst yaş grubunda (15-29) yaşamına devam etme olasılığıdır. Aynı şekilde, 1985 yılında 0-14 yaş grubunda bulunan bireylerin 2000 yılında 15-29 yaş grubundaki yaşam olasılığı 0.971 olarak elde edilmiştir. Belirtilen yıllar için hesaplanmış olan yaşa özgü yaşam oranı değerlerine ait en iyi belirleyici istatistik bu değerlerin aritmetik ortalamasıdır. Elde edilen ortalama değerler Leslie matrisinde kullanılacak olan yaşam oranı değerleridir.

Çalışmada kullanılan verilerin alındığı yıllara ait doğurganlık değerleri Tablo 2'de belirtilmiştir.

**Tablo 2.** Doğurganlık değerleri

Yıl	Doğurganlık (F)
1955	1.302
1970	1.153
1985	0.559

Tablo 2'de belirtilen 1955 yılına ait doğurganlık değeri (1.302) 1970 yılında 0-14 yaş grubunda yer alan bireylerin sayısının 1955 yılında 15-49 yaş gruplarındaki bireylerin toplamına oranıdır. Bu değerler Leslie matrisinde birinci satırı oluşturmaktadır.



Yapılan hesaplamalar sonucunda elde edilen yaşam oranı ve doğurganlık değerlerine göre oluşturulan Leslie matrisi aşağıdaki gibidir.

$$L = \begin{bmatrix} 0 & 1.302 & 1.153 & 0.559 & 0 \\ 0.961 & 0 & 0 & 0 & 0 \\ 0 & 0.976 & 0 & 0 & 0 \\ 0 & 0 & 0.898 & 0 & 0 \\ 0 & 0 & 0 & 0.857 & 0 \end{bmatrix}$$

Yukarıda belirtilen Leslie matrisi ile 2000 yılına ait yaş dağılım vektörünün çarpımı sonucunda bir sonraki döneme (2015 yılı = t+1) ait yaş dağılım tahmini elde edilir. Buna göre, 3 numaralı eşitlik dikkate alınarak yapılan işlem sonucunda t+1 zamanına ait yaş dağılımı,

$$\begin{bmatrix} 0 & 1.302 & 1.153 & 0.599 & 0 \\ 0.961 & 0 & 0 & 0 & 0 \\ 0 & 0.976 & 0 & 0 & 0 \\ 0 & 0 & 0.898 & 0 & 0 \\ 0 & 0 & 0 & 0.857 & 0 \end{bmatrix} \begin{bmatrix} 9767002 \\ 8960514 \\ 6843318 \\ 4061138 \\ 2470398 \end{bmatrix} = \begin{bmatrix} 21827112 \\ 9387041 \\ 8745462 \\ 6145300 \\ 3480395 \end{bmatrix}$$

olarak elde edilir. Burada eşitliğin sağ tarafında yer alan sonuç vektörüne göre, t+1 süresinde 0-14 yaş grubu için elde edilen tahmin değeri 21827112'dir. Bu değer 2000 yılındaki doğurgan nüfustan tahminlenen ve t+1 zamanındaki 0-14 yaş grubunda yer alan doğurgan olmayan birey sayısıdır. Sonuç vektöründe yer alan elemanların toplamı t+1 zamanındaki toplam kadın nüfusunun büyüklüğüne ait tahmini vermektedir.

Çalışmada belirtilmiş olan yıllara ait verilerin dağılımına göre büyüme oranı yaklaşık olarak 1.482 (%48.2) olarak elde edilmiştir. Bu değer 1'den büyük olması kadın nüfusunda artış meydana geldiğini göstermektedir. Yapılan çalışmada veriler 15 yıllık aralıklar ile değerlendirildiği için, hesaplanan nüfus büyüme oranı da 15 yıllık bir büyüme oranıdır. Buna göre yıllık büyüme oranı, bir numaralı eşitlik dikkate alınarak hesaplandığında, 1.024 (%2.4) olarak elde edilir.

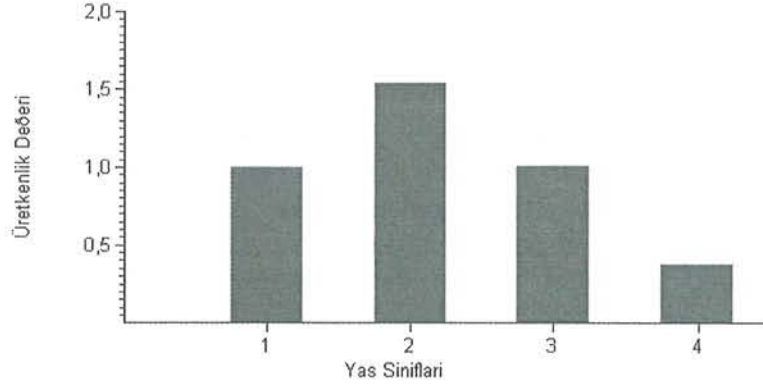
Leslie matrisindeki doğurganlık ve yaşam oranı değerlerine göre elde edilen üretkenlik değerleri Tablo 3'de verilmiştir.

**Tablo 3.** Yaş sınıflarına özgü üretkenlik değerleri

Yaş sınıfları	Üretkenlik Değeri
0-14	1.000
15-29	1.542
30-44	1.006
45-59	0.377
60+	0.000

İlk yaş sınıfındaki bir bireyin üretkenlik değeri göreceli olarak belirtilmiştir. Bu nedenle ilk yaş sınıfı için üretkenlik değeri daima 1.0'dır. Diğer yaş sınıflarındaki

üretkenlik değerleri incelendiğinde gelecek generasyonlar için en fazla üretkenliğin 15-29 yaş grubunda meydana geldiği sonucu elde edilmiştir. Doğurganlık değerleri hesaplanırken 15-49 yaş sınıfları dikkate alındığı için 4.yaş sınıfına ait üretkenlik değeri doğurgan özellikte olan 45-49 yaş sınıfında yer alan birey sayılarına göre hesaplanmıştır. Üretkenlik değerlerinin yaş sınıflarına göre dağılımları Şekil 3'de belirtildiği gibidir.



Şekil 3. Üretkenlik değerinin yaş sınıflarına göre dağılımı

Şekil 3'de görüldüğü gibi en fazla üretkenlik değeri 2. (15-29) yaş grubunda elde edilirken en az üretkenlik değerinin gözlemlendiği grup 4.yaş grubudur.

Sürdürülen analizler sonucunda yaşam oranı ve doğurganlık değerlerindeki değişim sonucunda elde edilen duyarlılık ve esneklik değerleri Tablo 4'de belirtildiği gibidir.

Tablo 4. Yaşa özgü duyarlılık ve esneklik değerleri

Yaş	Duyarlılık					Esneklik				
	0-14	15-29	30-44	45-59	60+	0-14	15-29	30-44	45-59	60+
0-14	0.3956	0.2566	0.1690	0.1024	0.0592	0.0000	0.2255	0.1315	0.0386	0.0000
15-29	0.6100	0.3956	0.2606	0.1579	0.0913	0.3956	0.0000	0.0000	0.0000	0.0000
30-44	0.3983	0.2583	0.1702	0.1031	0.0596	0.0000	0.1702	0.0000	0.0000	0.0000
45-59	0.1492	0.0968	0.0638	0.0386	0.0223	0.0000	0.0000	0.0386	0.0000	0.0000
60+	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Tablo 4'deki en yüksek duyarlılık ve esneklik değerleri dikkate alındığında Türkiye'deki kadın nüfusunun büyüme hızına en fazla etkili olan değer birinci yaş grubuna ait yaşama oranıdır.

#### 4. TARTIŞMA ve SONUÇ

Nüfus yaş yapısına ait tahminlerde bulunmak amacıyla kullanılan Leslie matris modeli ile 1955-2000 yılları arasında Türkiye'deki kadın nüfusunun yaşa özgü dağılımları hakkında çeşitli bilgilere ulaşılmıştır.

Çalışmadan elde edilen tahmin sonuçlarına göre Türkiye'deki kadın nüfusunda yıllık olarak yaklaşık %2.4 oranında bir artış olduğu, yapılan esneklik ve duyarlılık



testleri sonucunda ise, nüfus büyüme hızına en fazla etki eden grubun 0-14 yaş sınıfı olduğu sonucu elde edilmiştir.

Doğurganlık demografik çalışmalarda kullanılan önemli parametrelerden biridir (Ricklefs, 1993; Krebs, 1994). Eğitim düzeyindeki artış, köyden kente göç, ekonomik nedenler ve aile planlaması gibi faktörler bu durumu etkileyen faktörlerin başlıcalarıdır (Karaoğlu ve ark., 2002; TNSA, 1993). Yapılan çalışma ile elde edilen doğurganlık değerlerine göre en fazla doğurganlığın 1955 yılında olduğu, bu değer diğer yıllarda azaldığı sonucu elde edilmiştir. Buna göre, belirtilen yıllarda 0-14 yaş sınıfında bulunan bireylerin artışında bir düşüş olduğu anlaşılmaktadır.

Leslie matris modelinde sadece yaşam oranları ve doğurganlık değerleri kullanılarak sonraki döneme ait nüfus büyüklüğü ve bu büyüklüğün yaş sınıflarına göre dağılımları, yaş sınıflarının gelecek generasyonlar üzerindeki etkileri, esneklik ve duyarlılık değerleri tahmin edilmiştir (Akçakaya ve ark.,1999). Buna göre 1955-2000 yılları arasındaki kadın nüfusu verilerinden yararlanılarak 2015 yılına ait tahminler yapılmıştır.

Yaşa özgü doğurganlık oranları, genç yaşlara yığılımlı bir dağılım göstermektedir. İleri yaşlarda ise, doğurganlık değerinde belirgin ölçülerde azalma meydana gelmektedir (Karaoğlu ve ark.,2002; TNSA, 1993). Yaşam oranı ve doğurganlık değerlerinin hesaplanmasından sonra, Leslie matris modeli ile yapılan değerlendirmeler sonucunda, gelecek generasyonlar için herhangi bir yaş sınıfına ait etkinin yaşa özgü ölçümünü ifade eden üretkenlik değerlerinin en fazla 15-29 yaş sınıfında olduğu sonucu elde edilmiştir. Daha ileri yaş sınıflarında ise, üretkenlik değerlerinin azaldığı görülmüştür.

#### KAYNAKLAR

- Akçakaya, H.R.. and Root, W. (1998), *RAMAS EcoLab, Applied Biomathematics*, 100 North Country Road, Setauket, NY 11733 USA.
- Akçakaya, H.R., Burgman, M.A. and Ginzburg, L.R. (1999), *Applied Population Ecology*, Sunderland, Massachusetts: Sinauer Associated, Inc.
- DİE (2001), *İstatistik Göstergeler,1923-1998*, T.C. Başbakanlık Devlet İstatistik Enstitüsü, Ankara, TÜRKİYE.
- DİE (2003), *Türkiye'de Kadın Bilgi Ağı*,Erişim: [[http://www.die.gov.tr/tkba/tkba\\_tr.htm](http://www.die.gov.tr/tkba/tkba_tr.htm)]. Erişim Tarihi: 12.11.2003.
- Hayward, G.D. and McDonald, D.B. (1997), *Matrix Population Models as a Tool in Development of Habitat Models*, 2nd Owl Symposium, 205-212.
- Holsinger, K. (2003), *Sensitivity and Elasticity Analyses*, Erişim: [<http://darwin.eeb.uconn.edu/ceb310/lecture-notes/pva/node8.html>].Erişim Tarihi: 23.10.2003.
- Johnson, R.A. and Wichern, D. W. (1988), *Applied Multivariate Statistical Analysis*, New Jersey, Prentice-Hall Inc.

Karaoğlu, L., Öztürk, C., Pehlivan, E. (2002), *Çorum İli Bir Sağlık Ocağı Bölgesinde Yaşayan Evli Kadınlarda Doğurganlık ve Etkileyen Faktörler*, İnönü Üniversitesi Tıp Fakültesi Dergisi, 9(1) 33-40.

Krebs, C. J. (1994), *Ecology*, New York: HarperCollins Collage Publishers.

Ricklefs, R.E. (1993), *The Economy of Nature: A Textbook in Basic Ecology*, England: W.H.Freeman and Company.

Sharov, A. (1996), *Quantitative Population Ecology*. Erişim: [<http://www.ento.vt.edu/~sharov/alexei.html>]. Erişim Tarihi: 12.05.2003.

TNSA (1993), *Türkiye Nüfus ve Sağlık Araştırması*, Manisa Sağlık Müdürlüğü. Erişim: [<http://www.manisa.saglik.gov.tr/tnsa1.htm>]. Erişim Tarihi: 05.02.2004.

## **Estimation of Some Parameters for Woman Population in Turkey: A Leslie Model Approach**

### **ABSTRACT**

*The model of Leslie matrix is a method used for making various estimations that belong to age structure by using survival rate that belongs to a population which is broken up in the classes of age and by using the rate of productiveness.*

*In this study, by using State Institute of Statistics data estimation of some parameters, which belongs to woman population in Turkey between 1955-2000, was made. Woman population which is choosen for fifteen years old groups gathered in 5 classes and the groups of 0-14 and 50+ age considered to be the classes of age which are no fecund. In study, the model of Leslie matrix used in order to estimation of growing rate, reproductive value, fecundity, sensitivity, elasticity and quantity of population at  $t+1$  time.*

**Key Word:** *Leslie matrix model, population growth rate, reproductive value, survival rate, fecundity.*



## Kümelenmiş Verilerde Bağımlı İki Oranın Karşılaştırılması

Derya GÖKMEN*	Yasemin GENÇ*	Yıldır ATAKURT*	Banu YAĞMURLU**
---------------	---------------	-----------------	-----------------

### ÖZET

*Bir serbestlik dereceli ki-kare dağılımına sahip olan McNemar testi, bağımlı örneklemelerden tahmin edilen iki oranın karşılaştırılmasında kullanılır. Bu testin temel varsayımı, her bir gözlem çiftinin bağımsız olmasıdır. Fakat, bazı çalışmalarda bir bireyden birden fazla ölçüm alınır ve "kümelenmiş veri" olarak adlandırılan bu verilerde bağımsızlık varsayımı sağlanmaz. Yapılan benzetim çalışmaları, bu varsayımın bozulmasının I.tip hata oranında artışa neden olduğunu göstermiştir. Kümelenmiş verilere genellikle radyoloji, diş ve göz hastalıkları alanlarında yapılan çalışmalarda rastlanır. Çalışmamızda, kümelenmiş verilerde bağımlı iki oranın karşılaştırılması için Eliasziw ve Donner, Obuchowski ve Durkalski ile arkadaşları tarafından önerilen yöntemler tanıtılmıştır. Bu yöntemler, radyoloji alanında yapılan bir çalışmanın verileri üzerinde uygulanmış; sonuçlar standart McNemar testinden elde edilenle karşılaştırılmıştır.*

*Anahtar Kelimeler: Kümelenmiş veri, McNemar testi*

### 1. GİRİŞ

McNemar testi, bağımlı gözlemlerden tahmin edilen iki oranın karşılaştırılmasında kullanılır. Bu testin temel varsayımı, her bir gözlem çiftinin bağımsız olmasıdır. Fakat, bazı çalışmalarda bir bireyden birden fazla ölçüm alınır ve "kümelenmiş veri" olarak adlandırılan bu verilerde bağımsızlık varsayımı sağlanmaz. Kümelenmiş verilere genellikle diş ve göz hastalıkları alanlarında yapılan çalışmalarda rastlanır.

Örneğin, katarakt başlangıcı olan hastalara uygulanan bir tedavinin, görme keskinliği (GK) üzerine etkisini incelemek amacıyla bir çalışma düzenlendiğini varsayalım. Araştırmacı, 100 hastanın sol ve sağ gözlerinin GK değerlerini; tedaviden önce ve tedaviden sonra ölçsün. Bu örnekte,  $Y_{1i}$  ve  $Y_{2i}$  bir gözün, tedaviden önceki ve tedaviden sonraki durumunu göstermek üzere, GK, 20/50 ya da daha kötü (bozulmuş) olduğunda 1; 20/40 ya da daha iyi (normal) olduğunda 0 değerini alsın. Bir bireyin gözleri karakteristik olarak benzer olduğu için, 200 çiftin ( $Y_{1i}$ ,  $Y_{2i}$ ) örneklem sonuçları bağımlı olur. (Eliasziw ve Donner, 1991).

\*Ankara Üniversitesi, Tıp Fakültesi, Biyoistatistik A.B.D. Sıhhiye/ANKARA

\*\* Ankara Üniversitesi Tıp Fakültesi, Radyodiagnostik A.B.D. Cebeci/ANKARA



Çalışmamızda, kümelenmiş verilerde bağımlı iki oranın karşılaştırılması için Eliasziw ve Donner, Obuchowski ve Durkalski ile arkadaşları tarafından önerilen yöntemler tanıtılacak ve elde edilen sonuçlar bir örnek üzerinde karşılaştırılacaktır.

Kümelenmiş verilerde bağımlı iki oranın karşılaştırılması için, “ilişkili binom modeli”, “Fisher’in permütasyon testi”, “ağırlıklandırılmış binom kestiricileri” ve “genelleştirilmiş tahmini denklemleri kullanan marjinal regresyon modelleri” de kullanılan diğer yöntemler arasındadır. Bununla beraber, yukarıda sözü geçen üç yöntem uygulama kolaylıkları nedeniyle, daha kompleks olan bu yöntemlere tercih edilmektedir.

## 2. YÖNTEMLER

### 2.1. Standart McNemar Testi

Standart McNemar testi, her bir gözlem çiftinin bağımsız olduğu varsayımına dayanır. Bu gözlem çifti, aynı özelliklere sahip iki denekten birer ölçüm ya da aynı denekten iki ölçüm alınarak oluşturulabilir. Bu testte amaç,  $p_1 = p_2$  eşitliğinin araştırılması olup; bu değerler testlerin her biri için pozitif sonuç elde etme olasılığını gösterir. Uyumlu (concordant) eşler (a,d) testler arasındaki farklılığa ilişkin çok az bilgi verdiğinden, test istatistiği sadece uyumsuz (discordant) eşlerden (b,c) elde edilen bilgiyi kullanarak; başarı olasılıkları ya da uyumsuz çift sayıları cinsinden aşağıdaki gibi bulunabilir:

$$\chi_{MC}^2 = \frac{(\hat{p}_1 - \hat{p}_2)^2}{\text{var}(\hat{p}_1 - \hat{p}_2)} = \frac{(b-c)^2}{b+c} \quad (1)$$

Standart McNemar testi, asimptotik olarak bir serbestlik derecesi ile ki-kare dağılır.

### 2.2. Kümelenmiş Verilerde McNemar Testi

#### 2.2.1. Eliasziw ve Donner’ın uyumsuz çiftlerden elde edilen ilişki katsayısına dayalı McNemar testi

Snedecor ve Cochran (1980) tarafından önerilen küme içi korelasyon katsayısı ( $\rho$ ), beta-binom dağılımı ile modellenir. Bu model altında, bağımlı örneklem her biri için pozitif sonuca sahip olma olasılığı ( $p_1$  veya  $p_2$ ) binom dağılırken; bu olasılıklar, kümeler arasında beta dağılımına göre değişkenlik gösterir. Küme içi korelasyon katsayısı, genellikle gözlemciler arası uyumun araştırılmasında kullanılıp; Cohen’in kappa istatistiğine alternatif olarak verilir.

Donald ve Donner (1991), kümeleme nedeniyle olduğundan daha küçük kestirilen varyansı düzeltmek için, küme içi korelasyon katsayısının tutarlı bir tahminini ( $\hat{\rho}$ ) kullanarak  $1 + (n_c - 1)\hat{\rho}$  eşitliği ile verilen bir şişirme faktörü elde etmişlerdir. Eliasziw ve Donner (1991), kümelenmiş veriler için düzeltilmiş McNemar test istatistiğini, standart McNemar test istatistiğinin, bu şişirme faktörüne bölünmesi

biçiminde tanımlanmışlardır. Düzeltilmiş McNemar test istatistiği, bir serbestlik dereceli ki-kare dağılımına sahip olup;  $\rho=0$  olduğunda standart McNemar test istatistiğine dönüşür.

$H_0: p_1=p_2$ 'nin test edilmesi için, Eliasziw ve Donner tarafından önerilen düzeltilmiş McNemar test istatistiği Tablo 1'deki gösterimler kullanılarak aşağıdaki eşitlikler yardımıyla elde edilir:

**Tablo 1.** k. küme için uygun veri planı

	II. Test (T <sub>2</sub> )		
I. Test (T <sub>1</sub> )	Sonuç (+)	Sonuç (-)	Toplam
Sonuç (+)	$a_k$	$b_k$	$a_k+b_k$
Sonuç (-)	$c_k$	$d_k$	$c_k+d_k$
Toplam	$a_k+c_k$	$b_k+d_k$	$n_k$

Yukarıdaki tabloya göre,  $p_1=(a_k+b_k)/n_k$  ve  $p_2=(a_k+c_k)/n_k$  olarak verilir. k. ( $k=1, 2, \dots, K$ ) küme içindeki b tipi uyumsuz çiftlerin oranı  $p_k = b_k / S_k$  olup;  $S_k$  ( $S_k=b_k+c_k$ ) uyumsuz çiftlerin toplam sayısıdır.  $K_d$ , en azından bir uyumsuz çifte sahip küme sayısını gösterdiğinde (yani  $S_k \geq 1$ ),

$$\bar{p} = \frac{\sum_{k=1}^K b_k}{\sum_{k=1}^K S_k} \text{ ve } \bar{S} = \frac{1}{K_d} \sum_{k=1}^K S_k \text{ olmak üzere, } S_0 = \bar{S} - \frac{\sum_{k=1}^K (S_k - \bar{S})^2 - (K - K_d)\bar{S}^2}{K_d(K_d - 1)\bar{S}} \text{ 'dır.}$$

$HKO_{(\bar{O}A)}$ , ölçümler arası hata kareler ortalaması ve  $HKO_{(KI)}$  küme içi hata kareler ortalaması olmak üzere, tahmin edilen küme içi korelasyon katsayısı, küme içindeki ölçümler arası ilişkinin derecesini belirler:

$$\hat{\rho} = \frac{HKO_{(\bar{O}A)} - HKO_{(KI)}}{HKO_{(\bar{O}A)} + (S_0 - 1)HKO_{(KI)}} \quad (2)$$

$$HKO_{(\bar{O}A)} = \frac{1}{K_d} \sum_{k=1}^K \left\{ \begin{array}{l} \frac{(b_k - S_k \bar{p})^2}{S_k} \quad S_k \geq 1 \text{ ise} \\ 0 \quad S_k = 0 \text{ ise} \end{array} \right.$$

$$HKO_{(KI)} = \frac{1}{K_d(\bar{S} - 1)} \sum_{k=1}^K \left\{ \begin{array}{l} \frac{b_k(S_k - b_k)}{S_k} \quad S_k \geq 1 \text{ ise} \\ 0 \quad S_k = 0 \text{ ise} \end{array} \right.$$

$n_c$ ,  $n_c = S_0 + K_d(\bar{S} - S_0)$  uyumsuz eş sayısının düzeltilmiş ortalamasını göstermek üzere, düzeltilmiş McNemar test istatistiği aşağıdaki eşitlik yardımıyla elde edilir:

$$\chi_{ED}^2 = \frac{\chi_{MC}^2}{1+(n_c-1)\hat{\rho}} \quad (3)$$

### 2.2.2. Eliasziw ve Donner'ın tüm çiftlerden elde edilen ilişki katsayısına dayalı McNemar testi

Bölüm 2.2.1.'de bahsedilen  $\bar{p}$  0.5'den çok düşükse ve/veya  $S_k$  küçükse ve/veya  $\bar{S}$ , 1 civarında ise,  $\hat{\rho}$  değerine güvenilmez.  $\bar{p}=0$  ya da 1 olması durumunda ise,  $\hat{\rho}$  hesaplanamaz. Bu durumda, tüm çiftleri analize katan aşağıdaki  $\rho$  kestiricisinin kullanılması önerilmiştir:

$$\bar{n} = \frac{1}{K} \sum_{k=1}^K n_k \text{ olmak üzere, } n_0 = \bar{n} - \frac{\sum_{k=1}^K (n_k - \bar{n})^2}{K(K-1)\bar{n}} \text{ 'dır.}$$

$$\hat{P}_{11} = \sum_{k=1}^K a_k / \sum_{k=1}^K n_k, \quad \hat{P}_{10} = \sum_{k=1}^K b_k / \sum_{k=1}^K n_k, \quad \hat{P}_{01} = \sum_{k=1}^K c_k / \sum_{k=1}^K n_k, \quad \hat{P}_{00} = \sum_{k=1}^K d_k / \sum_{k=1}^K n_k,$$

$$TOPLAM \ HKO_{(\bar{O}A)} = \frac{1}{K} \sum_{k=1}^K \left[ \frac{(a_k - n_k \hat{P}_{11})^2 + (b_k - n_k \hat{P}_{10})^2 + (c_k - n_k \hat{P}_{01})^2 + (d_k - n_k \hat{P}_{00})^2}{n_k} \right]$$

$$TOPLAM \ HKO_{(Ki)} = \frac{1}{K(\bar{n}-1)} \sum_{k=1}^K \left[ \frac{a_k(n_k - a_k) + b_k(n_k - b_k) + c_k(n_k - c_k) + d_k(n_k - d_k)}{n_k} \right]$$

$$\tilde{\rho}^* = \frac{TOPLAM \ HKO_{(\bar{O}A)} - TOPLAM \ HKO_{(Ki)}}{TOPLAM \ HKO_{(\bar{O}A)} + (n_0 - 1)TOPLAM \ HKO_{(Ki)}} \text{ olmak üzere:}$$

$$\tilde{\rho} = \frac{1}{1 + \hat{P}_{10}(1 - \tilde{\rho}^*) / \tilde{\rho}^* + \hat{P}_{01}(1 - \tilde{\rho}^*) / \tilde{\rho}^*} \quad (4)$$

dır.  $\tilde{\rho}$ , hem uyumlu hem de uyumsuz çiftlerin bilgisini kullandığı için, genellikle  $\hat{\rho}$ 'dan daha etkilidir. Eşitlik 3'teki  $\hat{\rho}$ 'nın  $\tilde{\rho}$  ile yer değiştirmesi sonucu elde edilen düzeltilmiş McNemar test istatistiği  $\tilde{\chi}_{MD}^2$  ile gösterilir:

$$\tilde{\chi}_{ED}^2 = \frac{\chi_{MC}^2}{1+(n_c-1)\tilde{\rho}} \quad (5)$$

### 2.2.3. Obuchowski'nin örnekleme yaklaşımına dayalı McNemar testi

Obuchowski (1998) tarafından önerilen bu yöntem, ölçümlerin kümelerden örnekleme ile seçildiği varsayımı altında kullanılır. Önerilen yöntemde, k. kümesinde ( $k=1, 2, \dots, K$ )  $n_k$  ölçümün bulunduğu bir kitleden K kümelik bir örneklemin seçildiği



varsayılır. Her bir ölçüm birimine I test ya da tedavi uygulanır (Çalışmamız, sadece iki test ya da tedavinin olduğu durumla sınırlandırılacaktır).

$x_{ik}$ , k. küme için i. testte (i=1,2) elde edilen pozitif (veya negatif) sonuca sahip ölçüm sayısını,  $p_i$ , rasgele seçilen bir ölçüm biriminin, i. testte pozitif sonuç verme olasılığını gösterebilir. Rao ve Scott (1992),  $p_i$ 'nin bir tahmini olarak  $\hat{p}_i$ 'yi;

$$\hat{p}_i = \frac{\sum_{k=1}^K x_{ik}}{\sum_{k=1}^K n_k} \quad (6)$$

olarak vermişlerdir. Küme sayısı büyük olduğunda,  $\hat{p}_i$ 'nin küme içi korelasyonu dikkate alan varyans tahmini aşağıdaki gibidir:

$$\text{var}(\hat{p}_i) = K(K-1)^{-1} \left[ \sum_{k=1}^K (x_{ik} - n_k \hat{p}_i)^2 \right] / \left( \sum_{k=1}^K n_k \right)^2 \quad (7)$$

K kümelik bağımlı örneklemelerden tahmin edilen oranları karşılaştırmak için, küme içi korelasyonun yanı sıra, oranlar arasındaki ilişkinin de incelenmesi gerekmektedir.  $\hat{p}_1$  ve  $\hat{p}_2$  arasındaki kovaryansın bir tahmini aşağıdaki gibi önerilmiştir:

$$\text{côv}(\hat{p}_1, \hat{p}_2) = K(K-1)^{-1} \left[ \sum_{k=1}^K (x_{1k} - n_k \bar{p})(x_{2k} - n_k \bar{p}) \right] / \left( \sum_{k=1}^K n_k \right)^2 \quad (8)$$

Burada,  $\bar{p}$ ; her bir test için pozitif sonuç oranlarının ortalaması olup;  $\bar{p} = (\hat{p}_1 + \hat{p}_2) / 2$  eşitliği ile verilir. Bu durumda,  $\hat{p}_1$  ve  $\hat{p}_2$  arasındaki farklılığa ait varyans tahmini aşağıdaki gibidir:

$$\text{vâr}(\hat{p}_1 - \hat{p}_2) = \text{vâr}(\hat{p}_1) + \text{vâr}(\hat{p}_2) - \text{côv}(\hat{p}_1, \hat{p}_2) \quad (9)$$

$H_0: p_1 = p_2$  biçimindeki yokluk hipotezini test etmek için uygun test istatistiği asimptotik olarak bir serbestlik dereceli ki-kare dağılımına sahiptir:

$$\chi^2_O = (\hat{p}_1 - \hat{p}_2)^2 / \text{Vâr}(\hat{p}_1 - \hat{p}_2) \quad (10)$$

Her bir kümede sadece bir ölçüm olduğunda, Eşitlik (10)'da verilen test istatistiği, sonlu düzeltme faktörü,  $K/(K-1)$  ile çarpıldığında standart McNemar test istatistiğine dönüşür.

#### 2.2.4. Durkalski ve arkadaşlarının momentler yöntemine dayalı McNemar testi

Rasgele bir değişkenin dağılımı bilindiğinde, varyans rasgele değişkenin fonksiyonu kullanılarak tahmin edilebilir, fakat, dağılıma bağlı olarak hesaplamalar zor olabilir. Bunun yanında, kümelenmiş verilerde uyumsuz çiftler arası bilinmeyen ilişki

yapısından dolayı, K küme ve küme içindeki  $n_k$  ölçüm üzerinden toplanan uyumsuz çiftler arasındaki farklılığın gerçek dağılımını bulmak zor olabilir. Alternatif bir yaklaşım, varyansın momentler kestirimini yapan ve dağılımsal varsayımlara gerek duymayan sağlam (robust) bir metod kullanmaktır (Royall, 1986).

Durkalski ve arkadaşları (2003), Royall'ın iki aşamalı küme örneklemeinde varyansın tahmini için geliştirdiği momentler kestiricisi yöntemini bağımlı iki oran arasındaki farka ait varyansın tahmininde kullanmışlardır. Momentler kestiricisi yöntemi, örneklem büyüklüğü arttıkça, daha tutarlı varyans tahminleri verir. Bu yaklaşıma göre, oranlar arasındaki farka ait varyansın tutarlı bir tahmini ve önerilen test istatistiği aşağıdaki gibidir:

$$Var(\hat{p}_{10} - \hat{p}_{01}) = \frac{1}{K^2} \sum_{k=1}^K \left[ \frac{(b_k - c_k)}{n_k} \right]^2 \quad (11)$$

$$\chi_D^2 = \frac{\left[ \sum_{k=1}^K \frac{1}{n_k} (b_k - c_k) \right]^2}{\sum_{k=1}^K \left[ \frac{(b_k - c_k)}{n_k} \right]^2} \quad (12)$$

Küme sayısı arttıkça, farklılığa ilişkin tahmin yaklaşık olarak normal dağılım gösterir. Oranlar arası farklılığın, kendi tahmini standart hatasına oranının karesi, Wald istatistiğidir. Bu yüzden, test istatistiğinin, asimptotik olarak bir serbestlik derecesi ile ki-kare dağılımına sahip olduğu varsayılabilir. Ayrıca, her bir kümede sadece bir ölçüm olduğunda, önerilen test istatistiği, standart McNemar testine dönüşür.

### 3. UYGULAMA

Ankara Üniversitesi Tıp Fakültesi Radyodiagnostik Anabilim Dalında yapılan bir çalışmada, bel ağrısı şikayeti olan 60 hastanın 5 lumbosakral omurgası, disk patolojisini test etmeye yönelik olarak farklı parametre ve sekanslardan oluşan opposed phase (OP) ve konvansiyonel (K) MR testleri ile incelenmiş ve aşağıdaki sonuçlar elde edilmiştir. “Periferik disk taşması”, “disk protrüzyonu” veya “disk hernisi” olan bireyler “hasta” grubuna, bu bulgulardan hiçbirine rastlanmayanlar ise “sağlıklı” grubuna dahil edilmiştir.

**Tablo 2.** Verilerin dağılımı

		K		Toplam
		Hasta	Sağlıklı	
OP	Hasta	129	28	157
	Sağlıklı	8	135	143
Toplam		137	163	300



Çalışmamızda, her bir hastanın 5 lumbosakral omurgası incelendiğinden küme büyüklükleri değişkenlik göstermemektedir. Her bir test için, her bir kümede, hasta tanısı koyma olasılıkları 0 ile 1 aralığında değişmektedir.

Eliasziw ve Donner'ın sadece uyumsuz çiftlerden elde edilen ilişki katsayısına dayalı olarak önerdikleri McNemar testinde kullanılmak üzere, toplam uyumsuz çift sayısı 38 (%12), en azından bir uyumsuz çifte sahip küme sayısı 28, en az bir uyumsuz çifte sahip küme başına düşen ortalama uyumsuz çift sayısı 1.29, tüm kümeler üzerinden b tipi uyumsuz çift oranı 0.78 olarak bulunmuştur. Küme içi korelasyon katsayısı 1 olup; test istatistiği 7.36, buna ilişkin p değeri ise 0.007 olarak elde edilmiştir. Fakat, toplam uyumsuz çift oranının (%12) düşük ve en az bir uyumsuz çifte sahip küme başına düşen ortalama uyumsuz çift sayısının 1 değerine çok yakın olması nedeniyle sadece uyumsuz çiftlerden elde edilen ilişki katsayısına güvenilmez. Bu durumda, tüm çiftlerden elde edilen ilişki katsayısına dayalı olarak önerilen McNemar testi kullanılmalıdır. Bu test sonuçlarına göre, küme içi korelasyon katsayısı -0.08 olup, test istatistiği 11.56, ilişkin p değeri ise 0.0007'dir.

Eliasziw ve Donner tarafından önerilen yöntemler, her bir küme için uyumlu ve uyumsuz çift sayısının bilinmesini gerektirir. Dolayısıyla bu testlerin uygulanması için Tablo 2'deki bilgiler yeterli değildir.

Obuchowski'nin önerdiği örnekleme yaklaşımında, OP yönteminin hasta tanısı koyma olasılığı 0.523; standart hatası 0.027'dir. K yönteminde ise bu değerler, sırasıyla 0.457 ve 0.03'tür. OP ve K yöntemlerinin hasta tanısı koyma olasılıkları arasındaki farklılığın varyans tahmini 0.0006'dır. Bu yöntemle ait test istatistiği 7.28; buna ilişkin p değeri 0.007 olarak elde edilmiştir.

Durkalski ve arkadaşları tarafından önerilen yöntemde test istatistiği 7.41 ve p değeri 0.006 olarak bulunmuştur.

Son olarak, küme içi korelasyon göz ardı edildiğinde elde edilecek olan standart McNemar test istatistiği ise 11.11 ve p değeri 0.0008'dir.

Yukarıda bahsedilen test istatistiklerinin tümü, OP ve K testlerinin hasta tanısı koyma olasılıklarının farklı olduğunu göstermektedir.

#### 4. TARTIŞMA VE SONUÇ

McNemar testinin temel varsayımı, her bir gözlem çiftinin bağımsız olmasıdır. Fakat, bazı çalışmalarda bir denekten birden fazla ölçüm olarak kümelenmiş veriler üzerinden analiz yapmak gerekebilir. Yapılan Monte Carlo benzetim çalışmalarında, bağımsızlık varsayımının bozulmasının I.tip hata oranında 10 kat artışa neden olabildiği gösterilmiştir (Eliasziw and Donner, 1991).

Eliasziw ve Donner'ın önerdikleri yöntem, uygulamasının kolay olması bakımından diğerlerine kıyasla en sık kullanılan yöntemdir. Fakat, örneklemede sadece bir tür uyumsuz çift olduğunda, sadece uyumsuz çiftlerden elde edilen ilişki katsayısına dayalı olan McNemar testi uygulanamaz. Ayrıca, testin gücünün artması bakımından,



Eliasziw ve Donner, tüm çiftlerden elde edilen ilişki tahminine dayalı olan yöntemin kullanılmasını önermişlerdir. Donald ve Donner (1987), bu yöntemin küme-içi korelasyon küme büyüklüğüne bağlı olarak değişmediğinde ya da pozitif sonuç elde etme olasılığı, küme içi ölçümler arasında sabit kaldığı durumda, çok iyi sonuç verdiğini belirtmişlerdir.

Obuchowski, önerdiği yöntem ile Eliasziw ve Donner tarafından önerilen yöntemi bir benzetim çalışması ile karşılaştırmıştır. Obuchowski, “küme büyüklüklerinin değişkenlik gösterdiği” veya “testler arasındaki ilişkinin, küme içindeki tüm ölçümler arasında eşit olmadığı” durumlarda, önerdiği yöntemin, testin gücünde yüzde üçlük bir kayba neden olsa da, Eliasziw ve Donner’ın yöntemi kadar iyi sonuçlar verdiğini, küme büyüklükleri arasındaki farklılık çok büyük olduğunda ise kendi yönteminin tercih edilmesi gerektiğini belirtmiştir. Bu yöntemin bir diğer avantajı da, her bir kümedeki uyumsuz çift sayısının bilinmesine gerek duyulmaması, dolayısıyla uygulanmasının daha kolay olmasıdır.

Scott ve Wu (1981) Obuchowski’nin önerdiği örnekleme yaklaşımının, küme sayısı arttıkça, tutarlı bir varyans tahmini verdiğini göstermişlerdir.

Durkalski ve arkadaşları (2003) yaptıkları Monte Carlo benzetim çalışması sonucunda, önerdikleri yöntemin; kümelerde pozitif sonuç elde etme olasılığının eşit olduğu ve küme büyüklükleri arasındaki farkın 5 birimi geçmediği durumlarda Obuchowski’nin yöntemi ile benzer sonuçlar verdiğini göstermişlerdir. Fakat, küme büyüklükleri arasındaki fark 1 ile 10 arasında değiştiğinde, önerdikleri yöntem testin gücünde yüzde altılık bir kayba neden olurken, Obuchowski’nin önerdiği yöntem en iyi sonucu vermiştir. Kümelerde pozitif sonuç elde etme olasılıkları değişkenlik gösterdiğinde de benzer sonuçlar bulunmuştur.

## KAYNAKLAR

- DONALD, A., DONNER, A. (1987), *Adjustments to the Mantel-Haenszel Chi-Square Statistics and Odds Ratio Variance Estimator When the Data Are Clustered*, *Statistics in Medicine*, 6, 491-499.
- DURKALSKI, V. L., PALESCH, Y.Y., Lipsitz, S. R., Rust, P. F. (2003), *Analysis of Matched-Pair Data*, *Statistics in Medicine*, 22, 2417-2428.
- ELIASZIW, M., DONNER, A. (1991), *Application of the McNemar Test to Non-Independent Matched Pair Data*, *Statistics in Medicine*, 10, 1981-1991.
- OBUCHOWSKI, N. (1998), *On the Comparison of Correlated Proportions for Clustered Data*, *Statistics in Medicine*, 17, 1495-1507.
- RAO, JNK, SCOTT, AJ. (1992), *A Simple Method for the Analysis of Clustered Binary Data*, *Biometrics*, 48, 577-585.
- ROYALL, R.M. (1986), *The Prediction Approach to Robust Variance Estimation in Two-Stage Cluster Sampling*, *Journal of the American Statistical Association*, 81, 119-123.

SCOTT, A.J., WU, C.F.J. (1981), *On the Asymptotic Distribution of Ratio and Regression Estimators*, Journal of the American Statistical Association, 76, 98-102.

SNEDECOR, G., COCHRAN, WG. (1980), *Statistical Methods*, 7th edn, Iowa State University Pres: Ames.

## Comparison of Two Correlated Proportions for Clustered Data

### ABSTRACT

*McNemar's one degree of freedom chi-square test is used to compare two proportions estimated from dependent samples. The main assumption of this test is that the paired responses are independent. However, in some studies more than one observation is taken from an individual and this assumption is not hold. This data structure is called as "clustered data. The simulation studies have been shown that a violation of this assumption could lead to an increase in the type I error rate. Examples of clustered data can be given from studies made on radyology, periodontology and ophthalmology. In this paper, methods proposed by Eliasziw and Donner, Obuchowski and Durkalski et.al has been introduced. These methods has been applied on a data set of a study made in radyology and compared to the result of standard McNemar test.*

**Key Words:** Clustered data, McNemar test

## Lojistik ve Cox Regresyon Modellerinin İncelenmesi ve Karşılaştırılması

Durdu SERTKAYA\*

### ÖZET

*Bu çalışmada, epidemiyolojik çalışmalarda herhangi bir olayın risk faktörlerini belirlemek için kullanılan lojistik regresyon ve Cox regresyon modelleri ele alınmıştır. Bu modeller hakkında bilgi verilmiş, aralarındaki matematiksel ilişkiden ve karşılaştırmalardan bahsedilmiştir. Mide kanseri verileri kullanılarak bir uygulama yapılmıştır.*

*Anahtar Kelimeler:* Lojistik Regresyon, Cox Regresyon, Yaşam Çözümlemesi

### 1. GİRİŞ

Son yıllarda epidemiyolojik çalışmalardaki veri analizlerinde birçok farklı yaklaşımlar kullanılmaktadır (Green ve Symons, 1983). Son on yıldır klinik çalışmalarda herhangi bir olayın risk faktörlerini belirlemek için göreceli risk (relative risk) regresyon modelleri vazgeçilmez araçlar olmuşlardır. Klinik araştırmalarda hala birbiriyle yarışan iki regresyon modeli vardır. Bunlar lojistik regresyon modeli ve Cox regresyon modelidir. Risk faktörlerini belirlemede bu iki modelden hangisinin daha uygun olduğu hala tartışılmaktadır. Risk faktörlerini tanımlamada, tıp literatüründe görülen istatistiksel analizlerin çoğu lojistik modele dayanmaktadır. Cox regresyon modeli ise onkoloji ve hematoloji gibi tıp alanlarında daha çok görülmektedir (Chevret, 2001). Lojistik regresyon ve Cox regresyon modelleri kronik hastalık uygulamalarında risk faktörlerini belirlemek için epidemiyolojik çalışmaların analizinde kullanılmaktadır. Green ve Symons (1983) bu iki model arasındaki teorik ilişkiyi ortaya koymuşlardır (Green ve Symons, 1983). Gerçekte, lojistik ve Cox regresyonlarının her ikisi de log-doğrusal olarak, belirlenmiş bir olay için risk faktörlerinin kümesiyle ilgili olarak matematiksel modelleme yaklaşımlarıdır (Chevret, 2001). Her iki yöntemde, risk faktörleri kümesi ile belirlenmiş bir olay (genellikle ölüm ya da hastalığın ortaya çıkması) arasındaki ilişkiyi değerlendirmektedir (Ingram ve Kleinman, 1989).

Tıp literatüründe risk faktörlerini belirlemede daha çok lojistik regresyon modeli ile karşılaşılmaktadır. Oysa ki, özellikle epidemiyolojik çalışmalarda hastalar uzun süre takip edildiğinden Cox regresyon modeli daha uygun olmaktadır. Çünkü lojistik

\* Öğr. Gör. Dr., Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Beytepe-Ankara.  
durdu@hacettepe.edu.tr



modelde izleme süresi sabit varsayılmaktadır. Bu durumda da gözlem farklılıkları dikkate alınmamaktadır.

Bu çalışmanın amacı, lojistik ve Cox regresyon modellerini tanıtmak, aralarındaki matematiksel ilişkiyi vermek ve bu iki modeli karşılaştırmaktır.

## 2. REGRESYON MODELLERİ

### 2.1. Lojistik Regresyon Modeli

Lojistik regresyon, bağımlı değişkenin ikili (binary, dichotomous), üçlü ve çoklu kategorilerde gözlemlendiği durumlarda açıklayıcı değişkenlerle neden-sonuç ilişkisini belirlemede yararlanılan bir yöntemdir. Açıklayıcı değişkenlere göre yanıt değişkeninin beklenen değerinin olasılık olarak elde edildiği bir regresyon yöntemidir. Bağımlı değişken üzerinde açıklayıcı değişkenlerin etkileri olasılık olarak elde edilerek risk faktörlerinin olasılık olarak belirlenmesi sağlanmaktadır (Özdamar, 1997).

İkili yanıt değişkeni (bağımlı değişken) Y,

$$Y = \begin{cases} 1, & \text{izleme periyodu boyunca olay meydana gelmişse} \\ 0, & \text{ö.d.} \end{cases} \quad (1)$$

biçiminde tanımlansın. Lojistik model,

$$P(Y | \mathbf{z}, T) = \{1 + \exp[-(\beta_0 + \sum \beta_i z_i)]\}^{-1}$$

biçiminde olmaktadır. Burada  $z_i$  ( $i=1,2,\dots,r$ ) açıklayıcı değişkenler, T, bütün bireyler için aynı olduğu varsayılan toplam izleme uzunluğu,  $\beta_0$  sabit terim ve  $\beta_i$  ( $i=1,2,\dots,r$ ) regresyon katsayıları olarak ifade edilmektedir. Buradan

$$\frac{P(Y = 1 | \mathbf{z}, T)}{P(Y = 0 | \mathbf{z}, T)} = \exp(\beta_0 + \sum \beta_i z_i) \quad (2)$$

olmaktadır. İkili açıklayıcı değişken durumu için (2) eşitliği

$$\frac{P(Y = 1 | \mathbf{z}, T)}{P(Y = 0 | \mathbf{z}, T)} = \begin{cases} e^{\beta_0} & \mathbf{z} = 0 \text{ ise} \\ e^{\beta_0 + \beta} & \mathbf{z} = 1 \text{ ise} \end{cases} \quad (3)$$

ifadesine indirgenmekte ve odds oranı  $e^\beta$  olmaktadır.

(2) ve (3) eşitliklerinden de görüldüğü gibi lojistik model ölüm süresini ve farklı izleme uzunluklarını dikkate almamaktadır. Farklı izleme uzunlukları olduğunda iki yaklaşım uygulanabilmektedir. İlki, eğer farklı izleme uzunluklarının parametre tahminleri üzerine etkilerinin az oldukları varsayılırsa, izleme süreleri dikkate alınmadan bütün bireyler analize dahil edilebilir. Diğer bir yaklaşım ise, T izleme

süresinden daha küçük olan yani yaşayanları analizden çıkarmaktır. Bu yaklaşımların her ikisi de odds oranının yanlı tahmini ile sonuçlanabilmektedir (Ingram ve Kleinman, 1989).

## 2.2. Cox Regresyon Modeli

Birçok durumda incelenen yaşam süresinin başka faktörler tarafından da etkilenebileceği gözönünde bulundurulduğunda, bağımlı değişken olan yaşam süresi üzerinde açıklayıcı değişkenlerin de etkilerinin modellendiği regresyon modelleri yaşam çözümlemesinde önemli bir yer tutmaktadır. Orantılı hazard modeli olarak da ifade edilen bu modeller ilk kez Cox (1972) tarafından ele alınmıştır (Lawless, 1982).

Cox (1972)'un incelediği model

$$h(t, \mathbf{z}) = h_0(t) \exp(\boldsymbol{\beta}' \mathbf{z}) \quad (4)$$

biçimindedir. Burada  $\mathbf{z}$ , açıklayıcı değişkenler vektörü,  $\boldsymbol{\beta}$ , regresyon katsayıları (bilinmeyen katsayılar) vektörü ve  $h_0(t)$ ,  $\mathbf{z} = 0$  olan bir birimin temel hazard fonksiyonu olmaktadır (Cox, 1972; Cox ve Oakes, 1984).  $h(t, \mathbf{z})$ ,  $\mathbf{z}$  açıklayıcı değişkenlerle  $t$  zamanında bir birey için hazard hızını ifade etmektedir.  $h_0(t)$  için özel bir biçim varsayımı yapılmamakta ancak farklı açıklayıcı değişken kümeleri için hazardların orantılı oldukları varsayılmaktadır.  $\boldsymbol{\beta}$ 'nin tahmini

$$L(\boldsymbol{\beta}) = \prod_{k=1}^K \frac{\exp(\boldsymbol{\beta} z_k)}{\sum_{j \in R_i} \exp(\boldsymbol{\beta} z_j)}$$

kısmi olabilirlik fonksiyonu yardımıyla elde edilebilmektedir. Burada  $K$ , ortaya çıkan olay sayısını,  $R_i$  ise  $i$ . olayın meydana gelme zamanında riskte olan bireylerin kümesini göstermektedir (Kalbfleisch ve Prentice, 1980; Ingram ve Kleinman, 1989; Schemper, 1992). Cox regresyon modeli, risk kümesinden ayrılan bireyler tarafından tamamlanmış izleme için de uygun olmaktadır.

İkili açıklayıcı değişken için (4) eşitliği,

$$h(t, \mathbf{z}) = \begin{cases} h_0(t) & \mathbf{z} = 0 \text{ ise} \\ h_0(t) e^{\boldsymbol{\beta}} & \mathbf{z} = 1 \text{ ise} \end{cases}$$

biçimine indirgenmektedir. Hazard oranı ya da görel risk ise  $e^{\boldsymbol{\beta}}$  biçiminde ifade edilmektedir (Ingram ve Kleinman, 1989). Lojistik regresyondaki odds oranı ve Cox regresyondaki görel risk, elde edilen regresyon modelinin veri kümesini ne kadar iyi açıkladığını değerlendirebilmek için gerekli olan özet ölçümlerdir. Görel risk ve odds oranı, önemli bulunan değişken düzeyinin referans kategorisine göre kaç kat daha riskli olduğu yorumunu vermektedir (Hosmer ve Lemeshow, 1999).

### 3. LOJİSTİK ve COX REGRESYON MODELLERİ ARASINDAKİ İLİŞKİ VE BU MODELLERİN KARŞILAŞTIRILMASI

N tane birey olsun. r tane risk faktörü  $\mathbf{z} = (z_1, z_2, \dots, z_r)$ , herbiri T uzunluğundaki izleme periyodunun başlangıcında ölçülmüş olsun. İkili yanıt değişkeni Y, (1) eşitliğindeki gibi tanımlansın. Bu durumda  $\beta_0$  ve  $\beta$  katsayıları ile  $\mathbf{z}$  risk faktörlerinin çoklu lojistik fonksiyonu

$$P(Y=1 | \mathbf{z}, T) = \{1 + \exp[-\beta_0 - \beta' \mathbf{z}]\}^{-1} \quad (5)$$

biçiminde verilmektedir. T uzunluğundaki izleme periyoduna karşılık gelen yaşam olasılığı

$$P(Y=0 | \mathbf{z}, T) = S(T | \mathbf{z}, \beta) = \exp[-\beta_0 - \beta' \mathbf{z}] \{1 + \exp[-\beta_0 - \beta' \mathbf{z}]\}^{-1} \quad (6)$$

olmaktadır. (5) ve (6) eşitliklerinde izleme süresi T, bütün bireyler için aynı olmayı gerektirmektedir.

Cox (1972) tarafından önerilen

$$h(t, \mathbf{z}) = h_0(t) \exp(\alpha' \mathbf{z}) \quad (7)$$

modeli pozitif t zamanındaki hazarda dayalıdır. Bu modeldeki  $\alpha$  katsayı vektörü, (5) ve (6) eşitliklerindeki  $\beta_0$  ve  $\beta$  katsayılarıyla aynı olmayabilmektedir. (7) eşitliği ile verilen hazard, orantılı (proportional) olarak adlandırılmaktadır. Çünkü hazard, temel hazard olan  $h_0(t)$ 'nin ve hastalığın risk faktörlerinin bir fonksiyonu olan  $\exp(\alpha' \mathbf{z})$ 'nin çarpımı olarak ifade edilmektedir. Hazard modeline karşılık gelen yaşam (survival) fonksiyonu,

$$S(T | \mathbf{z}, \alpha) = \exp\left[-\int_0^T h_0(t) \exp(\alpha' \mathbf{z}) dt\right] \quad (8)$$

biçiminde verilmektedir. İzleme süresi T, Cox modelde herbir birey için farklı olabilmektedir.

Lojistik regresyon ile Cox modelin,  $h_0(t) = h$  özel durumu arasındaki ilişki belirtilmiştir. Lojistik model ve  $S(T | \mathbf{z}, \beta)$ 'nin (6) eşitliğinde verilen formu

$$\begin{aligned} -\ln[S(T | \mathbf{z}, \beta)] &= \ln\{[1 + \exp(-\beta_0 - \beta' \mathbf{z})][\exp(-\beta_0 - \beta' \mathbf{z})]^{-1}\} \\ &= \ln\left(\frac{1+u}{u}\right) = \ln\left(1 + \frac{1}{u}\right) \end{aligned}$$

biçiminde yazılabilmektedir. Burada  $u = \exp[-\beta_0 - \beta' \mathbf{z}]$  olmaktadır. Eğer  $u^{-1}$  küçük ise Taylor serisi açılımı ile



$$\ln\left(\frac{1+u}{u}\right) \cong \frac{1}{u}$$

yazılabilmektedir.

$$-\ln[S(T | \mathbf{z}, \boldsymbol{\beta})] \cong \exp(\beta_0 + \boldsymbol{\beta}'\mathbf{z}) \quad (9)$$

yaklaşımı,  $\exp(\beta_0 + \boldsymbol{\beta}'\mathbf{z})$ 'nin küçük olmasını sağlamaktadır. Böylece, (9) ifadesi ile verilen lojistik yaşam fonksiyonu ve (8) ifadesi ile verilen Cox model yaşam fonksiyonu

$$\exp(\beta_0 + \boldsymbol{\beta}'\mathbf{z}) \cong \int_0^T h_0(t) \exp(\boldsymbol{\alpha}'\mathbf{z}) dt$$

olduğunda yaklaşık olarak denk olmaktadır. Eğer temel hazard sabit ise ( $h_0(t) = h$ )

$$\exp(\beta_0 + \boldsymbol{\beta}'\mathbf{z}) \cong hT \exp(\boldsymbol{\alpha}'\mathbf{z}) \quad (10)$$

olmaktadır. (10) ifadesi ile verilen yaklaşık ilişki,  $\beta_0 = \ln(hT)$  ve  $i = 1, 2, \dots, r$  için

$$\beta_i \cong \alpha_i$$

olacaktır. İzleme uzunluğu  $T$ , lojistik regresyondaki  $\beta_0$  katsayısını doğrudan etkilemektedir. Lojistik regresyondaki risk faktörlerinin katsayıları, sabit temel hazard ve risk faktörlerinin orantılı log-doğrusal fonksiyonu ile Cox modelin katsayılarına yaklaşık olarak eşit olmaktadır. Bu sonuç,  $\exp(\beta_0 + \boldsymbol{\beta}'\mathbf{z})$  küçük olduğu zaman sağlanmaktadır. Bu yaklaşım, olayın meydana gelmesi seyrek olduğunda ve risk faktörlerinin etkileri çok büyük olmadığı kısa izleme periyotları ile daha da iyileşebilmektedir. Diğer taraftan, daha genel olaylar için, uzun izleme periyotları durumunda ya da risk faktörlü olaylar için bu yaklaşım kötüleşmektedir (Green ve Symons, 1983). Yani, uzun izleme süresi durumunda Cox regresyon modeli, lojistik modelden daha çok bilgi vermekte ve ilgilenilen olay ve risk faktörleri arasındaki ilişkinin daha güçlü değerlendirmesini sağlamaktadır (Ingram ve Kleinman, 1989). Ayrıca, Cox regresyondan elde edilen katsayıların standart hataları azalmaktadır (Green ve Symons, 1983).

Lojistik regresyon modeli ile Cox regresyon modeli arasında iki ana fark vardır: Birincisi, her ikisi de farklı metrik kullanarak birlikteliği (ilişkiyi) ölçmektedirler. Cox regresyon süreyi dikkate alarak görel riski ölçerken, lojistik regresyon ise sabit izleme süresi varsayımı altında görel odds'u ölçmektedir. Bu durumda, olayın ortaya çıkma hızı yüksek olduğunda görel odds, görel riskten çok farklı olabilmektedir. İkinci fark ise, Cox model farklı izleme sürelerine izin vermektedir. Örneğin, durdurulmuş (censored) (bazı olayların ortaya çıkma zamanlarının bilinmediği fakat sadece gözlem periyodundan daha büyük olduğunun bilindiği durum) veriler için Cox regresyon modeli uygun olmaktadır. Lojistik model ise ikili (binary, dichotomous) (iki değer alıyor, örneğin, hastalık var-yok, yaşıyor-öldü gibi) olaylarla ilgili olarak bütün bireyler için izleme süresinin sabit olduğunu varsaymaktadır (Ingram ve Kleinman, 2001). Ancak bu

epidemiyolojik çalışmalarda çok nadir bir durumdur. Genellikle bireyler farklı gözlem periyotlarına sahip olmaktadır (Estévez vd, 2001).

Lojistik regresyonda zaman sabit alındığından izleme periyodunun başlangıcına yakın meydana gelen olayla, izleme periyodunun sonuna yakın meydana gelen olaya aynı ağırlık verilmektedir. Oysaki, gerçekte kabul edilen risk faktörünün bir hastalık riskine katkısının önemliliği, izleme periyodunun uzunluğuna bağlı olabilmektedir (Green ve Symons, 1983). Cox (1972), süreyi ve herbir birey üzerinden ölçülen açıklayıcı değişkenleri (covariates) de içeren bir model önermiştir. Cox modeli, tam başarısızlık zamanını kullanmakta ve durdurmayı da dikkate almaktadır (Chevret, 2001).

Bu farklılara rağmen, olayın ortaya çıkması seyrek olduğunda, risk faktörlerinin etkileri zayıf olduğunda ve izleme periyodu kısa olduğunda iki modelden elde edilen parametre tahminleri benzer olabilmektedir (Green ve Symons, 1983; Hauck, 1985; Ingram ve Kleinman, 1989; Chevret, 2001; Estévez, 2001). İzlemenin erken safhalarında iki model hemen hemen aynı sonucu vermektedir. Ancak, daha uzun izleme süreleri için iki modelden elde edilen regresyon katsayıları farklılık göstermektedir ve Cox regresyondan elde edilen katsayıların standart hataları azalmaktadır (Green ve Symons, 1983).

Lojistik model standart bir yaklaşım olarak görülürken son zamanlarda ki yeni istatistiksel yaklaşımlar Cox modelin kullanımını yaygınlaştırmışlardır. Epidemiyolojik çalışmalarda hastalar genellikle ölene kadar ya da takip edilebildiği kadar izlenmektedir. Bu da hastalar arasında gözlem periyotlarında büyük farklar yaratmaktadır. Lojistik modelde bu süre sabit varsayıldığından, Cox taraftarları, gözlem farklılıklarının lojistik modelde kullanılmadığını ifade etmişlerdir (Chevret, 2001). Cox regresyondaki katsayıların üstelliği ( $e^{\beta}$ ) hazard oranlarına (hazard ratios) karşılık gelmektedir. Bu da risk skorunu belirtmekte ve yüksek skorlar, olayın ortaya çıkması için yüksek riskleri göstermektedir (Estevez vd, 2001).

#### 4. MİDE KANSERİ İLE İLGİLİ BİR UYGULAMA

Lojistik ve Cox regresyon modellerinin uygulanması ve karşılaştırılması için Ankara Onkoloji Hastanesi'nde mide kanseri tanısı konan 118 hastaya ait veriler kullanılmıştır. Açıklayıcı değişkenler olarak yaş, cinsiyet, ağırlık kaybı, anemi, tümörün midedeki lokalizasyonu, lenf nodu diseksiyonunun genişliği, hastalığın evresi, tümörün mide duvarındaki invazyon derinliği, tümörün venöz, lenfatik ve perinöral invazyon yapması, adjuvant kemoterapi alınmıştır. 118 hastanın 52'si (%44.1) ölmüş, 66'sı (%55.9) ise çalışmanın sonunda yaşamını sürdürmektedir. Mide kanserinden ölümler olay olarak alınmış, halen yaşayanlar ise durdurulmuş olarak kabul edilmiştir. Ortalama yaş  $56,7 \pm 11,34$  (29-84 yıl) olarak bulunmuştur. 3 yıllık yaşam olasılığı %42.47'dir. Ortanca yaşam süresi ise  $29 \pm 6$ 'dır.

Kullanılan değişkenler ve düzeyleri Tablo 1.'de verilmiştir.



Lojistik ve Cox regresyon modellerinde herbir değişken için ilk düzey referans kategorisi olarak alınmıştır. Bu ise değişken düzeylerini yorumlamada önemli olmaktadır. Bütün testler için p değeri 0.05'den küçük ise önemli olarak kabul edilmiştir.

Başarısızlığı etkileyen risk faktörlerini belirlemek için öncelikle tek değişkenli lojistik ve Cox regresyon modelleri incelenmiştir. Modellerdeki değişken için  $\beta$  parametresi, standart hatası, p-değeri,  $\beta$  parametresinin risk oranı ( $\exp(\beta)$ ) ile önemli bulunan değişken düzeyleri için risk oranının alt ve üst sınırları verilmiştir.  $\beta$  parametresinin pozitif değer olması bu düzeyin referans kategorisine göre daha fazla riskli olduğunu,  $\beta$  parametresinin negatif değer olması ise bu düzeyin referans kategorisine göre daha az riskli olduğunu göstermektedir. Lojistik regresyonda odds oranı, Cox regresyonda ise risk oranı olarak ifade edilen  $\exp(\beta)$  değeri ise önemli bulunan düzeyin, referans kategorisine göre kaç kat (ya da % ne kadar) daha riskli olduğu yorumunu getirmektedir. Lojistik regresyondan elde edilen sonuçlar Tablo 2.'de, Cox regresyondan elde edilen sonuçlar ise Tablo 3.'de verilmiştir.

**Tablo 1.** Açıklayıcı Değişkenler ve Düzeyleri

Değişken	Değişken Düzeyleri	n	%
Cinsiyet (CİNS)	1. Erkek	69	58.5
	2. Kadın	49	41.5
Kilo Kaybı (KİLO)	1. Yok	90	76.3
	2. Var	28	23.7
Anemi (ANEMİ)	1. Yok	31	26.3
	2. Var	87	73.7
Lenf nodu diseksiyonunun genişliği (DİSEKSİYON)	1. D <sub>0-1</sub>	62	52.5
	2. D <sub>2-3</sub>	56	47.5
Tümörün midedeki lokalizasyonu (LOKAL)	1. Üst üçlük	21	17.8
	2. Orta üçlük	25	21.2
	3. Alt üçlük	62	52.5
	4. Tüm mide	10	8.5
Tümörün mide duvarındaki invazyon derinliği (DERİNLİK)	1. T <sub>1</sub> +T <sub>2</sub>	17	14.4
	2. T <sub>3</sub>	57	48.3
	3. T <sub>4</sub>	44	37.3
Hastalığın evresi (EVRE)	1. Evre1+Evre2	30	25.4
	2. Evre3	65	55.1
	3. Evre4	23	19.5
Tümörün venöz invazyon yapması (VENÖZ)	1. Yok	87	73.7
	2. Var	31	26.3
Tümörün lenfatik invazyon yapması (LENFATİK)	1. Yok	84	71.2
	2. Var	34	28.8
Tümörün perinöral invazyon yapması (PRİNÖRAL)	1. Yok	102	86.4
	2. Var	16	13.6
Adjuvant kemoterapi (KEMOTERAPİ)	1. Yok	34	28.8
	2. Var	84	71.2



**Tablo 2.** Tek Değişkenli Lojistik Regresyon Sonuçları

Değişken	$\beta$	Standart Hatası	p	Exp( $\beta$ )	Alt sınır-Üst sınır
YAŞ	-.028	.0169	.0958	.972	-
CİNS	-.369	.379	.3300	.691	-
KİLO	.313	.434	.4699	1.368	-
ANEMİ	.672	.440	.1263	1.959	-
DİSEKSİYON	-1.558	.404	.0001	.210	.095 - .465
LOKAL			.0235		
ORTA ÜÇLÜK	1.098	.617	.0751	3.000	-
ALT ÜÇLÜK	.0953	.533	.8582	1.100	-
TÜM MİDE	2.079	.916	.0232	7.999	1.328 - 48.182
DERİNLİK			.0375		
T3	1.978	.797	.0131	7.227	1.514 - 34.499
T4	2.013	.810	.0130	7.485	1.529 - 36.645
EVRE			.0018		
EVRE3	2.087	.592	.0004	8.062	2.526 - 25.727
EVRE4	1.958	.680	.0040	7.085	1.868 - 26.866
VENÖZ	1.977	.487	.0000	7.224	2.781 - 18.765
LENFATİK	1.377	.432	.0014	3.965	1.699 - 9.256
PERİNÖRAL	.867	.554	.1175	2.381	-
KEMOTERAPİ	-.842	.415	.0422	.431	.191 - .971

Tablo 2'deki p değerleri incelendiğinde DİSEKSİYON, LOKAL, DERİNLİK, EVRE, VENÖZ, LENFATİK ve KEMOTERAPİ değişkenlerinin mide kanserinden başarısızlık (ölüm) için önemli risk faktörleri olduğu söylenebilmektedir ( $p < 0.05$ ). Önemli bulunan bu değişkenlerin herbir düzeyine karşılık gelen p değerlerine bakılarak önemli değişken düzeyleri belirlenebilmektedir. Herbir değişken için ilk düzeyler referans kategorisi olarak alındığından çizelgede yer almamaktadır. DİSEKSİYON değişkeni önemli bulunmuştur. Buna karşılık gelen katsayı da negatif olduğundan, lenf nodu diseksiyonunun  $D_{2-3}$  olmasının  $D_{0-1}$  olmasına göre başarısızlık riskinin %21 ( $\exp(\beta) = .21$ ) daha az olduğu söylenebilmektedir. LOKAL değişkeninin ise dördüncü düzeyi olan TÜM MİDE önemli olarak bulunmuştur. Buradan tümörün midedeki lokalizasyonu tüm mide olduğunda üst üçlük olmasına göre 8 kat ( $\exp(\beta) = 7.999$ ) daha fazla riskli olduğu yorumlanabilmektedir. Diğer değişkenler ve düzeyleri içinde aynı yorumlar yapılabilmektedir.

**Tablo 3.** Tek Değişkenli Cox Regresyon Sonuçları

Değişken	$\beta$	Standart Hatası	p	Exp( $\beta$ )	Alt sınır-Üst sınır
YAŞ	-.023	.013	.0671	.977	-
CİNS	-.304	.288	.2924	.738	-
KİLO	.162	.313	.6062	1.175	-
ANEMİ	.585	.353	.0979	1.795	-
DİSEKSİYON	-1.041	.313	.0009	.3532	.191 - .653
LOKAL			.0307		
ORTA ÜÇLÜK	.695	.459	.1302	2.004	-
ALT ÜÇLÜK	.199	.434	.6470	1.220	-
TÜM MİDE	1.269	.521	.0149	3.558	1.281 - 9.885
DERİNLİK			.0301		
T3	1.619	.732	.0270	5.050	1.202 - 21.213
T4	1.931	.742	.0093	6.894	1.610 - 29.514
EVRE			.0007		
EVRE3	1.952	.534	.0003	7.045	2.471 - 20.084
EVRE4	2.154	.595	.0003	8.617	2.684 - 27.669
VENÖZ	1.252	.280	.0000	3.498	2.021 - 6.055
LENFATİK	.842	.283	.0030	2.321	1.331 - 4.046
PERİNÖRAL	.867	.554	.1175	2.381	-
KEMOTERAPİ	-.451	.285	.114	.636	.364 - 1.115

Tek değişkenli Cox regresyonu sonucunda elde edilen ve Tablo 3’de verilen p değerleri incelendiğinde DİSEKSİYON, LOKAL, DERİNLİK, EVRE, VENÖZ ve LENFATİK değişkenlerinin mide kanserinden başarısızlık için önemli risk faktörleri olduğu söylenebilmektedir ( $p < 0.05$ ). DİSEKSİYON değişkeni önemli bulunduğundan lenf nodu diseksiyonunun D<sub>2-3</sub> olmasının D<sub>0-1</sub> olmasına göre başarısızlık riskinin %35 ( $\text{exp}(\beta) = .3532$ ) daha az olduğu söylenebilmektedir. LOKAL değişkeninin ise dördüncü düzeyi olan TÜM MİDE önemli olarak bulunmuştur. Buradan tümörün midedeki lokalizasyonu tüm mide olduğunda üst üçlük olmasına göre 3.5 kat ( $\text{exp}(\beta) = 3.558$ ) daha fazla riskli olduğu yorumlanabilmektedir. Diğer değişkenler ve düzeyleri içinde aynı yorumlar yapılabilmektedir.

Tek değişkenli lojistik regresyondan elde edilen sonuçlarla tek değişkenli Cox regresyondan elde edilen sonuçları karşılaştırdığımızda KEMOTERAPİ değişkeni dışında aynı değişkenlerin önemli bulunduğunu ancak Cox regresyondan elde edilen katsayıların standart hatalarının daha küçük olduğu dikkat çekmektedir. Bu da lojistik regresyonla Cox regresyonu ayıran önemli bir özellik olmaktadır.

Bundan sonra ise tüm değişkenler birlikte ele alınarak adımsal seçim yöntemiyle lojistik ve Cox regresyon çözümlenmeleri yapılmış ve son adımda elde edilen sonuçlar sırasıyla Tablo 4. ve Tablo 5.’de verilmiştir.

**Tablo 4.**Adımsal Lojistik Regresyon Sonuçları

Değişken	$\beta$	Standart Hatası	p	Exp( $\beta$ )	Alt sınır-Üst sınır
DİSEKSİYON	-2.497	.658	.0001	.082	.023 - .298
LOKAL			.0061		
ORTA ÜÇLÜK	3.698	1.138	.0012	40.377	4.337 - 375.872
ALT ÜÇLÜK	1.794	.906	.0477	6.012	1.018 - 35.498
TÜM MİDE	3.553	1.433	.0132	34.918	2.103 - 579.801
EVRE			.0016		
EVRE3	3.093	.889	.0005	22.047	3.859 - 125.938
EVRE4	3.465	1.071	.0012	31.965	3.917 - 260.819
VENÖZ	2.791	.757	.0002	16.293	3.694 - 71.858
KEMOTERAPİ	-3.304	.873	.0002	.036	.007 - .203

**Tablo 5.**Adımsal Cox Regresyon Sonuçları

Değişken	$\beta$	Standart Hatası	P	Exp( $\beta$ )	Alt sınır-Üst sınır
DİSEKSİYON	-.821	.333	.0137	.440	.229 - .845
LOKAL			.0552		
ORTA ÜÇLÜK	1.168	.490	.0172	3.216	1.229 - 8.409
ALT ÜÇLÜK	.883	.457	.0536	2.417	-
TÜM MİDE	1.390	.546	.0109	4.016	1.377 - 11.712
EVRE			.0012		
EVRE3	1.641	.561	.0034	5.162	1.718 - 15.505
EVRE4	2.322	.635	.0003	10.200	2.937 - 35.425
VENÖZ	1.016	.307	.0009	2.762	1.513 - 5.044
KEMOTERAPİ	-1.190	.317	.0002	.304	.163 - .567

Adımsal seçim yöntemiyle elde edilen sonuçları karşılaştırdığımızda, her iki çözümlenin son adımında DİSEKSİYON, LOKAL, EVRE, VENÖZ ve KEMOTERAPİ değişkenlerinin olduğu modele ulaşılmış ve bu değişkenler önemli olarak bulunmuştur. Ancak lojistik regresyonda LOKAL değişkeninin ALT ÜÇLÜK düzeyi önemli olarak bulunurken Cox regresyonda bu düzey anlamlı olarak bulunmamıştır.

Elde edilen sonuçlardan Cox regresyonunun lojistik regresyona göre daha tutucu olduğunu söyleyebiliriz. Ayrıca Cox regresyondan elde edilen katsayıların standart hatalarının lojistik regresyondan elde edilen katsayıların standart hatalarına göre daha küçük olduğunu ve Cox regresyonu sonucunda elde edilen değerlerin daha güvenilir olduğunu söyleyebilmekteyiz.



## 5. SONUÇ

Bu çalışmada, lojistik ve Cox regresyon modelleri hakkında bilgi verildikten sonra aralarındaki matematiksel ilişkiden bahsedilmiş ve bu iki model karşılaştırılmıştır. Mide kanseri verileri kullanılarak da bir uygulama yapılmıştır. Bu modellerle başarısızlığı etkileyen önemli açıklayıcı değişkenler belirlenebilmekte, açıklayıcı değişkenlerin düzeyleri karşılaştırılabilmekte ve başarısızlık riskinin hangi düzeyde daha fazla olduğu yorumu yapılabilmektedir.

Olayın ortaya çıkması seyrek olduğunda, risk faktörlerinin etkileri zayıf olduğunda ve izleme periyodu kısa olduğunda lojistik ve Cox regresyon modellerinden elde edilen parametre tahminlerinin benzer olduğu, ancak daha uzun izleme süreleri için iki modelden elde edilen regresyon katsayılarının farklılık gösterdiği ve Cox regresyondan elde edilen katsayıların standart hatalarının azaldığı söylenebilmektedir.

Uygulamada, mide kanseri üzerinde etkili olduğu düşünülen yaş, cinsiyet, ağırlık kaybı, anemi, tümörün midedeki lokalizasyonu, lenf nodu diseksiyonunun genişliği, hastalığın evresi, tümörün mide duvarındaki invazyon derinliği, tümörün venöz, lenfatik ve perinöral invazyon yapması, adjuvant kemoterapi değişkenleri kullanılmıştır.

Başarısızlık üzerinde etkili olan değişkenleri belirlemek için yapılan tek değişkenli lojistik ve Cox regresyon modelleri sonucunda KEMOTERAPİ değişkeni dışında aynı değişkenlerin (DİSEKSİYON, LOKAL, DERİNLİK, EVRE, VENÖZ ve LENFATİK) ve değişken düzeylerinin önemli olarak bulunduğu, ancak Cox regresyondan elde edilen katsayıların standart hatalarının daha küçük olarak elde edildiği görülmüştür.

Lojistik ve Cox regresyonda tüm değişkenler birlikte ele alınarak adımsal seçim yöntemi uygulandığında her iki modelin son adımında aynı değişkenlerin (DİSEKSİYON, LOKAL, EVRE, VENÖZ ve KEMOTERAPİ) olduğu modele ulaşıldığı ve bu değişkenlerin önemli olarak bulunduğu görülmüştür. Sadece lojistik regresyonda LOKAL değişkeninin ALT ÜÇLÜK düzeyi anlamlı bulunurken Cox regresyonda bu düzey anlamlı olarak bulunmamıştır. Ayrıca Cox regresyondan elde edilen katsayıların standart hataları lojistik regresyondan elde edilen katsayıların standart hatalarına göre daha küçük elde edilmiştir.

## KAYNAKLAR

- CHEVRET, S. (2001), *Logistic or Cox Model to Identify Risk Factors of Nosocomial Infection: Still a Controversial Issue*, Intensive Care Med., 27, 1559-1560.
- COX, D.R. (1972), *Regression Models and Life Tables*, Journal of Royal Statistical Society, Series B, 34, 187-202.
- COX, D.R. and OAKES, D. (1984), *Analysis of Survival Data*, Chapman and Hall.

- ESTÉVEZ, J.I., CONCHA, D.M., MOLINA, C.D., CALLES, J.M., CASTILLO, A.S. AND NAVAJAS, R.F.C. (2001), *Comparison of Different Methodological Approaches to Identify Risk Factors of Nosocomial Infection in Intensive Care Units*, *Intensive Care Med.*, 27, 1254-1262.
- GREEN, M.S. and SYMONS, M.J. (1983), *A Comparison of the Logistic Risk Function and the Proportional Hazards Model in Prospective Epidemiologic Studies*, *J. Chron. Dis.*, 36, 10, 715-724.
- HAUCK, W.W. (1985), *A Comparison of the Logistic Risk Function and the Proportional Hazards Model in Prospective Epidemiologic Studies*, *J. Chron. Dis.*, 38, 1, 125-126.
- HOSMER, D.W. and LEMESHOW, S. (1999), *Applied Survival Analysis: Regression Modelling of Time to Event Data*, John Wiley and Sons.
- INGRAM, D.D. and KLEINMAN, J.C. (1989), *Empirical Comparisons of Proportional Hazards and Logistic Regression Models*, *Statistics in Medicine*, 8, 525-538.
- KALBFLEISCH, J. D. and PRENTICE, R.L. (1980), *The Statistical Analysis of Failure Time Data*, John Wiley and Sons.
- LAWLESS, J.F. (1982), *Statistical Models and Methods for Lifetime Data*, John Wiley and Sons.
- SCHEMPER, M. (1992), *Cox Analysis of Survival Data with Nonproportional Hazards Functions*, *The Statistician*, 41, 455-465.
- ÖZDAMAR, K. (1997), *Paket Programlar ile İstatistiksel Veri Analizi I*, Anadolu Üniversitesi.

## Review of Logistic and Cox Regression Models

### ABSTRACT

*In this study, Logistic regression and Cox regression models used to identify risk factors of any event in epidemiologic studies are investigated. These models and mathematical relationship and comparisons between the them are explained. An application is carried out by using gastric cancer data.*

**Key Words:** *Logistic Regression, Cox Regression, Survival Analysis*



## Üniversitelerarası Kurul Yabancı Dil Sınavı(ÜDS) Sonuçlarının Loglineer Analiz İle İncelenmesi

Dilek ALTAŞ\*

Esen ZEREN YILDIRIM\*\*

### ÖZET

*Bu çalışmada, ÜDS sonuçları dört yönlü bir çapraz tablo üzerinde sınıflandırılarak, kategorik veri analizine imkan sağlayan loglineer analizle incelenmiştir. Analiz kapsamında aralarındaki ilişki yapısı incelenen faktörler, adayın dili, ilgili olduğu alan, sınava giriş amacı ve aldığı ÜDS puanıdır. Elde edilen loglineer model ve parametre tahminlerine dayalı bulgular uygulama kısmında sunulmuştur.*

*Anahtar Kelimeler: Çok yönlü tablolar, Loglineer modeller, ÜDS sonuçları*

### 1. GİRİŞ

ÜDS, Üniversitelerarası Kurul'un 28.07.2000 tarihli toplantısında alınan kararla, Öğrenci Seçme ve Yerleştirme Merkezi(ÖSYM) tarafından yılda iki kez Mart ve Aralık aylarında olmak üzere uygulanan bir yabancı dil sınavıdır. Lisansüstü eğitim görmek isteyenler ile doçent adaylarının başvurabildiği bu sınavda, ÖSYM'nin belirlediği ağırlıklarda kelime bilgisi, dil bilgisi, okuduğunu anlama, Türkçe'den yabancı dile, yabancı dilden Türkçe'ye çeviri soruları yer almaktadır.

ÜDS'ye başvurular, Fen Bilimleri, Sosyal Bilimler ve Sağlık Bilimleri olmak üzere üç alanda; İngilizce, Almanca ve Fransızca olmak üzere üç dilde yapılabilmektedir.

Üniversitelerarası Kurul kararı gereği, ÜDS'den doçent adaylarının en az 65, doktora ve sanatta yeterlik öğrencilerinin en az 50 puan alması gerekmektedir.

Bu çalışmada, ÜDS sonuçları dört yönlü bir çapraz tablo üzerinde sınıflandırılarak, kategorik veri analizine imkan sağlayan loglineer analizle incelenmiştir. Analiz kapsamında aralarındaki ilişki yapısı incelenen faktörler, adayın dili(Almanca, Fransızca ve İngilizce), adayın ilgili olduğu alan(Fen Bilimleri, Sağlık Bilimleri, Sosyal Bilimler), sınava giriş amacı(Yüksek Lisans adayı, Doktora adayı, Doçent adayı) ve aldığı ÜDS puanı(50'nin altı, 50 ile 65 arası, 65 ve üstü)'dir.

\* Yrd.Doç. Dr., Marmara Üniversitesi, İ.İ.B.F., Ekonometri Bölümü, e-posta: [dilekaltas@marmara.edu.tr](mailto:dilekaltas@marmara.edu.tr)

\*\* Araş. Gör., Marmara Üniversitesi, İ.İ.B.F., Ekonometri Bölümü, e-posta: [eyildirim@marmara.edu.tr](mailto:eyildirim@marmara.edu.tr)  
(Haberleşme adresi)



## 2. GENELLEŞTİRİLMİŞ LİNEER MODELLER

İstatistiksel incelemelerin çoğunda, verilere en iyi uyum sağlayan modeli bulmak, temel amaç olarak benimsenmektedir. Bulunan modelin yapısal biçimi ve parametre tahminleri, değişkenler arasındaki ilişki yapısı hakkında bilgiler vermektedir.

Bağımlı değişkenin(Y), açıklayıcı değişkenlerin ( $X_j$ ) bir doğrusal kombinasyonu ile açıklandığı modellere *genelleştirilmiş lineer modeller* denir (Nelder, J.A. and Wedderburn, R.W.M., 1972).

$$Y = \alpha + \sum_{j=1}^k \beta_j X_j = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Genelleştirilmiş lineer modeller, üç bileşenden oluşur(Agresti, A., 1996):

- i) Y bağımlı değişkenini tanımlayan ve Y'nin bir olasılık fonksiyonuna sahip olduğunu varsayan *tesadüfi bileşen*
- ii) Modelin tahmincisi olarak kullanılan  $X_j$  açıklayıcı değişkenlerini ifade eden *sistemik bileşen*
- iii) Tesadüfi bileşenin beklenen değeri ile sistemik bileşen arasındaki fonksiyonel ilişkiyi gösteren *bağlantı bileşeni*.

Genelleştirilmiş lineer modellerde, n büyüklüğünde bir örneklem için  $y_i$  gözlemlerinin birbirinden bağımsız ve bir olasılık yoğunluk fonksiyonuna sahip olduğu varsayılır. Her bir  $y_i$  gözlemi iki şıklı bir sonuca sahip ise, Y'nin bir binom dağılımına uyduğu düşünülür.  $y_i$  değerleri çok sonuçlu ve farklı başarı olasılıkların sahip ise, multinomial dağılımın olasılık yoğunluk fonksiyonu kullanılır. Çapraz sınıflandırma tablosundaki göze frekanslarında olduğu gibi  $y_i$ , negatif olmayan değerler alıyorsa, verilerin Poisson dağılımına uyduğu düşünülür(Agresti, A. 1996).

Y'nin beklenen değeri, olasılık dağılımının ortalamasıdır,  $E(Y) = \mu$ . Tesadüfi ve sistemik bileşen arasındaki bağlantı fonksiyonu, doğrudan  $\mu$ 'yü ya da  $\mu$ 'nün monoton bir fonksiyonu olan  $g(\mu)$ 'yü ifade eder.

Normal dağılıma uymayan veriler analiz edilirken geleneksel yöntem, çeşitli dönüştürmelerle normal dağılıma yaklaştırmayı sağlamak ve klasik en küçük kareler yöntemini kullanmaktır. Fakat, bazı veri setleri için normale yaklaştırma sağlanamayabilir ya da bulunan modelin verilere uyumu zayıf olabilir. Genelleştirilmiş lineer modellerle birlikte normal dağılıma yaklaştırma için dönüşüm yapma zorunluluğu ortadan kalkmıştır. Çünkü, bu tür modellerde, dağılım seçimi normallikle sınırlı olmadığından maksimum benzerlik yöntemi kullanılabilir ve  $\mu$ 'nün değişik fonksiyonları ile çalışılabilmektedir.

### 3. LOGLINEER MODELLER

Genelleştirilmiş lineer modellerde,  $g(\mu)=\log(\mu)$  fonksiyonu sıkça kullanılan bir bağlantı fonksiyonudur. Poisson dağılımlı veriler için geliştirilen ve  $\log(\mu)$  bağlantı fonksiyonu ile ifade edilen lineer modellere *loglineer model* adı verilir(Vermunt, J.K., 1997).

$$\log(\mu) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

İki yönlü çapraz sınıflandırma tabloları için değişkenler arasındaki ilişki yapısı, ki-kare analizi ile incelenebilmektedir. Fakat, değişken sayısının üç ya da daha fazla olması durumunda ki-kare analizi uygulanamamaktadır. Ki-karenin yetersiz kaldığı bu tür durumlarda çok yönlü tablolar, loglineer analiz ile incelenebilmektedir.

Fischer'in öne sürdüğü fikirlerden yola çıkarak, Birch(1963) tarafından ortaya atılan loglineer analiz, Goodman, Haberman ve Andersen'in katkılarıyla geliştirilmiştir. 1970'lerde sosyal araştırmacılar tarafından sıkça kullanılmaya başlanan bu analiz, sadece ilişki yapısını çözmekle kalmayıp verilere uyumlu bir model vermesi yönüyle de kategorik veri analizinde söz sahibi olmuştur(Andersen, E.B., 1980).

X, Y ve Z şeklindeki üç değişken için, IxJxK boyutlu bir çapraz sınıflandırma tablosu olsun. Böyle bir üç yönlü tablo için loglineer model aşağıda verilmiştir.

$$\ln f_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$$

$$i=1, 2, \dots, I \quad j=1, 2, \dots, J \quad k=1, 2, \dots, K$$

En yüksek dereceli parametre dikkate alınarak, bu model  $\{XYZ\}$  modeli biçiminde ifade edilir. Model parametrelerinin yorumu en yüksek dereceli terime ( $\lambda_{ijk}^{XYZ}$ ) yönelik olarak yapılır.

Modeldeki tek değişken parametreleri( $\lambda_i^X, \lambda_j^Y, \lambda_k^Z$ ), değişkenlerin ana etkilerini verir. İkinci dereceden etkileşim terimleri( $\lambda_{ij}^{XY}, \lambda_{ik}^{XZ}, \lambda_{jk}^{YZ}$ ), ikişerli olarak değişkenler arasındaki kısmi ilişkileri gösterir. Modeldeki  $\lambda_{ijk}^{XYZ}$  parametresi, üç değişken arasındaki etkileşimi ifade eder.  $\lambda$  genel etki terimi ise, loglineer parametrelerin tüm kategorileri üzerinden sıfır toplamı olmalarını sağlar(Vermunt, J.K., 1997).

Tüm ana etki ve etkileşim terimlerini içeren bu tür modellere, *doymuş(saturated) loglineer model* adı verilir. Doymuş modeldeki parametrelerden bir ya da birkaç tanesinin hariç tutulduğu modellere *doymamış(unsaturated) loglineer model* denir. Doymamış modeller bazı parametreler üzerine konan kısıtlarla, daha sade ve kolay yorumlanabilir bir yapıya sahip olduklarından doymuş modellere tercih edilir(Vermunt, J.K., 1997).

Verilere en iyi uyum sağlayan doymamış model araştırılırken, modeldeki *hiyerarşik yapıya* dikkat edilmelidir. Hiyerarşik yapı, yüksek dereceli bir terimin



modelde bulunduğu hallerde, daha düşük dereceli tüm terimlerin de modelde yer aldığı türde ilişkileri ifade eder. Hiyerarşik bir loglineer modelde, bir  $\lambda$  parametresi sıfıra eşit ise, daha yüksek dereceli tüm  $\lambda$  parametreleri de sıfıra eşit olur. Aynı zamanda, bir  $\lambda$  parametresi sıfırdan farklı ise, daha düşük dereceli tüm parametreler de sıfırdan farklıdır (Le, C. T., 1998).

Loglineer modellerin çözümünde kullanılan yöntemlerin ve bilgisayar programlarının hemen hepsi, modelin hiyerarşik olduğu varsayımına dayanır. Çünkü, çoğu uygulamada bir etkileşim teriminin sıfıra eşit olması durumunda, daha yüksek dereceli etkileşim terimlerini yorumlamak anlamlı olmamaktadır.

Üç yönlü bir çapraz sınıflandırma tablosu için tahmin edilen model, üçüncü dereceden etkileşim parametresini içermezse,  $\{XY, XZ, YZ\}$  şeklinde gösterilen homojen ilişki modeli elde edilir.

$$\ln f_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$$

İkili etkileşim parametrelerinin bir ya da bir kaçının modelden çıkarılması da mümkündür. Hiç bir etkileşim parametresi içermeyen  $\{X, Y, Z\}$  bağımsızlık modeli ise, değişken çiftleri arasında ilişki olmadığını gösterir (Agresti, A. 1996).

$$\ln f_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$$

Modeldeki parametre sayısının çapraz sınıflandırma tablosundaki gözeye sayısından fazla olması halinde, serbestlik derecesi negatif çıkamayacağından tahmin edilecek parametre sayısını azaltmak gerekir. Bu amaçla bazı kodlama yöntemlerine başvurulur (Le, C. T., 1998).

Sözkonusu kodlama yöntemlerinden biri, parametrelerin etkilerini, her bir değişkenin belirli bir kategorisindeki etkilerden sapmalar şeklinde ifade etmektir. Regresyon analizinde de kullanılan bu yöntem *dummy kodlaması* denir. Genellikle ilk satır ve ilk sütuna ya da son satır ve son sütuna ilişkin parametre sıfıra eşitlenir. Parametrelerin etkileri, bu gözeye göre yorumlanacağı için seçilen gözeye *referans gözesi* denir.

Varyans analizinde de kullanılan bir diğer kodlama yöntemi ise, her bir etkiyi ortalama etkiden sapmalar şeklinde tanımlayarak,  $\lambda$ 'ların toplamını sıfır kabul etmektir. Bu tür kodlamaya ise, *etki kodlaması* adı verilir.

Her iki kodlama yöntemi ile de aynı beklenen frekans değerlerine ulaşılmakla birlikte, hesaplanan parametre tahminleri farklı çıkmaktadır. Bu nedenle parametreler yorumlanırken, kullanılan kodlama yöntemi dikkate alınmalıdır (Long, J.S., 1982).

Loglineer modelin parametre tahminlerinin bulunmasında, en sık başvurulan yöntem, maksimum benzerlik (en çok olabilirlik) yöntemidir. Maksimum benzerlik yöntemi, ağırlıklı en küçük kareler, minimum ki-kare, minimum diskriminant bilgisi



gibi alternatifleri arasında, olasılık dağılımı bilinen tesadüfi değişkenler için güçlü bir tahmin yöntemi olması nedeniyle tercih edilmektedir(Vermunt, J.K., 1997).

Farklı loglineer modellerin karşılaştırılması yoluyla, değişkenler arasında ilişkileri ifade eden parametrelerin anlamlılık testi yapılabilir. Bu amaçla, karşılaştırılan modellerden birinde test edilecek parametre/parametreler yer alırken, diğerinde bu parametre/parametrelere yer verilmez. Söz konusu parametrelerin sıfıra eşitliğini ifade eden  $H_0$  yokluk hipotezinin testi için, tam bölünebilme özelliğine sahip olan  $G^2$  benzerlik oran istatistiğinden yararlanır.

$$G^2 = -2 \sum n_{ij} \ln \left( \frac{n_{ij}}{f_{ij}} \right)$$

İki modelin  $G^2$  değerleri arasındaki fark, iki modelin serbestlik derecelerinin farkına karşı test edilir. Kuyruk olasılığının küçük çıkması,  $H_0$ 'ın reddedileceğine, yani incelenen ilişkinin anlamlı olduğuna işaret eder. Kuyruk olasılığının büyük çıkması ise, iki modelin uyum iyiliği yönünden farklılık göstermediği anlamına gelir. Bu durumda, test edilen ilişki terimlerinin modele katkısı olmayacağından, bu terimleri içermeyen daha az parametrelili model tercih edilir(Tests of Independence... [www document]).

Çok yönlü tablolarda test edilecek model sayısı oldukça fazla olacağından, hipotezlerin sistematik bir yöntemle test edilmesi gerekmektedir. Bu nedenle, hipotezleri anlamlı bir yorum çıkaracak şekilde gruplandırarak test etmek en akılcı yoldur. Goodman bağımsızlık, koşullu bağımsızlık gibi hipotezlerin yorumlanabileceğini, diğerlerinin yorumlanabilir olma özelliği taşımadığını göstermiştir. Haberman ve Andersen tarafından geliştirilen bu sisteme göre, üç yönlü tablolar için test edilmesi gereken başlıca hipotezler aşağıda verilmiştir(Andersen, E.B., 1980).

$$\begin{aligned} H_{XYZ} : \lambda_{ijk}^{XYZ} &= 0 \\ H_{XY} : \lambda_{ijk}^{XYZ} &= 0, \lambda_{ij}^{XY} = 0 \\ H_{XZ} : \lambda_{ijk}^{XYZ} &= 0, \lambda_{ik}^{XZ} = 0 \\ H_{YZ} : \lambda_{ijk}^{XYZ} &= 0, \lambda_{jk}^{YZ} = 0 \\ H_X : \lambda_{ijk}^{XYZ} &= 0, \lambda_{ij}^{XY} = \lambda_{ik}^{XZ} = \lambda_{jk}^{YZ} = 0, \lambda_i^X = 0 \\ H_Y : \lambda_{ijk}^{XYZ} &= 0, \lambda_{ij}^{XY} = \lambda_{ik}^{XZ} = \lambda_{jk}^{YZ} = 0, \lambda_j^Y = 0 \\ H_Z : \lambda_{ijk}^{XYZ} &= 0, \lambda_{ij}^{XY} = \lambda_{ik}^{XZ} = \lambda_{jk}^{YZ} = 0, \lambda_k^Z = 0 \\ H_0 : \lambda_{ijk}^{XYZ} &= 0, \lambda_{ij}^{XY} = \lambda_{ik}^{XZ} = \lambda_{jk}^{YZ} = 0, \lambda_i^X = \lambda_j^Y = \lambda_k^Z = 0 \end{aligned}$$

Hipotezlerin testi, en yüksek dereceli parametrelerin sıfıra eşitliğini ifade eden ilk hipotezle başlar. Bir sonraki hipoteze geçilebilmesi için o aşamaya kadarki tüm hipotezlerin kabul edilmiş olması gerekir. Arka arkaya gelen iki hipotezden biri kabul edilirken, diğeri red edilirse verilere en iyi uyan model, bu iki hipotezin arasında yer alır. Bu durumda, iki model arasındaki fark parametreleri teker teker test edilerek verilere en uygun model elde edilir.

Üç yönlü tablolar için ifade edilen loglineer modeller ve parametre yorumları, daha büyük boyutlu tablolar için genişletilebilir.

#### 4. UYGULAMA

Bu çalışmada 2002 Mart dönemine ilişkin ÜDS sonuçları dört yönlü bir çapraz sınıflandırma tablosu şeklinde düzenlenerek loglineer analizle incelenmiştir. Aralarındaki etkileşim yapısı incelenen dört kategorik değişken, *dil, alan, amaç ve ÜDS puanı* değişkenleridir. Dil değişkeni Almanca, Fransızca ve İngilizce şeklinde üç kategoriye ayrılmaktadır. Alan değişkeni, ÜDS'ye giren adayların ilgili oldukları alanları gösterecek şekilde, Fen Bilimleri, Sağlık Bilimleri ve Sosyal Bilimler olmak üzere üç kategoriye sahiptir. Amaç değişkeni, sınava giren kişilerin Yüksek Lisans, Doktora ya da Doçent adayı olma durumlarına göre üç kategoriye ayrılmaktadır. ÜDS puanı değişkeni ise, 50'nin altı, 50 ile 65 arası, 65 ve üstü şeklinde üç kategoriden oluşmaktadır.

2002 Mart döneminde ÜDS'ye giren 21336 kişinin 5406(%25)'si doçent adayı, 7894(%37)'ü doktora adayı, 8036(%38)'i yüksek lisans adaydır. Bu kişiler arasında 65 ve daha fazla puan alan kişi sayısı 2736(%13), 50 ile 65 arasında puan alan kişi sayısı 4679(%22), 50'nin altında puan alan kişi sayısı 13921(%65)'dir. Adayların 20460(%96)'ının dili İngilizce, 336(%1,5)'sının Fransızca, 540(%2,5)'inin Almanca'dır. Fen Bilimleri alanından sınava katılan kişi sayısı 8362(%39), Sağlık Bilimleri alanından sınava katılan kişi sayısı 4751(%22), Sosyal Bilimler alanından giren kişi sayısı 8223(%39)'dür.

Elde edilen dört yönlü çapraz sınıflandırma tablosu aşağıda verilmiştir.

**Tablo 1.** ÜDS 2002 Mart Dönemine İlişkin Veriler

DİL		İngilizce			Fransızca			Almanca		
AMAÇ	ALAN	Fen B.	Sağ. B.	Sos. B.	Fen B.	Sağ.B.	Sos. B.	Fen B.	Sağ. B.	Sos. B.
	PUAN									
Doçent	≥65	180	577	163	2	8	4	3	21	11
	50-65	557	808	331	4	4	27	12	9	28
	<50	904	736	826	13	8	84	22	10	54
Doktora	≥65	312	292	329	2	6	16	2	11	8
	50-65	616	503	453	1	2	17	9	12	22
	<50	2097	890	2072	10	6	69	23	10	104
Yüksek L.	≥65	241	122	383	1	2	13	9	4	14
	50-65	529	203	493	3	1	9	11	3	12
	<50	2770	498	2575	3	2	19	26	3	87

Verilerin analizinde SPSS for Windows 10.0 paket programı kullanılmıştır. Loglineer analizin ilk aşamasında ana etki terimleri ile ikinci, üçüncü ve dördüncü dereceden etkileşim terimlerinin anlamlılığı incelenmiştir.

Tests that K-way effects are zero



K	DF	L.R.Chisq	Prob	Pearson Chisq	Prob	Iteration
1	8	49864,178	,0000	85675,836	,0000	0
2	24	9809,303	,0000	4763,067	,0000	0
3	32	170,702	,0000	174,071	,0000	0
4	16	17,058	,3818	17,356	,3629	0

Yukarıdaki tablodan anlaşılacağı gibi dördüncü dereceden etkileşim parametreleri, 0.3629 olasılıkla istatistiksel olarak anlamsız bulunmuştur. Bu durumda, kurulacak loglineer modelin en fazla üçüncü dereceden etkileşim terimlerini içeren doymamış bir model olduğu anlaşılmaktadır.

Modelde yer alacak parametreleri tespit etmek üzere yapılan kısmi ilişki testi sonuçları aşağıda verilmiştir.

\*\*\*\*\* HIERARCHICAL LOG LINEAR \*\*\*\*\*

Tests of PARTIAL associations.

Effect Name	DF	Partial Chisq	Prob	Iter
AMAÇ*UDSPUAN*DİL	8	41,332	,0000	3
AMAÇ*UDSPUAN*ALAN	8	73,533	,0000	4
AMAÇ*DİL*ALAN	8	14,644	,0665	4
UDSPUAN*DİL*ALAN	8	32,340	,0001	3
AMAÇ*UDSPUAN	4	514,063	,0000	4
AMAÇ*DİL	4	169,034	,0000	4
UDSPUAN*DİL	4	5,581	,2327	4
AMAÇ*ALAN	4	1315,525	,0000	4
UDSPUAN*ALAN	4	688,142	,0000	4
DİL*ALAN	4	413,434	,0000	4
AMAÇ	2	644,705	,0000	2
UDSPUAN	2	9553,587	,0000	2
DİL	2	38404,240	,0000	2
ALAN	2	1261,636	,0000	2

Yukarıdaki sonuçlar incelendiğinde, amaç-dil-alan ile üdspuan-dil parametrelerinin istatistiksel olarak anlamsız olduğu görülmektedir. Bu durumda, amaç-dil-alan üçüncü dereceden etkileşim parametreleri kurulacak doymamış loglineer modelde yer almayacaktır. Fakat, diğer üçüncü dereceden etkileşim terimlerinin istatistiksel olarak anlamlı bulunması nedeniyle, modelin hiyerarşik yapısı gereği tüm ikinci dereceden etkileşim parametreleri ile ana etki parametreleri modelde yer alacaktır.

Geriye doğru eleme yöntemi sonucunda bulunan en iyi model, yukarıdaki sonucu desteklemektedir. Eleme aşamaları aşağıda sunulmuştur.

\*\*\*\*\* HIERARCHICAL LOG LINEAR \*\*\*\*\*

Backward Elimination (p=,050) for DESIGN 1 with generating class

AMAÇ\*UDSPUAN\*DİL\*ALAN

Likelihood ratio chi square = ,0000 DF = 0 P = 1,000



If Deleted Simple Effect is	DF	L.R.Chisq	Change	Prob.	Iter
AMAÇ*UDSPUAN*DİL*ALAN	16		17,058	,3818	3

Step 1: The best model has generating class

AMAÇ\*UDSPUAN\*DİL  
AMAÇ\*UDSPUAN\*ALAN  
AMAÇ\*DİL\*ALAN  
UDSPUAN\*DİL\*ALAN

Likelihood ratio chi square = 17,05808 DF = 16 P = ,382

If Deleted Simple Effect is	DF	L.R.Chisq	Change	Prob.	Iter
AMAÇ*UDSPUAN*DİL	8	41,332	,0000	3	
AMAÇ*UDSPUAN*ALAN	8	73,533	,0000	4	
AMAÇ*DİL*ALAN	8	14,644	,0665	4	
UDSPUAN*DİL*ALAN	8	32,340	,0001	3	

Step 2: The best model has generating class

AMAÇ\*UDSPUAN\*DİL  
AMAÇ\*UDSPUAN\*ALAN  
UDSPUAN\*DİL\*ALAN

Likelihood ratio chi square = 31,70187 DF = 24 P = ,135

If Deleted Simple Effect is	DF	L.R.Chisq	Change	Prob.	Iter
AMAÇ*UDSPUAN*DİL	8	46,621	,0000	4	
AMAÇ*UDSPUAN*ALAN	8	71,223	,0000	5	
AMAÇ*DİL*ALAN	8	32,347	,0001	4	

Step 3: The final model has generating class

AMAÇ\*UDSPUAN\*DİL  
AMAÇ\*UDSPUAN\*ALAN  
UDSPUAN\*DİL\*ALAN

Likelihood ratio chi square = 31,70187 DF = 24 P = ,135

Görüldüğü gibi, ilk adımda dört yönlü etkileşim parametreleri istatistiksel olarak anlamsız bulunduğundan, üçüncü dereceden etkileşim parametrelerini içeren loglineer model en iyi model olarak değerlendirilmiştir. İkinci adımda, amaç-dil-alan üçüncü dereceden etkileşim parametreleri modelden çıkarılmış, diğer üçüncü dereceden etkileşim terimlerini kapsayan loglineer model 0,135 olasılıkla en iyi model olarak tespit edilmiştir. İncelemenin üçüncü ve son aşamasında elde edilen final modeli, ikinci adımda bulunan modelin aynısı olduğundan süreç durdurulmuştur.

Elde edilen final modelinde, bir sabit terim, 12 ana etki parametresi, 54 ikinci dereceden etkileşim terimi, 81 üçüncü dereceden etkileşim terimi olmak üzere toplam 148 parametre yer almaktadır. SPSS paket programında loglineer modelin parametre tahminleri, değişkenlerin son kategorisini referans göze kabul eden Dummy kodlamasına dayanmaktadır.

*Dil* değişkeni için, 1: Almanca, 2: Fransızca, 3: İngilizce; *Alan* değişkeni için, 1: Fen Bilimleri, 2: Sağlık Bilimleri, 3: Sosyal Bilimler; *Amaç* değişkeni için, 1: Yüksek Lisans adayı, 2: Doktora adayı, 3: Doçent adayı; *ÜDS puanı* değişkeni için, 1: 50'nin altı, 2: 50 ile 65 arası, 3: 65 ve üstü olmak üzere, 58 parametre için tahmin değerleri, standart hata katsayıları, Z test istatistiği değerleri aşağıda sunulmuştur. Dummy kodlaması gereğince sıfıra eşitlenen parametrelere yer verilmemiştir.

	Parametre	Tahmin	Standart Hata	Z değeri
1	$\lambda$	1,2459	0,2356	5,29
2	$\lambda$ (alan=1)	3,2182	0,2255	14,27
3	$\lambda$ (alan=2)	2,0000	0,2464	8,12
4	$\lambda$ (dil=1)	4,9666	0,2329	21,32
5	$\lambda$ (dil=2)	-1,7714	0,4004	-4,42
6	$\lambda$ (üdspuan=1)	0,3434	0,3714	0,92
7	$\lambda$ (üdspuan=2)	-0,5350	0,3817	-1,40
8	$\lambda$ (amaç=1)	1,1162	0,1564	7,14
9	$\lambda$ (amaç=2)	0,9414	0,1390	6,77
10	$\lambda$ (dil=1)*(alan=1)	-1,5775	0,2221	-7,10
11	$\lambda$ (dil=1)*(alan=2)	-0,2858	0,2434	-1,17
12	$\lambda$ (dil=2)*(alan=1)	0,3177	0,3434	0,93
13	$\lambda$ (dil=2)*(alan=2)	-0,3388	0,4005	-0,85
14	$\lambda$ (üdspuan=1)*(alan=1)	-1,9977	0,3595	-5,56
15	$\lambda$ (üdspuan=1)*(alan=2)	-1,8857	0,4200	-4,49
16	$\lambda$ (üdspuan=2)*(alan=1)	-1,0967	0,3442	-3,19
17	$\lambda$ (üdspuan=2)*(alan=2)	-0,7678	0,3763	-2,04
18	$\lambda$ (amaç=1)*(alan=1)	-1,5221	0,0697	-21,84
19	$\lambda$ (amaç=1)*(alan=2)	-1,4987	0,0690	-21,73
20	$\lambda$ (amaç=2)*(alan=1)	-0,7974	0,0628	-12,70
21	$\lambda$ (amaç=2)*(alan=2)	-0,8624	0,0626	-13,77
22	$\lambda$ (üdspuan=1)*(alan=1)	-1,7579	0,3675	-4,78
23	$\lambda$ (üdspuan=1)*(alan=2)	0,7896	0,6186	1,28
24	$\lambda$ (üdspuan=2)*(alan=1)	-0,3565	0,3760	-0,95
25	$\lambda$ (üdspuan=2)*(alan=2)	0,1437	0,6900	0,21
26	$\lambda$ (amaç=1)*(dil=1)	-0,7321	0,1468	-4,99
27	$\lambda$ (amaç=1)*(dil=2)	1,7944	0,2689	6,67
28	$\lambda$ (amaç=2)*(dil=1)	-0,3598	0,1285	-2,80
29	$\lambda$ (amaç=2)*(dil=2)	1,0989	0,2636	4,17
30	$\lambda$ (amaç=1)*(üdspuan=1)	0,4541	0,3337	1,36
31	$\lambda$ (amaç=1)*(üdspuan=2)	0,8424	0,3046	2,77
32	$\lambda$ (amaç=2)*(üdspuan=1)	-0,5103	0,3381	-1,51
34	$\lambda$ (amaç=2)*(üdspuan=2)	0,3670	0,2982	1,23
35	$\lambda$ (üdspuan=1)*(dil=1)*(alan=1)	1,5010	0,3508	4,28
36	$\lambda$ (üdspuan=1)*(dil=1)*(alan=2)	0,8709	0,4116	02,12
37	$\lambda$ (üdspuan=1)*(dil=2)*(alan=1)	0,3965	0,5456	0,73
38	$\lambda$ (üdspuan=1)*(dil=2)*(alan=2)	-0,0069	0,7357	-0,009
39	$\lambda$ (üdspuan=2)*(dil=1)*(alan=1)	0,3208	0,3343	0,96
40	$\lambda$ (üdspuan=2)*(dil=1)*(alan=2)	0,0144	0,3676	0,04

41	$\lambda$	(üdspuan=2)*(dil=2)*(alan=1)	0,9019	0,5870	1,54
42	$\lambda$	(üdspuan=2)*(dil=2)*(alan=2)	0,2906	0,7105	0,41
43	$\lambda$	(amaç=1)*(üdspuan=1)*(dil=1)	0,7174	0,3246	2,21
44	$\lambda$	(amaç=1)*(üdspuan=1)*(dil=2)	-1,8979	0,5561	-3,41
45	$\lambda$	(amaç=1)*(üdspuan=2)*(dil=1)	0,1392	0,2920	0,48
46	$\lambda$	(amaç=1)*(üdspuan=2)*(dil=2)	-1,2122	0,4923	-2,46
47	$\lambda$	(amaç=2)*(üdspuan=1)*(dil=1)	0,8186	0,3257	2,51
48	$\lambda$	(amaç=2)*(üdspuan=1)*(dil=2)	-0,3120	0,5133	-0,61
49	$\lambda$	(amaç=2)*(üdspuan=2)*(dil=1)	-0,0391	0,2842	-0,14
50	$\lambda$	(amaç=2)*(üdspuan=2)*(dil=2)	-1,0673	0,5094	-2,10
51	$\lambda$	(amaç=1)*(üdspuan=1)*(alan=1)	-0,8658	0,1497	-5,78
52	$\lambda$	(amaç=1)*(üdspuan=1)*(alan=2)	-0,3617	0,1538	-2,35
53	$\lambda$	(amaç=1)*(üdspuan=2)*(alan=1)	-0,2031	0,1253	-1,62
54	$\lambda$	(amaç=1)*(üdspuan=2)*(alan=2)	0,1712	0,1200	1,43
55	$\lambda$	(amaç=2)*(üdspuan=1)*(alan=1)	-0,2393	0,1426	-1,68
56	$\lambda$	(amaç=2)*(üdspuan=1)*(alan=2)	0,2086	0,1489	1,40
57	$\lambda$	(amaç=2)*(üdspuan=2)*(alan=1)	-0,1835	0,1217	-1,51
58	$\lambda$	(amaç=2)*(üdspuan=2)*(alan=2)	0,0875	0,1189	0,74

## 5. SONUÇ

ÜDS sonuçlarının loglineer analizle incelenmesi sonucunda, adayın dili, sınava giriş amacı, başvurduğu alan ve aldığı ÜDS puanı arasında dörtlü etkileşim olmadığı gözlenmiştir. Elde edilen loglineer modelin doymuş modelle karşılaştırıldığında istatistiksel olarak anlamsız olduğunu ifade eden  $H_0$  yokluk hipotezi, 0.135 olasılıkla reddedilememiştir.

Üçüncü dereceden etkileşim terimleri incelendiğinde, amaç-ÜDS puanı-dil, amaç-ÜDS puanı-alan ve ÜDS puanı-dil-alan değişkenleri arasındaki üçlü etkileşimler istatistiksel olarak anlamlı bulunmuştur. ÜDS puanı faktörünün sabit tutulması durumunda, amaç-dil-alan değişkenleri arasındaki etkileşimi veren parametreler ise istatistiksel olarak anlamlı değildir.

İkinci dereceden etkileşim terimleri incelendiğinde, adayın amacı ile aldığı ÜDS puanı, amacı ile dili, amacı ile başvurduğu alan, dili ile başvurduğu alan, ÜDS puanı ile başvurduğu alan arasında anlamlı ilişkiler tespit edilmiştir. Bununla birlikte, adayın aldığı ÜDS puanı ile dili arasında anlamlı bir ilişkiye rastlanmamıştır.

Değişkenlerin ana etki parametreleri incelendiğinde, dört değişken arasındaki etkileşim yapısını ortaya koymada her birinin ayrı ayrı önemli katkıları olduğu tespit edilmiştir.



## KAYNAKLAR

- AGRESTİ, A. (1996), *An introduction to categorical data analysis*. New York: John Wiley & Sons. Sy. 16-70, 145-173
- ANDERSEN, E.B. (1980), *Discrete statistical models with social applications*. Amsterdam: North-Holland Publishing Company. Sy. 162-191.
- LE, C. T. (1998), *Applied categorical data analysis*. New York: John Wiley & Sons. Sy. 15-102
- LONG, J.S. (1982), *Estimable Functions in Loglinear Models*. Sociological Methods & Research, Vol. 12
- NELDER, J.A. and Wedderburn, R.W.M. (1972), *Generalized Linear Models*. Journal of the Royal Statistical Society, Series A, Vol. 135.
- Tests of Independence Using Multiway Contingency Tables in SPSS. (12 October 1999) [www document] Erişim: 31.07.2002
- VERMUNT, J.K. (1997), *Log-Linear Models for Event Histories*. California: Sage Publications. Sy. 297-298.

## Evaluation of The Results of the UDS By Loglinear Models

### ABSTRACT

*In the study, the UDS results as being classified on a four dimensional cross classified table, examined by the loglinear analysis which gives a possibility for the categorical data analysis. In the content of this study the parameters that examined by the point of their relation structures are language, interesting branch of candidates, the aim for attending the exam and the taken UDS grade. The findings based on the obtained loglinear model and parameters are presented the application part.*

**Key Words:** *Multidimensional tables, Loglinear models, the results of the ÜDS*

## Çok Değişkenli Veri Kümeleri Üzerinde Tanımlı Aykırı Değer Belirleme Tekniklerinin Simülasyon Çalışması ile Karşılaştırılması

Gülşen KIRAL\*

Nedret BİLLOR\*\*

### ÖZET

*Temel bileşenler analizi ilişkisiz değişkenler kümesinin kurulması ve/ya boyut indirgenmesi amacı ile kullanılan istatistiksel bir tekniktir. Bazen tek başına bir analiz olarak kullanıldığı gibi bazen de başka analizler için veri hazırlama tekniği olarak kullanılmaktadır. Veri boyutunda indirgeme yaptığından özellikle yüksek boyutlu verilerin analiz edilmesinde tercih edilen bir tekniktir. Fakat klasik varyans ve kovaryans matrisine dayalı olarak hesaplandığından aykırı değerlerin varlığı durumunda sağlıklı sonuç vermemektedir. Bu nedenle aykırı değer olması olasılığına karşın dayanıklı temel bileşenler analiz teknikleri kullanımı önerilmektedir.*

*Bu çalışmada temel bileşenler analizi çerçevesinde tanımlanan dayanıklı tekniklerden, Dayanıklı Temel Bileşenler Analizi (DTBA) (Hubert ve ark.,2002) ve BACON Dayanıklı Temel Bileşenler Analizi (BDTBA) (Kıral ve Billor, 2001) ve Klasik Temel Bileşenler Analizinin (TBA) performansları simülasyon çalışması yapılarak karşılaştırılmıştır.*

**Anahtar Kelimeler:** *Klasik Temel Bileşenler Analizi, Dayanıklı Temel Bileşenler Analizi, BACON Algoritması, Aykırı Değer*

### 1. GİRİŞ

Bir veri kümesinde gözlemlerin çoğu tarafından önerilen modele uymayan gözlemlere aykırı değer denir. Aykırı değerlerin belirlenmesi kullanılan istatistiksel analizin doğru yorumlanması açısından önemlidir. Tek boyutlu veri kümesindeki aykırı değerler tek değişkenli analiz yöntemleriyle kolayca belirlenebilir, ancak çok boyutlu veri kümesinde işlemler zorlaşır. Çünkü bu tip veriler hesaplama zorluğu getirmektedir ve pek çok istatistiksel analiz ilişkili değişkenlerin varlığında sağlıklı sonuç vermemektedir. Aykırı değer belirlemeye ilişkin üç farklı yaklaşım vardır: Klasik aykırı değer belirleme yöntemleri, dayanıklı yöntemler ve birleştirilmiş (*combined*) yöntemler.

**Klasik aykırı değer belirleme yöntemleri** aykırı değerlere hassas klasik kestiricilere dayalı olup etkin değildirler. Masking (aykırı değer olan bir gözlemin normal gözlem gibi görülmesi) ve bulandırma (*swamping*) (normal olan bir gözlemin aykırı değer gibi görülmesi) hataları yanlış yorumlara sebep olabilirler.

**Dayanıklı yöntemler** dayanıklı kestiricilere dayalı olup maskeleyen ve bulandırma problemlerinden etkilenmezler, fakat hesaplanması ve uygulaması zor yöntemlerdir.

\*Çukurova Üniv. İ.İ.B.F. Ekonometri Böl. Balcalı/ADANA

\*\*Auburn Univ. Discrete and Statistical Science, Auburn, USA



**Birleştirilmiş yöntemler** ise dayanıklı ve klasik aykırı değer belirleme tekniklerinin birleşimi ile tanımlanmışlardır. Dayanıklı ve/ya klasik kestiricilere dayalıdır. Dayanıklı yöntemlere göre uygulamada daha pratik olup bir çoğu maskeleyme ve bulandırma problemlerinden etkilenmemektedirler.

Veri kümeleri içerisinde karşılaşılabileceğimiz en önemli problemlerden biri büyük veri kümeleri ile çalışıldığında ortaya çıkmaktadır. Çok sayıda değişkenle çalışmak değişkenlerin bağımsızlık varsayımını bozup, işlem yükünü arttıracak gibi yapılacak analiz sonuçlarının yorumlanmasını da güçleştirebilmektedir. TBA birbirleri ile ilişkisiz olan yeni bir veri kümesi oluşturarak veri boyutunu indirgediğinden yüksek boyutlu verilerin incelenmesinde tercih edilen bir tekniktir. Yöntem örneklem kovaryans matrisinin özvektörlerinin belirlenmesi ile yürütülür. Ancak örneklem kovaryans matrisi örneklem ortalamasına dayalı olduğundan yöntem aykırı değerlerden büyük ölçüde etkilenmektedir. Bu nedenle aykırı değerlerin varlığı olasılığına karşın dayanıklı temel bileşenler analizinin kullanılması tercih edilmektedir.

Dayanıklı temel bileşenler analizlerinde amaç aykırı değerlerden etkilenmeyen temel bileşenlerin belirlenmesidir. Bu amaçla işlemler iki farklı mantıkla yürütülebilir. Birincisinde klasik kovaryans matrisi dayanıklı kovaryans matrisi ile yer değiştirerek analiz gerçekleştirilebilir (örnek: Campbell, 1980 ve Croux ve Haesbroeck, 2000). Bu grupta sonuçlar dayanıklı olarak elde edilir ama ne yazık ki yüksek boyutlu verilerde hesaplama problemleri nedeniyle kullanışlı değildirler. İkincisinde ise analizi izdüşüm takibi (*projection pursuit*) yöntemini kullanarak gerçekleştirmektedirler (Örnek: Lie ve Chen, 1985, Croux ve Ruiz Gazen, 2000, Hubert ve ark., 2002). Bu yöntemler yansıtılmış veri üzerine dayanıklı yayılım ölçüsünü en büyük yapacak şekilde ardışık adımlarla yeni doğrultuları bulmaya çalışır.

Dayanıklı temel bileşenler ile ilgili ilk çalışma Campbell (1980) tarafından yapılmıştır. Campbell çalışmasında dayanıklı M-kestiricisi kullanarak aykırı değerlerden etkilenmeyen temel bileşenleri belirlemiştir. Yöntem temel bileşenler analizi içerisinde dayanıklı M-kestiricisine ait varyans-kovaryans matrisinin kullanımı ile tanımlanmıştır. Yöntemde amaç; aykırı değerlerin etkisini ortadan kaldıracak gerçek ağırlıkları bularak tüm veri kümesini temsil eden gerçek varyans-kovaryans matrisini elde etmektir. Ardından Lie ve Chen (1985); izdüşüm takibi yöntemine dayalı bir çözüm önerdiler. Lie ve Chen'nin amacı; en büyük dayanıklı ölçeklemeye sahip izdüşümü alınmış gözlemlerin doğrultusunu belirlemektir. Birbirini izleyen adımlarda her yeni doğrultu önceki tüm doğrultulara dik olacak şekilde belirlenmektedir. Yüksek boyutlu veri kümelerinde hatta ve hatta parametre sayısı gözlem sayısından büyükken de dahil olmak üzere iyi sonuç veren bir algoritmadır. Fakat hesaplama problemleri içermektedir. İzdüşüm takibi yöntemine dayalı yöntemlerde karşılık gelen etki fonksiyonunun sınırlandırılmamış olması yerel dayanıklılıkta eksikliğe sebep olmaktadır. Bunun yanında izdüşüm takibi yöntemine dayalı kestiricilerin nasıl hesaplanacağı açık değildir. Bu problemleri ortaya çıkaran Croux ve Ruiz-Gazen kısıtlamalar altında bir maksimizasyon probleminin çözümünü önerdiler. C-R algoritması adını verdikleri yöntem küçük boyutlu veri kümelerinde iyi çalışmasına rağmen büyük boyutlu veri kümeleri için hesaplama problemleri içermektedir. Daha sonra Hubert ve ark. (2002) C-R algoritmasını biraz daha geliştirerek, daha hızlı iki adımlı algoritma (*a faster two-step algorithm, DTBA*) adını verdikleri yeni bir yöntem sunmuşlardır. Yöntem sayısal olarak C-R algoritmasından daha karardır ve veride gözlemden çok değişken olması durumunda da sağlıklı sonuç vermektedir. Diğer taraftan Caroni (2000); Campbell'in (1980) yöntemi üzerine bir



simülasyon çalışması yapmıştır. Bu çalışmada  $x_i \sim N_p(\mu, \Sigma)$   $i=1, \dots, n$  sıfır hipotezi altında DTBA içinde gözlemlerin ağırlıklarına bağlı olarak kritik değerler belirlenmektedir. Bu kritik değerlerin belirlenmesi ile düşük ağırlığa sahip olan gözlemler yöntem tarafından doğru bir şekilde belirlenir. Bu nedenle yöntem aykırı değerler için bir formal test olarak düşünülebilir (Caroni, 2000).

Bu çalışmanın ikinci bölümde klasik temel bileşenler analizi ve BACON Dayanıklı Temel Bileşenler Analizi (BDTBA) anlatıldıktan sonra üçüncü bölümde BDTBA yönteminin etkinliği bir veri kümesi üzerinde gösterilecektir. Dördüncü bölümde Dayanıklı Temel Bileşenler Analizi (Hubert ve ark., 2002) ve BACON Dayanıklı Temel Bileşenler Analizi (Kıral ve Billor, 2001) ve Klasik Temel Bileşenler Analizi yöntemlerinin performansları simülasyon çalışmaları yapılarak karşılaştırılacaktır. Nihai tartışma ve sonuç ise beşinci bölümde verilecektir.

## 2. YÖNTEMLER

### 2.1. Temel Bileşenler Analizi

Temel bileşenler analizi; değişkenler arası bağımlılık yapısının yok edilmesi ve (veya) boyut indirgenmesi ya da başka analizler için veri hazırlanması amaçları ile kullanılır. Analizde, veriyi temsil eden  $X_{n \times p}$  matrisine uygun bir dönüşüm yapılarak,  $X$  uzayındaki problemler düzeltilmeye çalışılır. Dönüşüm sonucu birbirleri ile ilişkisiz sütunlardan oluşan bir veri matrisi elde edilmiş olur.

İncelemede  $X_{n \times p}$  matrisinin sütunlarının birimleri arasında uyuma söz konusu değilse, bu matris yerine onun standartlaştırılmış formu kullanılır.

Aslında temel bileşenler  $p$  tane  $X_1, X_2, \dots, X_p$  rasgele değişkenin özel doğrusal kombinasyonudur. Geometrik olarak, bu doğrusal kombinasyonlar koordinat eksenleri  $X_1, X_2, \dots, X_p$  ler olan orijinal sistemin döndürülmesiyle elde edilen yeni bir koordinat sistemini temsil eder.

$X_{n \times p}$  çok değişkenli veri kümesinin varyans-kovaryans matrisi  $V$  nin öz değerleri ( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ ) ve bu öz değerlere karşılık öz vektörleri  $u_i$  olmak üzere  $(\lambda_1, u_1), (\lambda_2, u_2), \dots, (\lambda_p, u_p)$  çiftleri için  $i$ . temel bileşen

$$y_i = u_i' X \quad i=1, 2, \dots, p$$

eşitliği yardımıyla hesaplanır. Örneğin 1. temel bileşen,  $Var(y_1) = u_1' V u_1 = \lambda_1$ , maksimum varyanslı doğrusal kombinasyondur. Bu şekilde temel bileşenler belirlenerek  $X_{n \times p}$  matrisini temsil eden yeni temel bileşenler matrisi  $Y_{n \times p}$  belirlenmiş olur.

Klasik temel bileşenler analizi; verilerde homojenliğin bozulması durumunda (*yani aykırı değerlerin varlığında*) sağlıklı sonuçlar vermemektedir. Bu durumda aykırı değerlere karşı dayanıklı olan kestiriciler kullanılarak analizin yapılması gerekmektedir.

### 2.2. BACON Algoritması

Bu yöntemde esas amaç; aykırı değerlerden arındırılmış olacak şekilde gözlemlerin hemen hemen yarısını içeren temel alt kümeyi bulmak, hemen ardından da temel alt küme ile uyumlu gözlemleri bu kümeye dahil etmektir. İşlem sonunda temel alt küme dışında kalan

gözlemler aykırı değer olarak belirlenirler. Temel alt küme dışında hiç gözlem kalmamışsa "veri kümesi aykırı değer içermemektedir" denir.

BACON yönteminde (Billor ve ark. 2000); gözlemlerin çok değişkenli eliptik dağılımdan geldiği varsayılarak Mahalanobis uzaklığından yararlanılmakta, kritik değer olarak da düzeltilmiş ki-kare değeri kullanılmaktadır.

BACON yöntemi gözlemlerin bloklanması nedeniyle hesaplamalar için etkin bir yöntemdir. Diğer yöntemlere göre bu yöntemdeki tekrarlamaya sayısı daha azdır. Tekrarlamaların her biri kovaryans matrisinin hesaplanması ve tersinin alınmasını gerektirir. Fakat tekrarlamaya sayısı  $n$  örneklem büyüklüğünün artması ile büyümeye ve hesaplanan  $n$  uzaklığın sıralanmasını gerektirmez.

### 2.3. BACON Dayanıklı Temel Bileşenler Analizi (BDTBA)

Dayanıklı kestiricilerle yapılan işlemler çoğu zaman için sağlıklı sonuç verirler ama bilindiği gibi yapılması gereken işlemler problemlidir ve zaman alıcıdır. Gözlem ve parametre sayılarının artması durumunda hesaplamalar iyice artmaktadır. Bunun yanında kullanılan veri kümesine ve istatistiğe bağlı olarak etkinliklerinde değişikliklerin olabilmesi ve sadece belli tipteki aykırı değerleri ortaya çıkarıyor olmaları da karşılaşılabilecek problemlerdir. O halde bu problemlerden etkilenmeyen daha hızlı işleyip sağlıklı sonuç veren bir yöntem gereksinim duyulmaktadır. Bu amaçla; bu çalışmada Billor ve ark. (2000) tarafından tanımlanan BACON algoritması kullanılarak dayanıklı temel bileşenlerin belirlenmesini sağlayan bir algoritma tanımlanmıştır (BDTBA) (Kıral ve Billor, 2001).

BDTBA yöntemi; DTBA (Hubert ve ark., 2002) ya da Campbell (1980) tarafından tanımlanan yöntem içinde kullanılan dayanıklı M-kestiricisi yerine BACON algoritmasından elde edilen klasik ortalama ve kovaryans matrisinin kullanılmasına dayalı olarak yürütülmektedir.

BDTBA yönteminde ana düşünce büyük veri kümelerinde etkinliği ispatlanmış BACON algoritması (Billor ve ark., 2000) kullanmak ve hemen ardından temel bileşenler analizini uygulamaktır. Böylece analizci aykırı değerlerden arındırılmış  $X$  veri matrisini tanımlayan en önemli bileşenleri belirleyebilir.

#### BDTBA Algoritması

**Adım 1:** Temel altküme; BACON algoritmasında tanımlı yaklaşımlardan biri kullanılarak  $m=cp$ , ( $c=4$  veya  $5$ ) elemanlı olacak şekilde belirlenir.

**Adım 2:** Temel alt kümedeki gözlemlerin ortalama ve varyans-kovaryans matrisleri sırasıyla,  $\bar{X}_b$  ve  $S_b$  olmak üzere

$$d_i(\bar{X}_b, S_b) = \sqrt{(x_i - \bar{X}_b)' S_b^{-1} (x_i - \bar{X}_b)} \quad i=1, 2, \dots, n$$

uzaklıkları hesaplanır.

**Adım 3:**  $d_i(\bar{X}_b, S_b) < C_{npr} \cdot \chi_{p,\alpha}^2$  olan gözlemlerle yeni temel alt küme belirlenir.  $\chi_{p,\alpha}^2$ ;  $p$  serbestlik dereceli,  $1-\alpha$  yüzdelikli ki-kare değeri,  $C_{npr} = C_{np} + C_{hr}$  olan bir düzeltme faktörü,  $r$ ; şu an ki



temel alt kümede bulunan eleman sayısı,  $C_{hr} = \max\{0, (h-r)/(h+r)\}$  ve  $C_{np} = 1 + \frac{p+1}{n-p} + \frac{1}{n-h-p}$

olarak tanımlıdır ( $h = \lfloor (n+p+1)/2 \rfloor$ ).

**Adım 4:** 2. ve 3. adımlar temel alt kümede değişme olmayana kadar tekrarlanır.

**Adım 5:** Son adımda elde edilen temel alt küme dışında kalan gözlemler aykırı değer olarak tanımlanır.

**Adım 6:** Aykırı değer olarak belirlenen gözlemler veri kümesinden atılarak indirgenmiş veri kümesi elde edilir ( $X_{(l)}$ ).

**Adım 7:**  $X_{(l)}$  matrisinin öz değer ve öz vektör çiftleri  $(\lambda_i, u_i)$ ;  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$  olacak şekilde hesaplanır.

**Adım 8:** İndirgenmiş matrisin varyans-kovaryans matrisine karşılık gelen özvektörler  $U=(u_1, u_2, \dots, u_p)$  olarak tanımlanmak üzere yeni temel bileşenler

$$y = X_{(l)}U = (y_1^*, y_2^*, \dots, y_p^*)$$

elde edilir.  $y_i^*$ ,  $y$  nin  $i$ . sütunudur.

### **Grafiksel Yöntem:**

Kıral ve Billor (2001) aykırı değerlerin görsel olarak belirlenmesine yardımcı olmak amacı ile birkaç grafiksel yöntem önermiştir. Grafiklerin oluşturulmasında takip edilecek adımlar:

**Adım 1:** BACON Dayanıklı Temel Bileşenler Uzaklığı (BDTBU) aşağıdaki adımları takip edilerek hesaplanır.

- BTBA'den elde edilen temel bileşenlerin ( $y$ ) ortalama ( $\bar{y}$ ) ve kovaryans matrisi ( $s_y$ ) hesaplanır.
- Orijinal veri matrisine ( $X$ ) ait temel bileşenler  $x_i^*$  hesaplanır.
- Aşağıdaki eşitliği kullanarak dayanıklı uzaklıklar hesaplanır.

$$BDTBU_i = d_i = (x_i^* - \bar{y}) \cdot (s_y)^{-1} \cdot (x_i^* - \bar{y})' \quad i = 1, 2, \dots, n$$

**Adım 2:** Birinci adımda elde edilen bilgiler kullanılarak aşağıdaki grafikleri çizilir

- BDTBU'nun küp köküne ait Kantil-Kantil (*Quantile-Quantile*) ( $Q-Q$ ) grafiği ya da
- Klasik Mahalanobis uzaklığı (MD) karşın BDTBU'nun serpilme grafiği ya da
- BDTBU na ait indeks grafiği

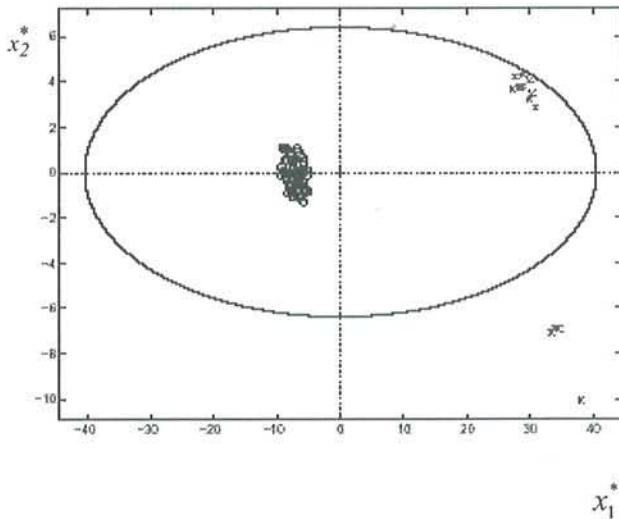
Birinci grafik verinin normal dağılımdan gelip gelmediği hakkında bilgi vermekte ve aykırı gözlemleri belirlemekte.

İkinci grafik BDTBA'ne dayalı ortalama ve kovaryans matrisini klasik ortalama ve kovaryans matrisi ile karşılaştırmaktadır. Bunları karşılaştırmada araştırmacı grafik içerisine  $x = C_{npr} \cdot \mathcal{X}_{p, \alpha/n}$  ve  $y = C_{npr} \cdot \mathcal{X}_{p, \alpha/n}$  doğrularını çizmelidir. Bu doğrular bir dikdörtgen belirlemektedir. Bu dikdörtgen içerisinde bulunan gözlemler temiz dışarıdakiler ise aykırı değerlerdir. Bu grafik yüksek boyutlu veride çok kullanışlı olmaktadır. Çünkü yüksek boyutta verinin ve veriyi temsil eden elipsoidin gözlemlenmesi problemlidir. Veride aykırı değerlerin olmaması durumunda tüm gözlemler dikdörtgen içerisinde birlikte bulunur.

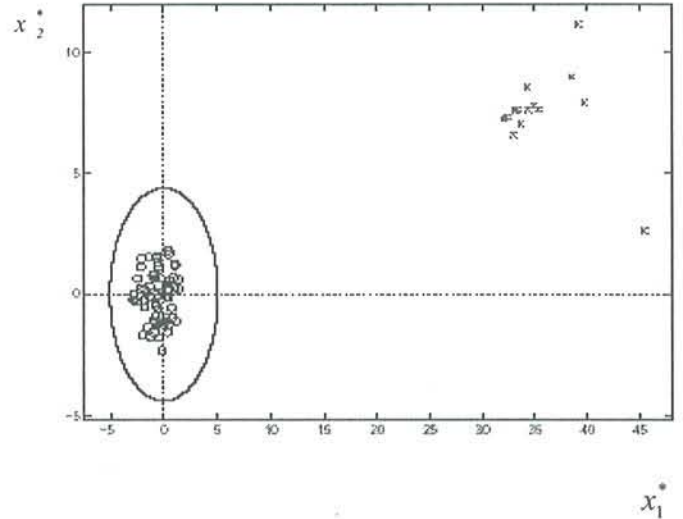
Üçüncü grafik aykırı değer belirlemek amacı ile kullanılmaktadır. Aykırı değerlerin belirlenebilmesi için  $y = C_{npr} \cdot \chi_{p,\alpha/n}$  doğrusunun çizilmesi gerekmektedir. Bu doğru üzerinde bulunan gözlemler aykırı değer olarak bilinirler. Araştırmacı aykırı değerlerin atılmasından sonra veride başka aykırı değer olup olmadığını araştırmak isterse y matrisinin bileşenlerine ait Q-Q grafiklerini de inceleyebilir.

### 3. UYGULAMA

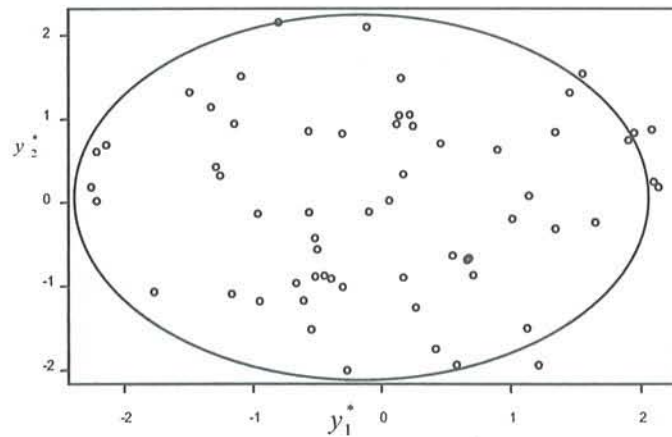
BDTBA nin etkinliği Hawkins, Bradu ve Kass (HBK) veri kümesi üzerinde gösterilecektir. HBK verisi (Hawkins ve ark., 1984) maskeleyen etkisinin incelenmesi amacı ile oluşturulan üç değişken ve toplam 75 gözlem üzerine kurulu tipik bir veri kümesidir. İlk on gözlem xy-uzayında ardından gelen dört gözlem ise x-uzayında aykırı değer olacak şekilde üretilmiştir.



Şekil 1. TBA den elde edilen HBK verisinin ilk iki bileşenine ait serpilme diyagramı



Şekil 2. DTBA ya dayalı HBK nin skorları



Şekil 3. HBK data için BDTBA den elde edilen ilk iki bileşene ait serpilme grafiği



Şekil 1.; tüm değişkenler ( $x_1, x_2, x_3$  ve  $y$ ) üzerinden elde edilen temel bileşenlerin ilk ikisine ait skorları göstermektedir. Bu grafik içerisine temel bileşenlerin %97.5 lik güven elipsoidini çizersek, klasik temel bileşenler yönteminin yalnızca 4 gözlemi aykırı değer olarak belirlediğini görebiliriz.

TBA yöntemi veri noktalarını ortalamamaktadır çünkü verinin ortalaması temiz gözlemler bulutunun dışarısında bulunmaktadır (Bkz. Şekil 1.). Diğer taraftan DTBA ve BDTBA yöntemleri veri noktalarını ortalamakta ve tüm aykırı değerleri doğru bir şekilde temiz gözlem kümesinden ayırmaktadır (Bkz. Şekil 2. ve 3.).

BDTBA yöntemi aykırı değer olan gözlemleri veri kümesinden attıktan sonra temel bileşenleri hesapladığından elipsoidin dışında gözlem bulunmamaktadır (Bkz. Şekil 3.). BDTBA nin indeks ve Q-Q grafiklerinden de aykırı değerlerin temiz veriden ayrıldıklarını görmekteyiz (Bkz. Şekil 4. ve 5.). Buna karşın TBA sonrası elde edilen Mahalanobis uzaklığına ait indeks ve Q-Q grafikleri aykırı değerleri tam olarak belirleyememektedir (Bkz. Şekil 6. ve 7.).

TBA klasik kestiricilere dayalı olduğundan aykırı değerlerin varlığında sağlıklı sonuçlar vermeyecektir. Klasik Mahalanobis uzaklığının küp küküne ait Q-Q grafiği ve indeks grafikleri bu problemi göstermektedir.

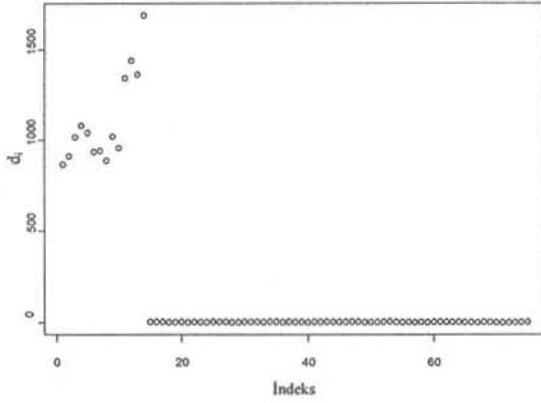
TBA klasik kestiricilere dayalı olduğundan aykırı değerlerin varlığında sağlıklı sonuçlar vermeyecektir. Klasik Mahalanobis uzaklığının küp küküne ait Q-Q grafiği ve indeks grafikleri bu problemi göstermektedir.

TBA, DTBA ve BDTBA yöntemlerin performanslarını karşılaştırmada incelenmek istenen yöntemler uygulandıktan sonra belirlenen verinin varyans-kovaryans matrisinin özdeğerleri hesaplandı. Sonuçlar Tablo 1 de görülmektedir. Bu özdeğerler ( $\lambda_i$ ) aynı zamanda  $i$ . temel bileşenin varyansına da karşılık gelmektedir. Tüm değerler bire eşitse orijinal X matrisinin sütunları ortogondur denir. Bir  $\lambda_i$  tam olarak sıfıra eşitse bu orijinal X matrisi sütunları arasında mükemmel bir doğrusal ilişki vardır. Bir ya da birden fazla sıfıra yakın değer ise çoklu iç ilişkinin varlığını vurgulamaktadır.

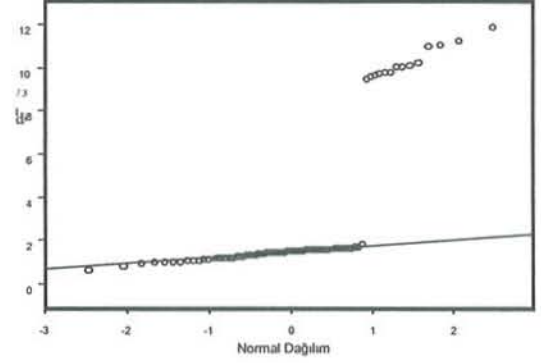
Aykırı değerler verinin yoğunlaştığı kısmın uzağında bulunduğundan TBA den elde edilen özdeğerler oldukça büyük elde edilmiştir (Bkz. Tablo 1). Diğer taraftan DTBA ve BDTBA'ne ait algoritmalar küçük özdeğerlere sahiptir. Fakat BDTBA'nın özdeğerleri DTBA'ninkilerden daha küçüktür. Bu sonuçlar gösteriyor ki BDTBA yöntemi diğerlerinden daha iyi sonuç vermektedir. BDTBA yöntemi aykırı değerleri eledikten sonra temel bileşenleri hesapladığından veri matrisinin boyutunun indirgenmesi daha güvenilir şekilde yapılmaktadır.

**Tablo1.** TBA, BDTBA, ve DTBA den elde edilen HBK verisinin özdeğerleri

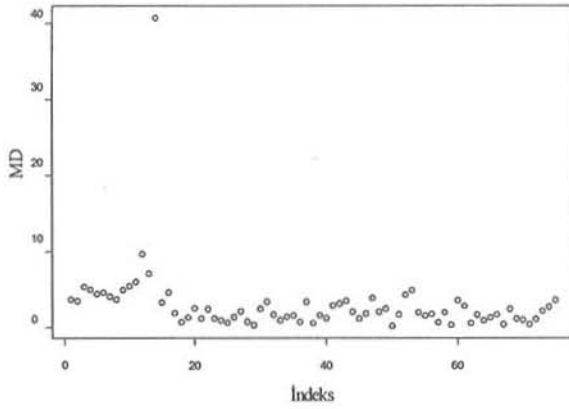
	TBA	DTBA	BDTBA
$\lambda_1$	223.12	3.47	1.326
$\lambda_2$	5.538	2.63	1.093
$\lambda_3$	1.688	2.47	0.956
$\lambda_4$	0.914	0.67	0.2957



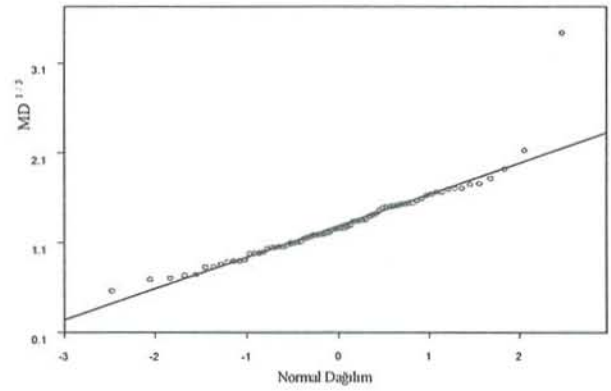
Şekil 4. HBK verisi için BDTBU'na ait indeks grafiği



Şekil 5. HBK verisinin BDTBU nun küpköküne ait Q-Q grafiği



Şekil 6. HBK data için elde edilen Mahalanobis uzaklığının indeks grafiği



Şekil 7. HBK verisi için elde edilen Mahalanobis uzaklığının küpköküne ait Q-Q grafiği

Tablo 2. Simülasyon çalışmasında kullanılacak faktörler

Gözlem Sayısı	100
Parametre sayısı	4 veya 20
Temiz gözlemin varyans kovaryans matrisi ( $\Sigma$ )	diag(4,3,2,1) (p=4 için) diag(5,4,3,2,1,0.15,0.14,...,0.02,0.01) (p=20 için)
Kirletmemiktarı ( $\varepsilon$ )	%0 veya %10
Kirletilen verinin ortalama vektörü ( $\mu^*$ )	(0,0,0,0) <sup>T</sup> veya (0,0,0,10) <sup>T</sup> (p=4 için) (0,0,0,0,0,1) <sup>T</sup> veya (0,0,0,0,1,0) <sup>T</sup> (p=20 için)
Kirletilen verinin varyans kovaryans matrisi ( $\Sigma^*$ )	$\Sigma$ veya $9\Sigma$ (p=4 için) $\Sigma$ veya $(1/20)\Sigma$ (p=20 için)



#### 4. SİMÜLASYON ÇALIŞMASI

Bu çalışmada TBA, BDTBA (Kıral ve Billor,2001) ve DTBA (Hubert, ve ark., 2002) yöntemleri altı farklı veri grubu üzerinde karşılaştırılacaktır.

Simülasyon çalışmasında kullanılan temiz ve kirletilmiş veri sırasıyla  $N(0, \Sigma)$  ve  $N(\mu^*, \Sigma^*)$  dağılımlarından üretilmektedir. BDTBA yönteminin performansını değerlendirme de kullanılan veri kümesine ait gözlem sayısı 100 ve parametre sayısı (düşük boyutlu veri için) 4 veya (büyük boyutlu veri için) 20 olarak alınmıştır. Simülasyon çalışması sonuçları 100 farklı veri kümesi üzerinden elde edilen sonuçların değerlendirilmesi ile elde edilmiştir. Kirletme seviyesi 1. ve 4. veri grupları için %0, 2., 3., 5. ve 6. veri grupları içinse %10 olarak belirlenmiştir. 2., 3., 5. ve 6. veri kümeleri içerisindeki kirletilen veriler sırasıyla  $N(0, \rho \Sigma)$ ,  $N((0,0,0,10)', \Sigma)$ ,  $N(10e_6, \Sigma)$  ve  $N(10e_5, (1/20)\Sigma)$ 'dan üretilmiştir, burada  $e_i$   $i$ . elemanı 1 diğerleri 0 olan birim vektördür.

##### Karşılaştırmada kullanılan performans ölçüleri:

- $MSE(\hat{\lambda}_i) = \frac{1}{100} \sum_{j=1}^{100} (\lambda_i - \hat{\lambda}_i^{(j)})^2$  değerleri (MSE: Ortalama kareler hatası)

( $\lambda_i$  :Veri kümesinin varyans-kovaryans matrisinin  $i$ . özdeğeri ve  $\hat{\lambda}_i^{(j)}$   $j$ . tekrarlar sonucunda elde edilen tahmini  $i$ . özdeğer),

- Veri kümesinin varyans-kovaryans matrisine ait özdeğerlerin ortalamasıdır.

Araştırmada kullanılan faktörler Tablo 2 de özetlenmiştir.

Tablo 3 ve 4 de sırasıyla  $p=4$  ve 20 için simülasyon sonuçlarını göstermektedir.  $p=4$  olması durumu sonuçlarını veren Tablo 3 de tüm özdeğer ve özvektörler için elde edilen sonuçlar verilmekte, fakat  $p=20$  için ki sonuçları gösteren Tablo 4 de sadece en büyük ilk beş özdeğer için elde edilen sonuçlar verildi.

**Tablo 3:**  $p=4$  için simülasyon sonuçları

	Veri grubu 1			Veri grubu 2			Veri grubu 3		
	TBA	BDTBA	DTBA	TBA	BDTBA	DTBA	TBA	BDTBA	DTBA
Ort $\hat{\lambda}_1$	4.190	4.187	4.23	8.197	4.565	5.43	10.278	4.162	5.01
Ort $\hat{\lambda}_2$	2.933	2.917	3.18	5.104	3.116	3.94	4.115	2.916	4.06
Ort $\hat{\lambda}_3$	1.915	1.909	2.11	3.225	2.029	2.67	2.879	1.867	3.25
Ort $\hat{\lambda}_4$	0.974	0.967	1.04	1.594	1.037	1.35	1.874	0.944	2.37
MSE $\hat{\lambda}_1$	0.352	0.356	0.39	20.38	0.811	2.61	39.905	0.327	1.38
MSE $\hat{\lambda}_2$	0.120	0.128	0.18	5.190	0.19836	1.20	1.507	0.154	1.36
MSE $\hat{\lambda}_3$	0.072	0.071	0.13	1.838	0.101	0.62	0.895	0.111	1.73
MSE $\hat{\lambda}_4$	0.017	0.017	0.03	0.477	0.040	0.19	0.821	0.038	2.00

Tablo 4 : p=20 için simülasyon sonuçları

	Veri grubu 4			Veri grubu 5			Veri grubu 6		
	TBA	BDTBA	DTBA	TBA	BDTBA	DTBA	TBA	BDTBA	DTBA
Ort $\hat{\lambda}_1$	5.361	5.361	5.12	9.519	5.532	5.30	10.05	10.05	5.55
Ort $\hat{\lambda}_2$	3.963	3.962	3.83	5.312	4.045	4.07	4.882	4.881	4.65
Ort $\hat{\lambda}_3$	2.887	2.887	2.87	3.940	2.867	3.25	3.627	3.623	3.81
Ort $\hat{\lambda}_4$	1.933	1.933	1.92	2.826	1.910	2.51	2.545	2.542	3.12
Ort $\hat{\lambda}_5$	0.953	0.952	0.96	1.875	0.952	1.78	1.695	1.695	2.24
MSE $\hat{\lambda}_1$	0.602	0.603	0.53	20.54	0.982	0.55	25.59	25.60	0.70
MSE $\hat{\lambda}_2$	0.179	0.179	0.25	2.074	0.293	0.21	1.076	1.070	0.64
MSE $\hat{\lambda}_3$	0.134	0.134	0.15	1.100	0.200	0.24	0.599	0.598	0.81
MSE $\hat{\lambda}_4$	0.063	0.063	0.12	0.793	0.104	0.41	0.401	0.397	1.43
MSE $\hat{\lambda}_5$	0.023	0.023	0.03	0.846	0.041	0.74	0.548	0.548	1.71

## 5. TARTIŞMA VE SONUÇ

Tablo 3 ve 4 de veri grubu 1-6 için elde edilen Ort  $\hat{\lambda}_i$  ve MSE  $\hat{\lambda}_i$  değerleri görülmektedir. İdeal olarak küçük Ort  $\hat{\lambda}_i$  ve MSE  $\hat{\lambda}_i$  değerleri bulmak arzulanır.

Kirletme oranının %0 olduğu, Veri grubu 1 de TBA ve BDTBA nin performans ölçü değerleri küçük ve hemen hepsinde benzer sonuçlar vermektedir. Ama DTBA yöntemi için elde edilen değerler TBA ve BDTBA için elde edilenlere göre daha yüksektir. Bu nedenle veride aykırı değer yokken DTBA yönteminin iyi işlemediğini söyleyebiliriz. Buna rağmen TBA bazı durumlar için iyi performansa sahip fakat kirletme varlığında (Veri grubu 2 ve 3) güvenilmeyen tahminler vermektedir. BDTBA yöntemi parametre sayısının 4 olması durumunda tüm kirletme seviyelerinde iyi sonuç vermektedir (Bkz. Tablo 3). Veri grubu 2 ve 3 de BDTBA ve DTBA yöntemlerinin performans ölçüleri oldukça küçük fakat BDTBA yönteminin değerleri DTBA yöntem değerlerinden biraz daha küçüktür. Bu nedenle veri grubu 1, 2 ve 3 için BDTBA diğer yöntemlerden daha üstün denilebilir.

Diğer taraftan Veri grubu 4 de TBA ve BDTBA yöntemleri hemen hemen aynı sonucu vermektedir. Yani kirletme yok ve parametre sayısı yüksek ise bu iki yöntem aynı performansı vermektedir denilebilir. DTBA değerleri Ort  $\hat{\lambda}_i$  için diğerleri ile hemen hemen aynı fakat DTBA'ne ait MSE  $\hat{\lambda}_i$  değerleri diğerlerine göre biraz daha yüksektir. Bu nedenle bu konfigürasyon içerisinde DTBA yöntemi diğer yöntemler kadar iyi sonuç vermemektedir.

Veri grubu 5 içinde pek çok bileşen içerisinde BDTBA; DTBA'dan daha iyi performansa sahiptir. Bu nedenle bu veri grubu içinde BDTBA yönteminin çoğunlukla DTBA'den daha iyi olduğu söylenebilir.

Veri grubu 6 da BDTBA'ne ait performans ölçüleri gösterir ki yöntem ilk önemli bileşen içinde aykırı değerlere karşı daha hassastır. Fakat diğer bileşenler için BDTBA yöntemi DTBA'den daha iyi sonuç vermektedir. Bu nedenle BDTBA yönteminin parametre sayısı ve kirletme



seviyesinin artması durumlarında TBA ve DTBA yöntemlerinden daha iyi sonuç verdiği güvenle söylenebilir.

Özet olarak Bu çalışmada dayanıklı temel bileşenleri elde etmede BDTBA yönteminin kullanımının yararları bir simülasyon çalışması yapılarak gösterilmiştir. Çok değişkenli veri kümeleri içerisinde çoklu aykırı değerleri bulmayı amaçlayan bu algoritma; büyük veri kümelerine (1 milyon gözlem için bile) uygulanabilmekte, model üzerinde çok küçük etkisi olabilecek gözlemleri belirleyebilmekte, hesaplama zorlukları içermemektedir. Dahası maskeleye ve bulandırma problemlerinden etkilenmemektedir. Bu nedenlerle şimdiye kadar yapılmış dayanıklı temel bileşenlere dayalı yöntemlere alternatif olarak önerilmektedir.

#### KAYNAKLAR

- BILLOR, N. , HADI, A. S. AND VELLEMAN, P. F. (2000). "BACON: Blocked Adaptive Computationally-Efficient Outlier Nominators", *Computational Statistics and Data Analysis*, 34, 279-298.
- CAMPBELL, N. A. (1980). "Robust Procedures in Multivariate Analysis I: Robust Covariance Estimation", *Applied Statistics*, 29 , 231-237.
- CARONI, C. (2000). "Outlier Detection by Robust Principal Components Analysis." *Commun. Statist.- Simula.*, 29(1), 139-151.
- CROUX, C. RUIZ-GAZEN A. (2000). "High Breakdown Estimators for Principle Components: The Projection-Pursuit Approach" revisited, under revision, <http://homepages.ulb.ac.be/~ccroux>.
- CROUX, C., AND HAESBROECK, G. (2000), "Principal Component Analysis Based on Robust Estimators of the Covariance or Correlation Matrix: Influence Functions and Efficiencies," *Biometrika*, 87, 603—618
- HAWKINS, D. M. , BRADU, D. , KASS, G. V. (1984). "Location of Several Outliers in Multiple Regression Data Using Elemental Sets". *Technometrics*, 26, 197-208.
- HUBERT, M., ROUSSEEUW, P.J., AND VERBOVEN, S. (2002). "A Fast Robust Method for Principal Components with Applications to Chemometrics". *Chemometrics and Intelligent Laboratory Systems*, 60, 101-111.
- HUBERT, M., ROUSSEEUW, P.J. (2002), "ROBPCA: A New Approach to Robust Principal Component Analysis".
- KIRAL, G. ve BILLOR, N. (2001). "BACON Temel Bileşenler Analizi" 5. Ulusal Ekonometri ve İstatistik Sempozyumu". 19-22 Eylül, ADANA
- KIRAL, G. (2003). "A Comparison of The Recent Algorithms for The Identification Of Outliers in Data". PhD Thesis. Department of Mathematics, Institute of Natural and Applied Sciences, Cukurova University.
- LIE, G. AND CHEN, Z. (1985). "Projection-Pursuit Approach to Robust Dispersion Matrices and Principal Components: Primary Theory and Monte Carlo", *J. Amer. Statistics. Assoc.*, 80, 759-766.

## **A Comparison of the Multiple Outlier Detection Method for Multivariate Data by Simulation Study**

### **ABSTRACT**

*Principle component analysis is a statistical technique used for reducing data dimension and/or constructing a set of uncorrelated variables. In some cases it is solely used as a technique of analysis itself while in other situations used as a data preparation technique for further analysis. In particular, it is preferred in testing high dimensional data since it provides dimensional reduction in data. But it is based on classical variance and covariance matrix therefore it is sensitive to outliers in data. So using robust principle component analysis is preferred to obtain reliable results in the existence of outliers.*

*In this study classical principal component analysis technique and two robust principle component techniques [Robust Principle Component Analysis (Hubert et al., 2002) and BACON Robust Principle Component Analysis (Kiral and Billor, 2001)] are compared using simulation.*

**Key Words :** *Classic Principle Component Analysis, Robust Principle Component Analysis, BACON Algorithm, Outlier*



## Türkiye’de 1980-1985 Dönemi Bölgelerin İl Merkezi, İlçe Merkezi İle Bucak Ve Köyler Bakımından Daimi İkametgaha Göre İç Göçler

Salih ÇELEBİOĞLU\*

### ÖZET

*Bu çalışmanın amacı, Türkiye’de son yıllarda ve özellikle 1985 sonrası âdeta “şiddetlenen” iç göç olayında şimdilik son “doğal dönem” olarak görebileceğimiz D.İ.E.’nin 1980-1985 dönemindeki “Daimî İkametgâha Göre İç Göçler” yayınının içerdiği bazı bilgileri göz önüne sermektir. Çalışmada Türkiye beş bölgeye ayrılmış ve bu ayrışimla eşzamanlı olarak her bir bölge kendi içinde il merkezi, ilçe merkezi ile bucak ve köyler biçiminde yerleşim merkezlerine göre ele alınmıştır. Ele alınan bölge-yerleşim merkezi türü durumları arasındaki göçler, geliş ve varış yerleri bakımından karşılaştırmalı olarak incelenmiş, ayrıca çok bölgeli demografinin kullandığı Markov zinciri kavramından yararlanılarak uzun dönemde ülke nüfusunun durumlara göre dağılımı tahmin edilmiştir. Bunlara ek olarak daha sonraki çalışmalarda kullanılabileceği düşünülerek, durumlar arası geçişlere ilişkin sayılar, net göçler, bölge içine ve bölge dışına olan göçlere de yer verilmiştir. Bunun yanında kentleşme olgusu göz önüne alınarak, Türkiye nüfusunun yerleşim merkezi türleri arasındaki iç göçüne de değinilmiştir.*

### I. GİRİŞ

Göç, nüfusu etkileyen üç temel hayati olaydan biridir. Özellikle iletişimin baş döndürücü bir hızla arttığı günümüzde ülkeler içi boyuttan ülkeler arası boyuta da taşınmış olan göç, sonucu olan kentleşmenin neden olduğu sorunlar yüzünden gelişmekte olan ülkelerin, her geçen yıl artan sığınmalar yüzünden de gelişmiş ülkelerin öncelikli problemlerinden biri haline gelmektedir. Ancak, göç olayına yalnızca bir problem olarak bakmak doğru değildir. Nitekim bu olayı, özellikle ülke içi göç veya kısaca iç göç olayında yönetimde ve yönlendirmede yerinde kararlar alındığında nüfus hedeflerine ulaşmada bir araç olarak görenler de vardır (Price, 1973).

Ülkemizde de İkinci Dünya Savaşı’ndan sonra kentlere göç biçiminde ortaya çıkan iç göç olayı, bugün önemini hâlâ korumaktadır. Ancak iç göç olayı son yıllarda gündeme geldiği ilk yıllardakinin tersine cinsiyet bakımından seçiciliğini âdeta yitirmiştir.

Özet girişte de belirtildiği gibi bu çalışmanın amacı, daha önce çeşitli varsayımlarla ele alınan Türkiye erkek nüfusunun iç göçü yerine (Çelebioğlu, 1988; 1991; 1997), (Çelebioğlu ve Bal, 1995), son yıllarda ve özellikle 1985 sonrası âdeta “şiddetlenen” iç göç olayında şimdilik son “doğal dönem” olarak görebileceğimiz D.İ.E.’nin 1980-1985 dönemindeki “Daimî İkametgâha Göre İç Göçler” yayınının içerdiği bazı bilgileri göz önüne sermektir.

Çalışmanın ikinci kesiminde iç göç olayı Markov zincirleri ile incelenmiş ve ayrıca her bir duruma girişte ve her bir durumdan gidişte en çok ve en az tercih edilen durumlara yer verilmiştir.

\*Doç. Dr., Gazi Üniversitesi, Fen-Edebiyat Fak., İstatistik Böl.

Üçüncü kesimde ülke genelinde durum türleri arasındaki göçler ele alınmış ve koşullar değişmediğinde bu göç örgüsüyle elde edilecek uzun dönemli davranışlara ilişkin bazı istatistiklere özet olarak değinilmiştir. Son kesimde, önceki kesimlerde elde edilen sonuçlar net göç bakımından ülke temelinde bölgeler arasında karşılaştırılmış ve yorumlanmıştır.

## II. BÖLGELERİN İL MERKEZİ, İLÇE MERKEZİ İLE BUCAK VE KÖYLER BAKIMINDAN DAİMİ İKAMETGAHA GÖRE İÇ GÖÇLERİ

Bu kesimde, ülkemizdeki bölgeler arasındaki daimî ikâmetgâha göre iç göçler, çok bölgeli demografide kullanılan inceleme araçlarından biri olan Markov zincirleriyle, yerleşim birimleri il merkezi, ilçe merkezi ile bucak ve köyler bakımından ele alınmıştır. Çok bölgeli demografide Markov zincirleri, göç olayında bölgeler arasındaki bağımlılık etkilerini dinamik bir biçimde modellemede kullanılmaktadır (Plane ve Rogerson, 1986). Çalışmada bölgeler sırasıyla aşağıdaki illeri içermektedir:

1. **Bölge-Orta Anadolu Bölgesi** : Afyon , Ankara, Çankırı, Çorum, Eskişehir, Kayseri, Kırşehir, Konya, Nevşehir, Niğde, Sivas, Tokat, Yozgat.
2. **Bölge-Karadeniz Bölgesi** : Amasya, Artvin, Bolu, Giresun, Gümüşhane, Kastamonu, Ordu, Rize, Samsun, Sinop, Trabzon, Zonguldak.
3. **Bölge-Marmara-Ege Bölgesi** : Aydın, Balıkesir, Bilecik, Bursa, Çanakkale, Edirne, İstanbul, İzmir, Kırklareli, Kocaeli, Kütahya, Manisa, Sakarya, Tekirdağ, Uşak.
4. **Bölge-Akdeniz Bölgesi** : Adana, Antalya, Burdur, Denizli, Gaziantep, Hatay, Isparta, İçel, Kahramanmaraş, Muğla.
5. **Bölge-Doğu-Güneydoğu Anadolu Bölgesi** : Adıyaman, Ağrı, Bingöl, Bitlis, Diyarbakır, Elâzığ, Erzincan, Erzurum, Hakkâri, Kars, Malatya, Mardin, Muş, Siirt, Tunceli, Şanlıurfa, Van.

Bu çalışmada Markov zincirinin durumları, bölge sırasındaki bölgelerin il merkezleri, ilçe merkezleri ile bucak ve köylerinden oluşmaktadır. Markov zincirinin durumlarını göstermede elde edilen sonuçları daha kolay yorumlamak için aşağıdaki kısaltmalardan yararlanılacaktır :

OA : Orta Anadolu Bölgesi,  
KD : Karadeniz Bölgesi,  
ME : Marmara-Ege Bölgesi,  
AD : Akdeniz Bölgesi,  
DG : Doğu-Güneydoğu Anadolu Bölgesi.

İM : İl Merkezleri,  
İÇ : İlçe Merkezleri,  
BK : Bucak ve Köyler.

Dolayısıyla bu kısaltmalara göre ME İÇ, Marmara-Ege Bölgesi'nin ilçe merkezlerini, DG BK, Doğu-Güneydoğu Anadolu Bölgesi'nin bucak ve köylerini gösterecektir.

1985 Türkiye Genel Nüfus Sayımı'ndan derlenen Daimî İkametgâha Göre İç Göçler kaynağından elde edilen durumlar arası geçiş sayıları sonraki sayfada yer alan matriste verilmiştir. Bu geçiş sayılarına 1980 daimî ikâmetgâhı il merkezi, ilçe merkezi ile bucak ve köyler biçiminde kesin olarak bilinmeyen nüfuslar dahil edilmemiştir. Bir sonraki sayfada da durumlar arası geçiş sayıları matrisinden hesaplanan geçiş olasılıkları matrisi verilmiştir.

Geçiş sayıları matrisinden bulunan durumlara geçiş sayıları ve yüzdeleri sırasıyla şöyledir:



**Türkiye’de 1980-1985 Dönemi Bölgelerin İl Merkezi, İlçe Merkezi İle Bucak Ve Köyler Bakımından Daimi İkametgaha Göre İç Göçler**

---

**Durum** : OA İM OA İÇ OA BK KD İM KD İÇ KD BK ME İM ME İÇ  
**Geçiş Sayısı** : 369365 174895 136960 90153 112553 114623 901137 350200  
**Geçiş Yüzdesi** : 11,0 5,2 4,1 2,7 3,3 3,4 26,8 10,4

**Durum** : ME BK AD İM AD İÇ AD BK DG İM DG İÇ DG BK  
**Geçiş Sayısı** : 218689 233265 150565 120908 158525 126942 107206  
**Geçiş Yüzdesi** : 6,5 6,9 4,5 3,6 4,7 3,8 3,2

Geçiş sayılarından ve dolayısıyla yüzdelerinden de görüldüğü gibi, 1980-1985 döneminde daimî ikâmetgâh olarak en çok ilgi çeken beş merkez sırasıyla, Marmara-Ege Bölgesi’nin il merkezleri, Orta Anadolu Bölgesi’nin il merkezleri, Marmara-Ege Bölgesi’nin ilçe merkezleri, Akdeniz Bölgesi’nin il merkezleri ve Marmara-Ege Bölgesi’nin bucak ve köyleri olmuştur. En az göç ise, Karadeniz Bölgesi’nin il merkezlerine olmuştur.

3.179.819	40.883	33.831	12.628	8.148	6.146	85.003	27.217	9.606	31.160	9.174	5.308	25.330	7.528	5.004
62.999	1.588.230	27.784	2.904	9.142	3.273	68.148	22.817	5.244	8.580	10.899	3.501	4.492	6.679	2.667
79.444	54.303	4.300.914	1.157	2.809	7.142	40.179	8.390	13.426	5.876	2.838	9.936	1.379	1.671	8.144
18.095	3.124	1.894	725.931	13.796	17.323	47.706	6.588	3.429	5.358	1.992	1.054	6.553	1.609	1.301
18.096	9.185	2.889	16.922	996.187	21.718	79.587	16.478	5.746	2.239	3.174	1.371	2.677	3.345	1.615
6.097	2.802	7.320	23.344	35.606	3.926.984	57.522	9.009	21.421	993	893	3.452	1.320	1.039	5.168
51.791	13.557	14.314	14.753	12.864	16.128	6.188.982	110.894	51.766	29.333	10.306	7.027	29.028	11.046	10.972
32.885	14.997	7.597	4.530	14.341	13.265	113.795	2.166.191	50.653	8.721	10.742	5.015	7.257	10.971	6.826
4.236	1.566	11.056	823	1.534	12.892	165.894	76.238	4.143.776	1.606	1.484	8.471	1.235	1.703	9.196
20.724	4.336	3.728	3.358	1.816	1.575	33.593	8.004	3.482	1.901.002	20.863	21.232	13.311	5.019	4.821
13.620	8.591	2.786	1.136	2.920	1.374	24.801	11.082	2.998	36.493	1.309.274	24.815	3.965	8.031	3.068
2.766	1.468	6.553	346	546	2.403	7.023	3.124	7.881	38.715	45.370	3.440.860	1.464	2.341	8.321
29.807	5.279	3.065	5.910	3.047	2.625	70.519	16.146	7.820	32.940	6.783	3.741	1.381.173	27.691	17.968
21.661	12.382	3.286	1.522	4.711	2.392	62.202	22.426	8.703	15.616	16.947	4.845	28.669	1.225.240	22.135
7.144	2.422	10.857	820	1.273	6.367	45.165	11.787	26.514	15.635	9.100	21.140	31.845	38.269	4.230.344

Geçiş Sayıları Matrisi.



**Türkiye’de 1980-1985 Dönemi Bölgelerin İl Merkezi, İlçe Merkezi İle Bucak Ve Köyler Bakımından Daimi İkametgaha Göre İç Göçler**

0,91196	0,01172	0,00970	0,00362	0,00234	0,00176	0,02438	0,00781	0,00276	0,00894	0,00263	0,00152	0,00727	0,00216	0,00143
0,03448	0,86914	0,01520	0,00159	0,00500	0,00179	0,03729	0,01249	0,00287	0,00470	0,00596	0,00192	0,00246	0,00365	0,00146
0,01751	0,01197	0,94784	0,00026	0,00062	0,00157	0,00885	0,00185	0,00296	0,00129	0,00063	0,00219	0,00030	0,00037	0,00179
0,02115	0,00365	0,00221	0,84829	0,01612	0,02024	0,05575	0,00770	0,00401	0,00626	0,00233	0,00123	0,00766	0,00188	0,00152
0,01532	0,00778	0,00245	0,01432	0,84335	0,01838	0,06738	0,01395	0,00486	0,00189	0,00269	0,00116	0,00227	0,00283	0,00137
0,00149	0,00068	0,00178	0,00569	0,00868	0,95711	0,01402	0,00220	0,00522	0,00024	0,00022	0,00084	0,00032	0,00025	0,00126
0,00788	0,00206	0,00218	0,00224	0,00196	0,00245	0,94161	0,01687	0,00788	0,00446	0,00157	0,00107	0,00442	0,00168	0,00167
0,00133	0,00608	0,00308	0,00183	0,00581	0,00537	0,04611	0,87779	0,02053	0,00353	0,00435	0,00203	0,00294	0,00445	0,00277
0,00095	0,00035	0,00249	0,00019	0,00035	0,00290	0,03735	0,01717	0,93292	0,00036	0,00033	0,00191	0,00028	0,00038	0,00207
0,01013	0,00212	0,00182	0,00164	0,00089	0,00077	0,01641	0,00391	0,00170	0,92874	0,01019	0,01037	0,00650	0,00245	0,00236
0,00936	0,00590	0,00192	0,00078	0,00201	0,00094	0,01705	0,00762	0,00206	0,02508	0,89987	0,01706	0,00273	0,00552	0,00211
0,00077	0,00041	0,00184	0,00010	0,00015	0,00067	0,00197	0,00087	0,00221	0,01085	0,01271	0,96405	0,00041	0,00066	0,00233
0,01846	0,00327	0,00190	0,00366	0,00189	0,00163	0,04368	0,01000	0,00484	0,02040	0,00420	0,00232	0,85547	0,01715	0,01113
0,01491	0,00852	0,00226	0,00105	0,00324	0,00165	0,04282	0,01544	0,00599	0,01075	0,01167	0,00333	0,01973	0,84340	0,01524
0,00160	0,00054	0,00244	0,00018	0,00029	0,00143	0,01013	0,00264	0,00595	0,00351	0,00204	0,00474	0,00714	0,00858	0,94879

Geçiş Olasılıkları Matrisi.

Şimdi her bir durumu kendisine en çok nüfus gönderen ilk beş merkez ile en az nüfus gönderen merkez bakımından dışarıdan alınan nüfus ve yüzdeleriyle birlikte inceleyelim.

**Tablo 1.** Orta Anadolu Bölgesi İl Merkezlerine Gelenler

Göç Derecesi	Gelenin Yer	Geçiş Sayısı	Geçiş Yüzdesi*
1	OA BK	79.444	21,5
2	OA İÇ	62.999	17,1
3	ME İM	51.791	14,0
4	ME İÇ	32.885	8,9
5	DG İM	29.807	8,1
En Az	AD BK	2.766	0,8

\* Dışarıdan bu merkeze olan göçler içindeki yüzde.

**Tablo 2.** Orta Anadolu Bölgesi İlçe Merkezlerine Gelenler

Göç Derecesi	Gelenin Yer	Geçiş Sayısı	Geçiş Yüzdesi
1	OA BK	54.303	31,1
2	OA İM	40.883	23,4
3	ME İÇ	14.997	8,6
4	ME İM	13.557	7,8
5	DG İÇ	12.382	7,1
En Az	ADBK	1.468	0,8

**Tablo 3.** Orta Anadolu Bölgesi Bucak ve Köylerine Gelenler

Göç Derecesi	Gelenin Yer	Geçiş Sayısı	Geçiş Yüzdesi
1	OA İM	33.873	24,7
2	OA İÇ	27.784	20,2
3	ME İM	14.314	10,5
4	ME BK	11.056	8,1
5	DG BK	10.857	7,9
En Az	KD İM	1.894	1,4

**Tablo 4.** Karadeniz Bölgesi İl Merkezlerine Gelenler

Göç Derecesi	Gelenin Yer	Geçiş Sayısı	Geçiş Yüzdesi
1	KD BK	23.344	25,9
2	KD İÇ	16.922	18,8
3	ME İM	14.753	16,4
4	OA İM	12.628	14,0
5	DG İM	5.910	6,6
En Az	AD BK	346	0,4

**Tablo 5.** Karadeniz Bölgesi İlçe Merkezlerine Gelenler

Göç Derecesi	Gelinen Yer	Geçiş Sayısı	Geçiş Yüzdesi
1	OA BK	35.606	31,6
2	OA İÇ	14.341	12,7
3	ME İM	13.796	12,3
4	ME İÇ	12.864	11,4
5	DG İM	9.142	8,1
En Az	AD BK	546	0,5

**Tablo 6.** Karadeniz Bölgesi Bucak ve Köylerine Gelenler

Göç Derecesi	Gelinen Yer	Geçiş Sayısı	Geçiş Yüzdesi
1	KD İÇ	21.718	19,0
2	KD İM	17.323	15,1
3	ME İM	16.128	14,1
4	ME İÇ	13.265	11,6
5	ME BK	12.892	11,3
En Az	AD İÇ	1.374	1,2

**Tablo 7.** Marmara-Ege Bölgesi İl Merkezlerine Gelenler

Göç Derecesi	Gelinen Yer	Geçiş Sayısı	Geçiş Yüzdesi
1	ME BK	165.894	18,4
2	ME İÇ	113.795	12,6
3	OA İM	85.003	9,4
4	KD İÇ	79.587	8,8
5	DG İM	70.519	7,8
En Az	AD BK	7.023	0,8

**Tablo 8.** Marmara-Ege Bölgesi İlçe Merkezlerine Gelenler

Göç Derecesi	Gelinen Yer	Geçiş Sayısı	Geçiş Yüzdesi
1	ME İM	110.894	31,7
2	ME BK	76.238	21,8
3	OA İM	27.217	7,8
4	OA İÇ	22.817	6,5
5	DG İÇ	22.426	6,4
En Az	AD BK	3.124	0,9



**Tablo 9.** Marmara-Ege Bölgesi Bucak ve Köylerine Gelenler

Göç Derecesi	Gelinen Yer	Geçiş Sayısı	Geçiş Yüzdesi
1	ME İM	51.766	23,7
2	ME İÇ	50.653	23,2
3	DG BK	26.514	12,1
4	KD BK	21.421	9,8
5	OA BK	13.426	6,1
En Az	AD İÇ	2.998	1,4

**Tablo 10.** Akdeniz Bölgesi İl Merkezlerine Gelenler

Göç Derecesi	Gelinen Yer	Geçiş Sayısı	Geçiş Yüzdesi
1	AD BK	38.715	16,6
2	AD İÇ	36.493	15,6
3	DG İM	32.940	14,1
4	OA İM	31.160	13,4
5	ME İM	29.333	12,6
En Az	KD BK	993	0,5

**Tablo 11.** Akdeniz Bölgesi İlçe Merkezlerine Gelenler

Göç Derecesi	Gelinen Yer	Geçiş Sayısı	Geçiş Yüzdesi
1	AD BK	45.370	30,1
2	AD İM	20.863	13,9
3	DG İÇ	16.947	11,3
4	OA İÇ	10.899	7,2
5	ME İÇ	10.742	7,1
En Az	KD BK	893	0,6

**Tablo 12.** Akdeniz Bölgesi Bucak ve Köylerine Gelenler

Göç Derecesi	Gelinen Yer	Geçiş Sayısı	Geçiş Yüzdesi
1	AD İÇ	24.815	20,5
2	AD İM	21.232	17,6
3	DG BK	21.140	17,5
4	OA BK	9.936	8,2
5	ME BK	8.471	7,0
En Az	KD İM	1.054	0,9

**Tablo 13.** Doğu-Güneydoğu Anadolu Bölgesi İl Merkezlerine Gelenler

Göç Derecesi	Gelinen Yer	Geçiş Sayısı	Geçiş Yüzdesi
1	DG BK	31.845	20,1
2	ME İM	29.028	18,3
3	DG İÇ	28.669	18,1
4	OA İM	25.330	16,0
5	AD İM	13.311	8,4
En Az	ME BK	1.235	0,8

**Tablo 14.** Doğu-Güneydoğu Bölgesi Anadolu İlçe Merkezlerine Gelenler

Göç Derecesi	Gelinen Yer	Geçiş Sayısı	Geçiş Yüzdesi
1	DG BK	38.269	30,2
2	DG İM	27.691	21,8
3	ME İM	11.046	8,7
4	ME İÇ	10.971	8,6
5	AD İÇ	8.031	6,3
En Az	KD BK	1.039	0,8

**Tablo 15.** Doğu-Güneydoğu Anadolu Bölgesi Bucak ve Köylerine Gelenler

Göç Derecesi	Gelinen Yer	Geçiş Sayısı	Geçiş Yüzdesi
1	DG İÇ	22.135	20,7
2	DG İM	17.968	16,8
3	ME İM	10.972	10,2
4	ME BK	9.196	8,6
5	AD BK	8.321	7,8
En Az	KD İM	1.301	1,2

Doğrudan geçiş olasılıkları matrisini gözlemleyerek, göçmenlerin ilk beş sırada ve son sırada varış yeri olarak tercih ettikleri yerleri Tablo 16-30'daki gibi belirlenmiştir:

**Tablo 16.** Orta Anadolu Bölgesi İl Merkezlerinden Gidenler

Göç Derecesi	Gidilen Yer	G e ç i ş		
		Sayısı	Yüzdesi*	Olasılığı
1	ME İM	85.003	27,7	0,02438
2	OA İÇ	40.883	13,3	0,01172
3	OA BK	33.831	11,0	0,00970
4	AD İM	31.160	10,2	0,00894
5	ME İÇ	27.217	8,9	0,00781
En Az	DG BK	5.004	1,6	0,00143

\* Merkez dışına göç eden nüfus içindeki yüzde.

**Tablo 17.** Orta Anadolu Bölgesi İlçe Merkezlerinden Gidenler

Göç Derecesi	Gidilen Yer	G e ç i ş		
		Sayısı	Yüzdesi	Olasılığı
1	ME İM	68.148	28,5	0,03729
2	OA İM	62.999	26,3	0,03448
3	OA BK	27.784	11,6	0,01520
4	ME İÇ	22.817	9,5	0,01249
5	AD İÇ	10.899	4,6	0,00596
En Az	DG BK	2.667	1,1	0,00146

**Tablo 18.** Orta Anadolu Bölgesi Bucak ve Köylerinden Gidenler

Göç Derecesi	Gidilen Yer	G e ç i ş		
		Sayısı	Yüzdesi	Olasılığı
1	OA İM	79.444	33,6	0,01751
2	OA İÇ	54.303	22,9	0,01197
3	ME İM	40.179	17,0	0,00885
4	ME BK	13.426	5,7	0,00296
5	AD BK	9.936	4,2	0,00219
En Az	KD İM	1.157	0,4	0,00026

**Tablo 19.** Karadeniz Bölgesi İl Merkezlerinden Gidenler

Göç Derecesi	Gidilen Yer	G e ç i ş		
		Sayısı	Yüzdesi	Olasılığı
1	ME İM	47.706	36,7	0,05575
2	OA İM	18.095	13,9	0,02115
3	KD BK	17.323	13,3	0,02024
4	KD İÇ	13.796	10,6	0,01612
5	ME İÇ	6.588	5,1	0,00770
En Az	KD İM	1.054	0,8	0,00123

**Tablo 20.** Karadeniz Bölgesi İlçe Merkezlerinden Gidenler

Göç Derecesi	Gidilen Yer	G e ç i ş		
		Sayısı	Yüzdesi	Olasılığı
1	ME İM	79.587	43,0	0,06738
2	KD BK	21.718	11,7	0,01838
3	OA İM	18.096	9,8	0,01532
4	KD İM	16.922	9,1	0,01432
5	ME İÇ	16.478	8,9	0,01395
En Az	AD BK	1.371	0,7	0,00116



**Tablo 21.** Karadeniz Bölgesi Bucak ve Köylerinden Gidenler

Göç Derecesi	Gidilen Yer	G e ç i ş		
		Sayısı	Yüzdesi	Olasılığı
1	ME İM	57.522	32,7	0,01402
2	KD İÇ	35.606	20,2	0,00868
3	KD İM	23.344	13,3	0,00569
4	ME BK	21.421	12,2	0,00522
5	ME İÇ	9.009	5,1	0,00220
En Az	AD İÇ	893	0,5	0,00022

**Tablo 22.** Orta Anadolu Bölgesi İl Merkezlerinden Gidenler

Göç Derecesi	Gidilen Yer	G e ç i ş		
		Sayısı	Yüzdesi	Olasılığı
1	ME İÇ	110.894	28,9	0,01687
2	OA İM	51.791	13,5	0,00788
3	ME BK	51.766	13,5	0,00788
4	AD İM	29.333	7,6	0,00446
5	DG İM	29.028	7,6	0,00442
En Az	AD BK	7.027	1,8	0,00107

**Tablo 23.** Marmara-Ege Bölgesi İlçe Merkezlerinden Gidenler

Göç Derecesi	Gidilen Yer	G e ç i ş		
		Sayısı	Yüzdesi	Olasılığı
1	ME İM	113.795	37,7	0,04611
2	ME BK	50.693	16,8	0,02053
3	OA İM	32.885	10,9	0,01333
4	OA İÇ	14.997	5,0	0,00608
5	KD İÇ	14.341	4,8	0,00581
En Az	KD İM	4.530	1,5	0,00183

**Tablo 24.** Marmara-Ege Bölgesi Bucak ve Köylerinden Gidenler

Göç Derecesi	Gidilen Yer	G e ç i ş		
		Sayısı	Yüzdesi	Olasılığı
1	ME İM	165.894	55,7	0,03735
2	ME İÇ	76.238	25,6	0,01717
3	KD BK	12.892	4,3	0,00290
4	OA BK	11.056	3,7	0,00249
5	DG BK	9.196	3,1	0,00207
En Az	KD İM	823	0,3	0,00019

**Tablo 25.** Akdeniz Bölgesi İl Merkezlerinden Gidenler

Göç Derecesi	Gidilen Yer	G e ç i ş		
		Sayısı	Yüzdesi	Olasılığı
1	ME İM	33.593	23,0	0,01641
2	AD BK	21.232	14,6	0,01037
3	AD İÇ	20.863	14,3	0,01019
4	OA İÇ	20.724	14,2	0,01013
5	DG İM	13.311	9,1	0,00650
En Az	KD BK	1.575	1,1	0,00077

**Tablo 26.** Akdeniz Bölgesi İlçe Merkezlerinden Gidenler

Göç Derecesi	Gidilen Yer	G e ç i ş		
		Sayısı	Yüzdesi	Olasılığı
1	AD İM	36.493	25,1	0,02508
2	AD BK	24.815	17,0	0,01706
3	ME İM	24.801	17,0	0,01705
4	OA İM	13.620	9,3	0,00936
5	ME İÇ	11.082	7,6	0,00762
En Az	KD İM	1.136	0,8	0,00078

**Tablo 27.** Akdeniz Bölgesi Bucak ve Köylerinden Gidenler

Göç Derecesi	Gidilen Yer	G e ç i ş		
		Sayısı	Yüzdesi	Olasılığı
1	AD İÇ	45.370	35,4	0,01271
2	AD İM	38.715	30,2	0,01085
3	DG BK	8.321	6,5	0,00233
4	ME BK	7.881	6,1	0,00221
5	ME İM	7.023	5,5	0,00197
En Az	KD İM	346	0,3	0,00010

**Tablo 28.** Doğu-Güneydoğu Anadolu Bölgesi İl Merkezlerinden Gidenler

Göç Derecesi	Gidilen Yer	G e ç i ş		
		Sayısı	Yüzdesi	Olasılığı
1	ME İM	70.519	30,2	0,04368
2	AD İM	32.940	14,1	0,02040
3	OA İM	29.807	12,8	0,01846
4	DG İÇ	27.691	11,9	0,01715
5	DG BK	17.968	7,7	0,01113
En Az	KD BK	2.625	1,1	0,00163

**Tablo 29.** Doğu-Güneydoğu Anadolu Bölgesi İlçe Merkezlerinden Gidenler

Göç Derecesi	Gidilen Yer	G e ç i ş		
		Sayısı	Yüzdesi	Olasılığı
1	ME İM	62.202	27,3	0,04282
2	DG İM	28.669	12,6	0,01973
3	ME İÇ	22.426	9,9	0,01544
4	DG BK	22.135	9,7	0,01524
5	OA İM	21.661	9,5	0,01491
En Az	KD İM	1.522	0,7	0,00105

**Tablo 30.** Doğu-Güneydoğu Anadolu Bölgesi Bucak ve Köylerinden Gidenler

Göç Derecesi	Gidilen Yer	G e ç i ş		
		Sayısı	Yüzdesi	Olasılığı
1	ME İM	45.165	19,8	0,01013
2	DG İÇ	38.269	16,8	0,00858
3	DG İM	31.845	13,9	0,00714
4	ME BK	26.514	11,6	0,00595
5	AD BK	21.140	9,3	0,00474
En Az	KD İM	820	0,4	0,00018

1980-1985 dönemindeki “Daimî İkametgâha Göre İç Göçler”den elde edilen bir diğer bilgi, 1985 yılındaki ülkemiz nüfusunun daimî ikâmetgâh açısından durumlara göre dağılımıdır :

Durum : OA İM OA İÇ OA BK KD İM KD İÇ KD BK ME İM ME İÇ  
Nüfus Oranı : 0,07912 0,04146 0,10296 0,01942 0,02680 0,09310 0,14914 0,05600

Durum : ME BK AD İM AD İÇ AD BK DG İM DG İÇ DG BK  
Nüfus Oranı : 0,10079 0,04645 0,03301 0,08099 0,03663 0,03294 0,10117.

Eğer koşullar aynı kalsaydı, yani durumlar arası daimî ikâmetgâha göre göç örgüsü böyle devam etseydi, nüfusun daimî ikâmetgâha göre dağılımı aşağıdaki gibi olurdu :

Durum : OA İM OA İÇ OA BK KD İM KD İÇ KD BK ME İM ME İÇ  
Nüfus Oranı : 0,09126 0,02866 0,05998 0,01353 0,01562 0,05899 0,29827 0,07673

Durum : ME BK AD İM AD İÇ AD BK DG İM DG İÇ DG BK  
Nüfus Oranı : 0,08534 0,07494 0,03479 0,07619 0,02575 0,01614 0,04381.

1985 yılındaki dağılımla bu son limit dağılım karşılaştırıldığında, uzun dönemde il merkezlerinden sırasıyla Marmara-Ege, Akdeniz ve Orta Anadolu bölgelerindekilerin nüfus oranının artacağı, ilçe merkezlerinden Marmara-Ege ve Akdeniz bölgelerindekilerin nüfus oranının artacağı, bucak ve köylerden ise tüm bölgelerdekilerin nüfus oranının azalacağı görülmektedir. Limit dağılımda göze çarpan bir diğer durum, Marmara-Ege Bölgesi’nin il merkezlerinin ülke nüfusunun % 30’ a yaklaşan (% 29,8) bir kısmının daimî ikâmetgâhı olacağıdır. Yine bölge temelinde bakıldığında Marmara-Ege Bölgesi’nin nüfusunun oransal olarak % 15,4’ün üzerinde artarak tüm yerleşim merkezlerinde nüfusumuzun % 46’sını aşan bir kısmının, Akdeniz Bölgesi’nin nüfusunun % 2,5’in biraz üzerinde artarak nüfusumuzun % 18,5’i aşan bir kısmının, Orta Anadolu Bölgesi’nin % 4,3’den fazla azalarak nüfusumuzun % 18’inin, Karadeniz Bölgesi’nin % 5,1’den fazla azalarak nüfusumuzun % 8,8’inin, Doğu-Güneydoğu Anadolu Bölgesi’nin %



8,5'den fazla azalarak nüfusumuzun % 17'sini aşan bir kısmının daimî ikâmetgâhı olacağı görülüyor. Nüfus kaybında Doğu-Güneydoğu Anadolu ilk sırada olup, 1985'deki nüfusuyla karşılaştırıldığında, uzun dönemde nüfusunun % 49,8'ini aşan bir kısmını kaybediyor.

### III. ÜLKE GENELİNDE İL MERKEZİ, İLÇE MERKEZİ İLE BUCAK VE KÖY ARASINDAKİ DAİMİ İKÂMETGAHA GÖRE GÖÇLER

Bir anlamda, yerleşim merkezleri, yani il merkezi, ilçe merkezi ile bucak ve köyler arasındaki daimî ikâmetgâha göre göçler ülkemizdeki kentleşme olgusunu daha yakından tanımamıza yardımcı olur. 1980-1985 dönemindeki "Daimî İkâmetgâha Göre İç Göçler"den sözkonusu yerleşim merkezleri arasındaki geçiş sayıları ve bu geçiş sayılarından elde edilmiş geçiş olasılıkları matrisi aşağıda verilmiştir :

Geçiş Sayıları:

	İM	İÇ	BK
İM	13.943.810	377.710	255.160
İÇ	643.517	7.504.982	235.566
BK	542.838	317.585	20.250.530

Geçiş Matrisi :

$$P = \begin{bmatrix} 0,95659 & 0,02591 & 0,01750 \\ 0,07675 & 0,89515 & 0,02810 \\ 0,02571 & 0,01505 & 0,95924 \end{bmatrix}$$

1980-1985 dönemindeki "Daimî İkâmetgâha Göre İç Göçler"den elde edilen yerleşim merkezlerine göre 1985'deki nüfus dağılımı aşağıda verilmiştir :

Yerleşim Merkezi :	İM	İÇ	BK
Nüfus Oranı :	0,33075	0,19024	0,47901.

Koşulların aynı kalması durumunda, yani böyle bir göç örgüsünde yerleşim merkezlerine göre daimî ikâmetgâh bakımından uzun dönemde nüfus dağılımının aşağıdaki biçimde olması beklenir :

Yerleşim Merkezi :	İM	İÇ	BK
Nüfus Oranı :	0,49796	0,17062	0,33142.

1985'deki dağılım ve limit dağılım karşılaştırıldığında, uzun dönemde il merkezlerinin daimî ikâmetgâh bakımından çekiciliğini koruyarak ülke nüfusunun yaklaşık yarısına ulaşacağı, ilçe merkezlerinin nüfusunun % 2'ye yaklaşan bir oranda azalacağı, bucak ve köylerinse nüfusunun üçte birine yaklaşan bir oranını ( % 30,8, ülke genelinin % 14,8'e ulaşan bir kısmını ) kaybederek ülke nüfusunun yaklaşık üçte birinin daimî ikâmetgâhı durumuna dönüşeceği görülüyor. Kentleşme açısından bakıldığında bu sonuç, pek sağlıklı olmasa da eğer ilçe merkezleri de kent olarak görülürse uzun dönemde ülke nüfusunun üçte ikisini aşan ( % 66,9 ) bir kısmının kentsel yerleşim birimlerini daimî ikâmetgâh olarak seçeceğini söyleyebiliriz. Limit dağılımdan ayrıca il merkezinden ayrılan bir kişinin il merkezine ilk geri dönüşünü ortalama yaklaşık 10 yıl sonra, ilçe merkezinden ayrılan bir kişinin ilçe merkezine ilk geri dönüşünü 29,3 yıl sonra ve bucak ve köyden ayrılan bir kişinin bucak ve köye ilk geri dönüşünü 15,1 yıl sonra yapacağı bilgisini de elde ediyoruz.

### IV. DİĞER SONUÇLAR

Önceki kesimlerde verilen tablolar dikkatle incelendiğinde aşağıdaki sonuçlara ulaşılır: bir geliş yeri olarak il merkezleri ele alındığında, tüm bölgelerde il merkezlerini ilk sırada seçenler o bölgelerin bucak ve köylerinden göç edenler olmaktadır. Aynı açıdan ilçe merkezleri incelendiğinde, Marmara-Ege Bölgesi dışında tüm bölgelerde ilçe merkezlerini ilk sırada seçenler yine o bölgelerin bucak ve köylerinden göç edenler olmaktadır. Marmara-Ege Bölgesi'ndeki ilçe merkezlerini en fazla tercih edenler, aynı

## Türkiye’de 1980-1985 Dönemi Bölgelerin İl Merkezi, İlçe Merkezi İle Bucak Ve Köyler Bakımından Daimi İkametgaha Göre İç Göçler

bölgenin il merkezlerinden göç edenlerdir. Geliş yeri olarak bucak ve köylere gelince, Marmara-Ege Bölgesi ile Orta

Anadolu Bölgesi’nde o bölgelerin il merkezleri ilk sırada yer alırken diğer bölgelerde ise o bölgelerin ilçe merkezleri ilk sıraya yerleşmektedir (Bkz Tablo 1-15).

Variş yeri (gidilen yer) olarak il merkezleri incelendiğinde, Akdeniz Bölgesi’nin ilçe merkezleriyle Akdeniz ve Orta Anadolu Bölgesi’nin bucak ve köyleri dışında, tüm bölgelerin tüm yerleşim merkezlerinden en fazla göç, Marmara-Ege Bölgesi’nin il merkezlerine olmaktadır. Üstelik, bulunduğu bölgenin yerleşim yeri dışına göç edenler en az % 19,8 olasılıkla Marmara-Ege Bölgesi’nin il merkezlerini seçmektedir. Bu olasılık Marmara-Ege Bölgesi’nin ilçe merkezlerinden göç edenlerde % 55,7’ye kadar çıkmaktadır. Bu bulguya ek olarak, Doğu-Güneydoğu il merkezleri dışında tüm bölgelerin yerleşim merkezlerinde öncelikle bölge içi göçler ilk üç sırada yer almaktadır. Doğu-Güneydoğu Anadolu Bölgesi il merkezlerinde ise, sonraki iki sırayı Akdeniz ve Orta Anadolu bölgelerinin il merkezleri almaktadır.

Bölge içi göçleri tüm bölgelerin yerleşim merkezi türleri arasındaki göçlerle karşılaştırmak için, her bir bölge içindeki yerleşim merkezleri arasındaki göç sayılarıyla bu sayıların ülke içindeki yerleşim merkezleri arasındaki göç sayılarına oranı aşağıda verilmiştir :

Bölge	:	OA	KD	ME	AD	DG
Geçiş Sayısı	:	299.244	128.709	569.240	187.488	166.577
Göç Oranı (%)	:	8,9	3,8	16,9	5,6	4,9.

Dolayısıyla bölge içi göçlerin toplam sayısı 1.351.258 olup, bu göçler ülke içindeki yerleşim merkezleri arasındaki toplam göçün ( 3.365.986 ), % 40,1’ini oluşturmakta, bölgelerin dışına yapılan göçler ise 2.014.728 sayısı ile % 59,9’una ulaşmaktadır. Bölge dışına yapılan göçlerin dağılımı da aşağıdaki gibi belirlenmiştir :

Bölge	:	OA	KD	ME	AD	DG
Geçiş Sayısı	:	483.545	362.141	414.068	232.375	522.599
Göç Oranı (%)	:	14,4	10,9	12,3	6,9	15,5.

Bu iki dağılımdan da görüldüğü gibi, Marmara-Ege Bölgesi bölge içi göçler bakımından, Doğu-Güneydoğu Anadolu Bölgesi de bölge dışına olan göçler bakımından ülkemizin en hareketli bölgesidir. Bölgelere gelenler açısından bakıldığında aşağıdaki dağılım elde edilmiştir :

Bölge	:	OA	KD	ME	AD	DG
Geçiş Sayısı	:	381.976	188.620	900.786	317.250	226.096
Göç Oranı (%)	:	11,3	5,6	26,8	9,4	6,7.

-16-

Dolayısıyla bölgelerin net göç değerleri aşağıdaki biçimde gerçekleşmiştir :

Bölge	:	OA	KD	ME	AD	DG
Net Göç	:	-101.569	-173.521	486.718	84.875	-296.503.

Bu net göç değerlerine göre Marmara-Ege ve Akdeniz bölgeleri göç alan ve diğer bölgeler göç veren durumundadır.

Daha ayrıntılı bir biçimde 1980-1985 döneminde bölge-yerleşim merkezlerinin net göç değerleri aşağıdaki gibi gerçekleşmiştir :

Durum	:	OA İM	OA İÇ	OA BK	KD İM	KD İÇ	KD BK	ME İM	ME İÇ
Net Göç	:	62.399	-64.234	-99.734	-39.669	-72.489	-61.363	517.358	48.605

Durum	:	ME BK	AD İM	AD İÇ	AD BK	DG İM	DG İÇ	DG BK
Net Göç	:	-79.245	87.403	4.855	-7.413	-74.816	-100.555	-121.123.



Bu değerlerden de görüldüğü gibi, Marmara-Ege ve Akdeniz bölgelerinin il ve ilçe merkezleriyle Orta Anadolu Bölgesi'nin il merkezleri dışında tüm bölgelerin yerleşim merkezleri göç verir durumdadır.

"Daimî İkâmetgâha Göçler"den bu çalışmada daha verilebilecek birçok bilgi vardır. Sözgelimi, beşer yıllık dönemlere göre bölge/yerleşim merkezleri nüfus oranları tahmini bunlardan biridir. Ancak bu ve benzer noktadaki açılımlar başka bir çalışmanın konusu olarak incelenebileceğinden bu çalışmanın kapsamı dışında kalmaktadır.

#### KAYNAKLAR

- ÇELEBİOĞLU, S., (1988, Doktora Tezi, Gazi Üni., Fen Bil. Ens. ), *Markov Zincirlerinde Durum Tahmini ve İç Göçe Uygulaması*
- ÇELEBİOĞLU, S., (1991), "*Nedensel Matrisler ve İç Göçler*", Gazi Üni. İktisadi ve İdari Bilimler Fakültesi Dergisi, 7(1-2), 153-161.
- ÇELEBİOĞLU, S. ve Bal, H., (1995), "Sürekli Parametrelili Sabit Nedensel Markov Zincirleri ve İç Göçe Uygulanışı", Gazi Üni. Fen-Edebiyat Fak. Fen Bilimleri Dergisi, 5, 105-118.
- ÇELEBİOĞLU, S., (1997), "İç Göçlerin Markov Zincirleriyle Modellenmesi Üzerine", Araştırma Sempozyumu '97 Bildirileri, T.C. Başbakanlık D.İ.E., 120-124.
- PLANE, D. A. ve Rogerson, P. A. (1986), "Dynamic Flow Modelling with Interregional Dependency Effects : An Application to Structural Change in the U.S. Migration System", *Demography*, 23 (1), 91-104.
- PRICE, C. A., (1973), "Migration as a Means of Achieving Population Targets", Seminar on Demographic Research in Relation to Population Growth Targets, 1974 World Population Year, Committee for International Coordination of National Research in Demography, Cired.
- Daimi İkametgaha Göre İç Göçler, T.C. Başbakanlık D.İ.E., 1985.

### **The Internal Migration By The Permanent Residence Among The Regions in Turkey in View Of Centers of Province, District and Subdistrict and Villages, in The Period 1980-1985**

#### ABSTRACT

*In the last years, especially after 1985, there is nearly an "intensified" internal migration fact in Turkey. For this reason, the period 1980-1985 can be seen as a "last natural period" for the present. The aim of this study is to bring some knowledges included in "Internal Migration by Permanent Residence" of D.İ.E. into the open. In this study, Turkey is divided into five regions, and simultaneously each region is separated as centers of provinces, centers of districts and subdistricts and villages. The migrations among the states, the regions-kinds of centers of population are taken up*



*comparatively in respect of origins and destinations of migrators. And furthermore, taking the advantage of concept of Markov chains in which the multiregional demography uses the long-run distribution into states of Turkey's population is estimated. In addition, under the expectation to use for the future studies, it is also given place to the frequencies of transitions among the states, net migrations, the migrations into the regions and outside the regions. At the same time, the migrations among the kinds of centers of population are included as well, as a means of understanding the urbanization in Turkey.*

**Key Words:** *Internal migration by permanent residence, Net migration, Markov chain*

## Whittaker Düzeltme Yöntemi ve Türkiye Ölüm Oranlarına Bir Uygulaması

Hatice TUZGÖL\* Süleyman GÜNAY\*\*

### ÖZET

*Bu çalışmanın amacı ölüm oranlarına uygulanan ve parametrik olmayan Whittaker Düzeltme Yöntemini incelemek ve Türkiye verilerine bir uygulamasını yapmaktır.*

*Uygulama bölümünde kullanılan veriler, Devlet İstatistik Enstitüsü'nün derlemekte olduğu il ve ilçe merkezleri ölüm ve nüfus istatistikleridir. Bu istatistiklerden hesaplanan ham ölüm oranlarına, Whittaker düzeltme yöntemini uygulamak amacıyla Microsoft Excel'de hazırlanmış bir program tekrar gözden geçirilip kullanılmıştır. Uygulama sonucunda ham ölüm oranları düzeltilmiş, Türkiye il ve ilçe merkezleri için hesaplanan düzeltilmiş ham ölüm oranları kullanılarak cinsiyet bazında 1990 ve 2000 yılları için ölüm tabloları hazırlanmış ve yaşam beklentileri hesaplanmıştır. Uygulamada ulaşılan sonuçlar değerlendirilmiştir.*

**Anahtar Kelimeler:** Ölüm oranı, düzeltme, Whittaker

### 1. GİRİŞ

Belirli bir topluluğa ilişkin yaşlar itibariyle oluşturulan ölüm oranları dizisine ölüm tablosu ya da yaşam tablosu denilmektedir. Bu tablolar herhangi bir yaşta, bir yıl içinde kaç kişinin hayatta kalacağını ya da kaç kişinin öleceğini göstermektedir. Türkiye'de henüz tam anlamıyla böyle bir ölüm tablosu oluşturulamamıştır. Sosyal güvenlik kurumlarının, özel sigorta şirketlerinin, özel sandıkların ve faaliyetlerine yeni başlamış olan bireysel emeklilik şirketlerinin aktüeryal hesaplamalarında böyle bir tabloyu kullanmaları kaçınılmaz bir gerekliliktir. Dolayısıyla Türkiye'de daha sağlıklı aktüeryal hesaplar yapılabilmesi için Türkiye'nin kendi demografik özelliklerine göre hazırlanmış bir ölüm tablosunun oluşturulması gerekmektedir.

Aktüerya alanında ölüm tabloları, primlerin, rezervlerin ve anüitelerin (primlere karşı belirli dönemlerde sigortalılara yapılan ödemelerin) hesaplanmasında kullanılmaktadır. Ölüm oranlarının düzenli ve sürekli değiştiği varsayımı, hesaplamaları yanlış sonuçlara götürebilir. Çünkü tablodaki yaşlar itibariyle değişken düzensizlikler, primlerin düzenli ilerlemesini etkiliyor olabilir. Bu nedenle, bu düzensiz yapılar ortaya çıkarılmalı ve anlamlı sonuçlara ulaşmak için düzeltme yöntemi uygulanmalıdır (London, 1985).

\* Sos. Sig. Uzm. Yard. Hatice Tuzgöl, SSK Başkanlığı, Finansman, Aktüerya ve Planlama Daire Başk., 06437, Sıhhiye/Ankara (htuzgol@ssk.gov.tr)

\*\* Prof. Dr. Süleyman Günay, Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06532, Beytepe/Ankara (sgunay@hacettepe.edu.tr)

Ölümün ölçüleri, gözlemlenmiş oranlar ve bunlardan elde edilen düzeltilmiş oranlar olarak ikiye ayrılabilir. Gözlemlenmiş oranlar genellikle, hesaplanması basit oranlardır ve kısa bir işlem ile doğrudan asıl veriden hesaplanır. Düzeltilmiş oranların hem hesaplanması hem de yorumu gözlemlenmiş değerlere göre daha karmaşıktır. Bunlar; verilen bir nüfus grubunun özel ölüm oranları dizisine dayandırılan özet ölçülerini çıkarmak için çeşitli varsayımları içeren ölüm seviyesinin kuramsal göstergeleridir (Shryock, 1971).

Ölüm tabloları hazırlanırken iki farklı veri tabanı kullanılabilir. Bunlardan birincisi, ölüm oranlarının aynı kaldığı bir yıl, üç yıl ya da iki nüfus sayımı arasındaki bir dönem (intercensal) gibi kısa bir sürenin deneyimine dayandırılmaktadır. Bunlara şimdiki zamana ait ya da dönemsel ölüm tabloları (the current or period life table) adı verilir. Dönemsel ölüm tabloları şu anki ölüm oranlarının bir “şip şak fotoğrafı” olarak görülebilir. Bu tablolar, bir yıldaki ya da kısa bir süre için hesaplanan ölüm oranının çok iyi bir özet tanımını sunmaktadır. Kullanılan ikinci veri tabanı ise belirli bir doğum neslinin, örneğin 1900 yılında doğan tüm kişilerin, deneyimlerinden elde edilen ölüm oranları üzerine dayandırılmaktadır. Bu şekilde hazırlanan tablolara nesil ya da kuşak tablosu (the generation and cohort life table) adı verilir. Buna göre nesildeki kişilerin ölüm deneyimleri, doğumlarından itibaren birbirini izleyen takvim yıllarında ardışık her yaş için ölüme kadar gözlenebilir. Açıkça, tek bir tablonun tamamlanması için uzun yılların verisi gerekmektedir ve günümüzde nesil tablolarını oluşturmak mevcut veriler ile olanaksızdır. Bu tür tablolar, ölüm oranlarının projeksiyonu, trend çalışmaları ve doğurganlık ölçümleri için yararlıdır (Shryock, 1971).

Ölüm tablosu oluşturulurken ham ölüm oranları elde edildikten sonra ikinci bir işlem bu oranların düzeltilmesidir. Ölüm oranları dizisinde, verilerin birbiriyle olan ilişkileri nedeniyle aşırı değerlerin düzeltilmesi gerekmektedir. Bu yöntem düzeltme yöntemi (graduation method) adı verilir. Düzeltmenin amacı, ardışık yaşlar için ölüm oranlarının “en olası” değerlerinin bir tablosunu üretmektir. Bu işlem, ham ölüm istatistiklerinin düzleştirilmesi ile yapılır (Verrall, 1993).

Ham veriler, yaşlar itibarıyla “ham” bir şekilde gelişme eğilimindedir (London, 1985). Çünkü;

- Ham ölüm oranları her yaş için ayrı ayrı ve birbirinden bağımsız olarak tahmin edilmekte ve
- Verilerin örnekleme ile seçilmesi durumunda, örnekleme hatası nedeniyle bir yan ortaya çıkabilmektedir.

Genel olarak dört düzeltme yöntemi bulunmaktadır. Bunlar: Grafikselsel düzeltme, standart bir tabloyu temel alan düzeltme, parametrik düzeltme ve parametrik olmayan düzeltme yöntemleridir. Bu yöntemlerden bazıları: Makeham-Gompertz (parametrik), Whittaker (parametrik olmayan) ve Bayesci düzeltme yöntemleridir.

Hiçbir düzeltme yöntemi için “en iyi yöntem” ya da “en doğru yöntem” ifadeleri kullanılmamaktadır. Hangi yöntemin kullanılacağı, araştırmacının amacının ne olduğuna ve nasıl bir veri kümesi ile çalışıldığına göre değişmektedir. Ancak belirli



durumlarda bir yöntem diğerine tercih edilebilmektedir. Yöntemin seçimini etkileyen önemli faktörler, verinin biçimi ve büyüklüğüdür. Bilgisayar teknolojisinin çok gelişmediği dönemlerde hesaplama yönteminin kolaylığı yöntem seçiminde önemli bir etken olarak görülmekte idi oysa artık günümüzde bu çok önemli değildir. Genellikle tek bir yöntemin önerilmemesine karşın Whittaker yönteminin çok yönlülüğü vurgulanmaktadır. Whittaker yöntemi, ölümlülük çalışmalarını da içeren çeşitli türdeki verilerle kullanılmaktadır. Ayrıca, geçmişten günümüze Whittaker düzeltme yönteminin iyi sonuçlar verdiği de söylenmektedir (London, 1985).

Bu çalışmada, dağılım varsayımı içermemesi nedeniyle parametrik olmayan düzeltme sınıfında yer alan ve kolay anlaşılabilir bir yapıya sahip olan Whittaker düzeltme yöntemi incelenecektir.

### 1.1 Düzeltmenin Amacı ve İstatistiksel Temeli

Bir aktüer için, sigorta değerlerinin hesaplandığı yaşam modellerinin oluşturulması önemlidir. Bu modeller, genellikle, Tablo biçiminde gösterilmekte ve “Ölüm Tabloları” olarak ifade edilmektedir (Daha iyimser bir ifade olması açısından “Yaşam Tabloları” da denilmektedir).

Model kurma sürecinde ilk aşama, gözlem altındaki bir grubun yaş bazında bir ölümlülük oranları dizisi yani “gözlemlenmiş dizi” elde edilmesidir. Bu oranların elde edilmesinden sonra, model kurmanın ikinci aşaması olan düzeltme süreci başlar. Verilen başlangıç verilerinden bilinmeyen temel olan ölümlülük örüntüsünü daha iyi temsil etmek amacıyla, sistematik bir revizyona başlanır. Düzeltme süreci de tamamlandıktan sonra, model kurmada üçüncü bir aşamadan söz edilmektedir. Bu aşama ölüm oranlarının projeksiyonudur. Ölümlülük oranları zamanla değişme - genellikle azalış - gösterdiğinden bu gibi değişikliklerin modele yansıtılması gerekmektedir. Çünkü, model geleceğe dönük uygulanabilir ekonomik ve demografik değerlerin hesaplanmasında kullanılabilir.

Ölüm oranları için yaşlar itibariyle düzenli gelişen bir tahmin kümesinin olması istenir. Herhangi bir yaştaki ham ölüm oranı tahmini, bir sonraki yaştaki değerler hakkında bilgi verir. Düzeltme formülündeki düzgünlük ölçütü, her yaştaki tahmini “geliştirmek” için, yakın yaşların verisinden yararlanmayı sağlamaktadır. Bu noktada şöyle bir soru sorulabilir: İlk verilere sahipken, neden bunlar düzeltilmelidir? Neden bu değerlerin, tahmin edilmek istenen bilinmeyen değerler için en iyi tahmin olduğu düşünülmez. Eğer tahmin yöntemi uygun ve geçerli ise neden değiştirilmektedir? Bu durum şöyle açıklanabilir:

Ölüm oranları, her yaş için ayrı ayrı tahmin edilerek bulunduktan sonra bu oranlara kullanılabilir olması amacıyla bir takım değişiklikler yapılmaktadır. Bu nedenle, oranların yaşlar itibariyle “düzgün bir şekilde” değiştiği beklenebilir. Bunun anlamı, oranların düzensiz bir seyir izlememesinin beklenmesi değildir, asıl istenen şey, oranların yaşlar itibariyle düzgün bir seyir izlemesidir. Bu nedenle ham ölüm oranları, aktüerlerin düzeltme olarak adlandırdığı yöntem kullanılarak düzeltilmektedir. Bir veri kümesi düzeltilirken, verinin içerdiği belirleyici özelliklerini yansıtmak şekilde düzeltmeye, beklenenin üzerinde ya da altında değerler elde etmemeye ya da açıklaması

güç yanlı bir yapının oluşmamasına dikkat edilmelidir. Düzeltme için iki yaklaşım bulunmaktadır. Bunlardan ilki, ölüm sayıları için bir olabilirlik fonksiyonunu bulmaktır. Diğer bir ifadeyle, her yaş için ölüm sayılarının ( $d_x$ ) dağılımına ya da ham ölüm oranlarının ( $\hat{q}_x$ ) dağılımına bakmaktır. Ölüm sayılarını kullanmak daha iyi sonuçlar vermesine rağmen genellikle ham ölüm oranları ve ağırlıklandırılmış en küçük kareler yaklaşımı kullanılmaktadır (<http://staff.city.ac.uk/r.j.verrall/dip104.html>, 2004).

## 2. WHITTAKER DÜZELTME YÖNTEMİ

Whittaker Düzeltme Yöntemi E.T. Whittaker (1923) tarafından geliştirilmiştir. Henderson (1924 ve 1925) Whittaker kuramına ve “Kuram uygulamaya nasıl geçirilir?” sorusuna önemli bir katkıda bulunmuştur. Bu yöntemi ilk bulan Whittaker olsa da, bu yöntem Whittaker - Henderson olarak bilinmektedir (London, 1985). Whittaker düzeltme yöntemi, parametrik olmayan bir yöntemidir.

Whittaker yönteminin uzun zamandır bilinmesi ve hala geçerli olması parametrik olmayan ve kolay anlaşılır bir yöntem olmasına bağlıdır. Yöntemi uygulayan kişi, ne yapabileceği ve yapamayacağı hakkında iyi bir fikir sahibidir. Bunlara ek olarak, Whittaker yönteminde, Bayesci yaklaşım kullanılarak bir istatistiksel türetme söz konusudur. Türetme problemi, tam bir türetmeden çok, bir tahmine dayanmaktadır. Bu, düzeltmeyi yapan kişiye bir karar vermeden önce bilinmesi gereken koşulları vermemektedir. Bu durumu göz önünde bulundurmadan düzeltme, düzgünleştirme parametreleri seçilene kadar tamamlanamaz ve bu her zaman analizi yapan kişinin kararına dayanmaktadır (Carlin ve Klugman, 1993).

Whittaker - Henderson düzeltmesi düzgünlük (smoothness) ve uyum (goodness of fit) olmak üzere iki ana ölçüt içermektedir. Bu iki ölçüt arasında uygun bir dengenin sağlanması gerekmektedir. Bu denge, araştırmacının düzeltilmiş verilerde ne düzeyde uyum ve ne düzeyde düzgünlük elde etmek istediğine bağlı olarak değişmektedir (Taylor, 1992).

### 2.1. Temel Whittaker Düzeltme Formülü

Temel Whittaker yöntemi, temel mantığı ileriki bölümlerde söz edilecek olan uyum ve düzgünlük ölçütlerinden gelmektedir. Bölüm 2.1.1 ve 2.1.2’de verilecek olan uyum ve düzgünlük ölçüleri için Eş. (2.3) ve Eş. (2.4) kullanılmaktadır. Bu ölçüler doğrusal bir şekilde bir araya getirilerek aşağıdaki gibi formüle edilebilir:

$$M = F + hS = \sum_{x=1}^n w_x (\tilde{q}_x - \hat{q}_x)^2 + h \sum_{x=1}^{n-z} (\Delta^z \tilde{q}_x)^2 \quad (2.1)$$

Düzeltilmiş değerler,  $\tilde{q}_x$ ,  $x = 1, 2, \dots, n$ , M ölçüsünü minimize eden değerler olarak alınır. Düzeltme sürecinde kullanılan temel notasyonlar aşağıdaki gibidir:

1.  $x$  genellikle yaşı göstermektedir ve 1’den  $n$ ’e kadar alınmıştır.
2.  $t_x$ , gerçek değerler dizisidir.
3.  $\hat{q}_x$ ,  $t_x$ ’in ilk tahmini ya da gözlenmiş değerlerdir.



4.  $\hat{Q}_x, t_x$  'in raslantı değişkenler tahmin edicisidir, bunların yansız binom oranları oldukları varsayılmaktadır.
5.  $\tilde{q}_x, t_x$  'in gözden geçirilmiş tahminleri yani düzeltilmiş tahminlerdir.
6.  $\tilde{Q}_x, \tilde{q}_x$  'in raslantı değişkenleridir.
7.  $z$  parametresi polinom derecesini oluşturan bir parametredir, genellikle araştırmacılar,  $z$ 'yi önceki deneyimlerine dayanarak 2, 3 ya da 4 olarak belirlemektedirler.
8. Sonlu farkların indisinde genel kullanım olan  $\Delta^z \tilde{q}_x$  'in en yüksek indekslenmiş değeri,  $x = n-z$ 'dir. Böylece  $S$ 'de  $F$ 'den daha az terim vardır.  $\Delta$ , her  $\tilde{q}_x$  için  $\Delta \tilde{q}_x = \tilde{q}_{x+1} - \tilde{q}_x$  olarak tanımlanan ileri fark alma operatörüdür, burada  $z$  fark alma operatörünün uygulama sayısını göstermektedir (Guerrero, Juarez ve Poncela, 2001).
9.  $w_x$  ağırlıkları pozitif gerçel sayılardır.
10.  $M$ 'nin minimizasyonunda gerçel bir sayı olan  $h$  parametresi,  $F$  ve  $S$  ile yakından ilişkilidir.
11.  $E_x$ , raslantı hata değişkenler tahminidir.  $e_x$  ise  $E_x$  'in bir gerçekleşmesidir.
12. Geçmişten günümüze, Eş. (2.1)'in minimizasyonu "B tipi" Whittaker düzeltmesi olarak gösterilir. "A tipi" Eş. (2.1)'in özel bir durumu olarak gösterilir, bu durumda  $w_x$  her  $x$  için 1'dir. Her iki tip için farklı minimizasyon süreci kullanılmaktadır (London, 1985).

Genel olarak, düzeltme sürecinde ağırlıklar kullanılmakta ve notasyon olarak  $w_x$  benimsenmektedir. Bölüm 2.2.3.'de verilecek olan ağırlıklar çoğunlukla, örneklem büyüklüğünün ( $n_x$ ) bazı fonksiyonlarından oluşmaktadır.

### 2.1.1. Uyum Ölçüsü

Whittaker düzeltme'nin temel ölçütlerinden biri uyum ölçüsüdür. Düzeltilmiş değerlerin ilk değerlerden çok fazla sapmaması istenir. Bu ikisi arasında uyumun bir derecesi ölçülebilir. Örneklem büyüklüğü çok küçük değilse, düzeltilen değerler ilk değerlere yakın değerler olabilir. Binom oranı, raslantı değişken büyüklüğü ile tersine bir varyansa sahip olduğundan düzeltilmiş bir değer, "büyük  $n$ " durumları için, gözlemlenmiş değerden daha az sapabilir. Dolayısıyla istatistiksel olarak, ilk tahminlerin daha büyük  $n$ 'e dayandırılarak düzeltilmesi sonucunda "daha iyi uyum" (daha küçük sapmaların) ortaya çıkacağı belirtilmektedir (London, 1985).

Sayısal uyum ölçüsü,

$$F_1 = \sum_x w_x (\tilde{q}_x - \hat{q}_x) \quad \text{olarak tanımlanmaktadır.} \quad (2.2)$$

Eş. (2.2)'e göre uyum ölçüsünde, sapmalar ağırlıklandırılarak örneklem büyüklüğü yansıtılmaktadır.  $F$  değerinin sıfır olması uyumun iyi olduğunu gösterdiğinden, büyük ağırlıklar ile sapmalar kendileri ile karşılaştırmalı olarak  $F$ 'in küçük kalmasını sağlamak için daha küçük olmalıdır. Eş. (2.2), yetersiz bir uyum ölçüsüdür. Çünkü, rasgele pozitif ve negatif ağırlıklı sapmaların birbirini götürmesi



sonucu  $F$ 'in değeri sıfır çıkacak ve bu uyumun çok zayıf olduğu yönünde değerlendirilecektir. Bu durum göz önünde bulundurularak uyum ölçüsü, aşağıdaki gibi ağırlıklı sapmalar karesi toplamı olarak tanımlanabilir (London, 1985):

$$F_2 = \sum_x w_x (\tilde{q}_x - \hat{q}_x)^2 \quad (2.3)$$

### 2.1.2. Düzgünlük Ölçüsü

Whittaker düzeltmenin temel ölçütlerinden diğeri ise düzgünlük ölçüsüdür. Düzeltmiş değerlerin grafiği çizilerek, sağlanmış düzgünlüğün derecesi gözlenebilir. Alternatif olarak düzgünlüğün sayısal ölçüsü hesaplanabilir. Bu genellekle, düzeltilmiş değerlerin son farkların birkaç sırası hesaplanarak yapılmaktadır. Eğer bu farkların bazı sıraları (genellikle 3 ya da 4) “küçük” ise, düzlemin derecesi elde edilebilir. Düzlemeyi göstermek için tek indeks numarası, bu farkların karesel toplamı alınarak elde edilebilir. Bu indeks tek başına bir anlam ifade etmeyebilir ancak çeşitli düzeltilmemiş ya da düzeltilmiş verilerin karşılaştırılmasında kullanılabilir (London, 1985).

$$S = \sum_{x=1}^{n-z} (\Delta^z \tilde{q}_x)^2 \quad (2.4)$$

Eş. (2.4) düzgünlük ölçüsü olarak kullanılmaktadır.  $z$  burada fark alma derecesini göstermekte,  $\tilde{q}_x$ ,  $x$  indeks değeri ile düzeltilmiş değeri ifade etmektedir. Örneğin  $z = 4$  için, eğer düzeltilmiş değerler kübik eğri biçiminde ise, bu değer alacağı değer sıfır olacağından dolayı düzlemin bir standardı olarak 3. derece polinom düşünülebilir.  $S$ 'yi olabildiğince küçük yapmak, düzeltilmiş değerleri kübik bir biçime yaklaştıracaktır. Ancak yaşlar itibariyle ölümlülük oranlarının, kübik dağılıma değil, tartışılan yaşın geniş aralığında, Makeham-türü üstel biçiminde bir dağılıma sahip olduğu iddia edilmektedir. Bu durumda farklı bir  $S$  kullanılması gerekmektedir. Üzerinde çalışılan veri kübik eğri biçiminde değilse, 3. fark yerine 4. fark tasarlanabilir. Genellikle 3. ve 4. farklar kullanılmaktadır. Alternatif olarak Giesecke (1984), çalışmasında regresyon kullanarak farklı bir düzgünlük ölçüsü geliştirmiştir. Düzgünlük ölçüsünde bulunan  $k$  matrisi için farklı seçenekler sunmuş ve sonuçları karşılaştırmıştır. Bu seçenek matrisler içerdiği satır sayısına, her bir satırdaki en fazla tahmin edici sayısına ve her bir satırda kullanılan en yüksek derecede polinom özelliklerine göre tanımlanmış matrislerdir. Farklı seçenekte düzgünlük matrislerinin kullanılmasının bir yararı, temel düzeltme yönteminin sonucu ile karşılaştırıp hangisinin iyi olduğuna karar vermede bir seçenek oluşturmasıdır. Bir başka faydası, düzeltmenin farklı bölümlerinde, farklı derecede eğriler kullanılmak istenmesi durumunda görülmektedir.

### 2.1.3. Ağırlıklar

Whittaker yöntemi, Eş. (2.3)'deki uyum ölçüsüne dayandırılmaktadır. Düzeltme uygulandıktan sonra uyumun derecesini görmek için bir uyum ölçüsü hesaplanmalıdır.

Ancak Eş. (2.3)'deki ağırlıklarda  $w_x = \frac{n_x}{\tilde{q}_x(1-\tilde{q}_x)}$  formülünü kullanabilmek için önce

$\tilde{q}_x$  değerlerinin bilinmesi gerektiğinden, bu durum  $M$ 'nin minimizasyon

hesaplamalarını güçleştirmektedir. Genellikle Eş. (2.3)'de ağırlıklar sadece örneklem büyüklüğü  $n_x$  olarak alınmaktadır. Bu F'in değerinin büyük çıkmasına yol açacak, dolayısıyla eğrinin düzgünlüğüne daha az ve eğrinin uyumuna gereğinden fazla önem vermekten kaçınmak için, h'in çok büyük bir değeri kullanılmalıdır. Seçenek olarak,

$$w_x = \frac{n_x}{\bar{n}} \quad (2.5)$$

sunulmaktadır, Eş. (2.5)'de  $\bar{n}$  her x için  $n_x$  'lerin ortalamasını göstermektedir. Burada amaç, daha büyük  $n_x$  'e sahip yaşlara, oransal olarak daha çok ağırlık vermektir.

#### 2.1.4. h parametresi

h, düzeltme sürecinde düzgünlüğü ayarlayan bir parametredir. h'in seçimine yönelik şu gözlemler yapılabilir:

1. S'nin değeri ne olursa olsun,  $h = 0$  ise  $hS = 0$ 'dır. O zaman  $F > 0$  olduğundan, M,  $F = 0$  için minimize edilir, bu düzeltme olmaması yani her x için  $\tilde{q}_x = \hat{q}_x$  olması anlamına gelmektedir. Böylece, genellikle h, 0'a yaklaştıkça,  $\tilde{q}_x$ ,  $\hat{q}_x$  'e yaklaşır ve uyum düzgünlük üzerinde belirlenir.
2. Diğer taraftan h çok büyük olduğunda minimizasyon süreci S'yi, sezgisel olarak büyük h'in etkisini gidermeye yönelik belirlemektedir.

Düzeltilme sürecinde h parametresi, bu durumlar dikkate alınarak belirlenmelidir.

#### 2.1.5. Whittaker Düzeltme Yönetiminin Olumlu ve Olumsuz Yanları

Yöntemin üstünlükleri:

1. Eğer verinin parametrik bir formül kullanılarak düzeltilmesi zor ise, kullanışlıdır.
2. Esnektir.
3. Ölüm oranlarından başka, örneğin hastalık ile ilgili bir veri kümesinin düzeltilmesinde de kullanılabilir.

Yöntemin olumsuz yanları ise şunlardır:

1. Veriyi ne kadar düzgünleştireceğini ayarlamak zordur.
2. Whittaker düzeltmenin bir diğer güçlüğü de, h düzgünleştirme parametresinin seçimine yol göstermede kuramsal olarak yetersiz kalmasıdır (Taylor, 1992).

#### 2.2. M fonksiyonunun minimizasyonu

M,  $\tilde{q}_x$  'in n bilinmeyen değerinin bir fonksiyonudur. M'nin minimizasyon değerini her bir  $\tilde{q}_x$  'e göre sıfır yapan  $\tilde{q}_x$  değerleri, M'nin kısmi türevlerinin sıfıra eşitlemesi ile elde edilen n denklemin çözümüdür (London, 1985).

Bu denklem,

$$\frac{\partial M}{\partial \tilde{q}_r} = 0, r = 1, 2, \dots, n \text{ olur.} \quad (2.6)$$



Buradan  $\tilde{q}$  değerleri,  $(I + hw^{-1}k'_z k_z)^{-1} \hat{q}$  'nun çözümü ile elde edilir (Taylor, 1992). İlk tahminler ( $\hat{q}$ ) ve düzeltilmiş tahminler ( $\tilde{q}$ ) vektörlerinin transpozları,  $\hat{q}' = [\hat{q}_1, \dots, \hat{q}_n]$  ve  $\tilde{q}' = [\tilde{q}_1, \dots, \tilde{q}_n]$  vektörleri ile,

$$\text{ve ağırlıkları içeren } n \times n \text{ köşegen matris, } w = \begin{bmatrix} w_1 & & & 0 \\ & \cdot & & \\ & & \cdot & \\ 0 & & & w_n \end{bmatrix} \text{ ile gösterilir,}$$

$k_z$ ,  $z$  derecesinin binom katsayılarını içeren özel bir matrisi ve  $k_z \tilde{q}$ ,  $\Delta^2 \tilde{q}_x$  değerlerini içeren bir çarpım vektörünü göstermektedir. Eğer  $\tilde{q}_x$  'in  $n$  değeri var ise  $\tilde{q}$ ,  $n \times 1$  bir vektör, bu durumda  $k_z$ ,  $(n - z) \times n$  köşegen olacak ve  $k_z \tilde{q}$   $(n - z) \times 1$  boyutlu bir vektör olacaktır.

Guerrero (2001) tarafından  $k_z$  matrisi şöyle tanımlanmaktadır:

$k_z$ ,  $(n - z) \times n$  matrisinin  $j$ . elementi,

$j < i$  ya da  $j > z + i$  için,  $k_z(i, j) = 0$ ,

$i = 1, \dots, n - z$  ve  $j = 1, \dots, n$  için,

$$k_z(i, j) = \frac{(-1)^{z+i-j} \cdot z!}{[(j-i)!(z-j+i)!]} \text{ 'dir.}$$

Örneğin,  $z = 2, n = 6$  için  $k_2$  matrisi ve  $z = 3, n = 7$  için  $k_3$  matrisi,

$$k_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}, k_3 = \begin{bmatrix} -1 & 3 & -3 & 1 & 0 & 0 & 0 \\ 0 & -1 & 3 & -3 & 1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -3 & 1 & 0 \\ 0 & 0 & 0 & -1 & 3 & -3 & 1 \end{bmatrix}$$

olarak gösterilir ve  $k_2 \tilde{q}$  ve  $k_3 \tilde{q}$  sonuçları sırasıyla  $\Delta^2 \tilde{q}_x$  ve  $\Delta^3 \tilde{q}_x$  değerlerinin istenen değerlerini oluşturmaktadır.

### 2.3. Bayesci Yaklaşım

Aktüerya da önsel bilginin elemanları ölüm oranlarının düzeltilmesinde çok sık kullanmakta, bu oranların düzgün diziden elde edilen gerçek oranlar olduğuna inanılmaktadır. Başlangıçta var olan sezgisel kanı, kesinlikle deneysel kanıtlarla desteklenmektedir.

Düzeltilme yönteminin orijinal gelişimi, Whittaker tarafından ve tekrar Whittaker ve Robinson tarafından verilmiştir. Yöntem, biçimsel olarak bir Bayes yaklaşımına benzemektedir. Whittaker düzeltme ile yapılan çalışmalar arasında Bayesci Whittaker Düzeltilmeye de rastlanmaktadır. Broffitt (1992) Bayesci riskin minimizasyonunda Whittaker düzeltmeyi incelemiştir. Taylor (1992) Whittaker düzeltmeyi, düzgünlük katsayısını etkin şekilde belirleyebilen Bayesci bir bağlamda sunmuştur. Whittaker



düzeltilme ile Bayesci yöntemler arasındaki ilişki Whittaker ve Robinson (1944) tarafından sunulan Whittaker türetmede ve Whittaker düzeltmeye ilişkin özel bir uygulama, Gersh ve Kitigawa (1988) tarafından verilmiştir (Carlin ve Klugman, 1993). Kimeldorf ve Jones (1967) Whittaker düzeltmeyi Bayesci bir açıdan ele almışlardır (Guerrero, Juarez ve Poncela, 2001). Yine Carlin ve Klugman (1993), önsel ve sonsal dağılımlar yardımı ile Sıralı Bayesci Whittaker yöntemini incelemiştir.

### 3. UYGULAMA

Yaşlar itibariyle oluşturulan ölüm olasılıkları tablosuna, “ölüm tablosu” adı verilir. Ölüm tabloları, hayat sigortası ve emeklilik gelirlerinin ücretlendirilmesinde ve karşılıklarının değerlendirilmesinde kullanılır.

Devlet İstatistik Enstitüsü'nün derlemekte olduğu il ve ilçe merkezleri ölüm ve nüfus sayıları dikkate alınarak oluşturulan ham ölüm oranlarına, Whittaker düzeltme yönteminin uygulanması ve bu düzeltilmiş değerlerden cinsiyet bazında ölüm tablolarının oluşturulması amaçlanmaktadır. Bu uygulama için, 1990 – 2000 yıllarına ilişkin veriler ele alınmıştır. Temel Whittaker biçimi değişim önerisine göre bir Excel minimizasyon programı hazırlanmıştır. Oluşturulan ham ölüm oranları ile bir uygulama yapılmıştır.

#### 3.1. Ham Ölüm Oranlarının Hesaplanması

Ham ölüm oranlarının hesaplanmasında temel varsayım, Türkiye'nin il ve ilçe merkezi nüfusunun yıllar itibariyle durağan bir seyir izlediğidir.

Diğer Varsayımlar:

1. Elde edilen veriler eksik bilgi değildir.
2. Ölümler son doğum gününde, 1 Ocak – 31 Aralık 1990 ve 1 Ocak – 31 Aralık 2000 takvim yıllarında ortaya çıkmaktadır,
3. Nüfus sayım tarihleri 25 Ekim 1990 ve 25 Ekim 2000 gününde yapılmıştır.
4. 1990 ve 2000 yıllarına ilişkin nüfus istatistikleri, sayım tarihleri dikkate alınarak interpolasyon yaklaşımı ile 1 Ocak için düzenlenmiştir.

Aşağıdaki formüller kullanılarak ham ölüm oranlarına ulaşılmıştır (Jordan, 1967):

$$m_x = \frac{d_x}{L_x}, \quad m_x = \frac{d_x}{l_x - \frac{1}{2}d_x} = \frac{\hat{q}_x}{1 - \frac{1}{2}\hat{q}_x}$$

Ölüm tablosundaki gösterim olan  $m_x$ ,  $L_x$  değerleri bilinmeden hesaplanamaz (Barclay, 1958).

$$\hat{q}_x = \frac{m_x}{1 + \frac{1}{2}m_x}, \quad 1,000\hat{q}_x = 1,000 \times \frac{m_x}{1 + \frac{1}{2}m_x},$$

Hesaplama dikkate alınan nüfus için yaşa özel ölüm oranı  $M_x$  ile gösterilirse,

$M_x = \frac{D_x}{P_x}$  şeklinde yazılabilir. Buradan,

$\hat{q}_x = \frac{M_x}{1 + \frac{1}{2}M_x}$  hesaplanarak ölüm oranı için bir yaklaşım yapılabilmektedir.

Formüllerde bulunan notasyonların tanımları şöyledir:

$m_x$  : x yaşa özel ölüm oranıdır.

$d_x$  : Bir takvim yılında x ve x+1 tam yaşları arasındaki ölümlerin sayısıdır.

$L_x$  :  $L_x$  değerleri, x ve x+1 yaşları arasında ölüm tablosu topluluğu yaşanmış yıllar sayısını göstermektedir. Bu değer, tam olarak hesaplanamamakta dolayısıyla birtakım yaklaşımlar yardımı ile hesaplanabilmektedir. Durağan bir nüfusta, bir takvim yılından diğer yıla sabit kalmaktadır. Türkiye'nin il ve ilçe merkezi nüfusu sabit bir nüfus değildir.

$l_x$  : Bir takvim yılda x yaşına varan kişi sayısıdır. Takvim yılı boyunca x yaşta kabul edilirler. Örneğin doğum günü 19.01.1936 olan bir kişi 01.01.2005 itibariyle hala 68 yaşında olduğu için 68 yaşındaki sayıya dahil edilir.

$M_x$ ,  $D_x$ ,  $P_x$  : Sırasıyla, mevcut veriden hesaplanan ölüm oranı, takvim yılı boyunca kaydedilmiş toplam ölümlerin sayısı (1 Ocak – 31 Aralık), yıl ortasındaki toplam nüfusu göstermektedir. Bu çalışmada yıl başı varsayımı yapılmıştır.

$\hat{q}_x$  :  $M_x$  kullanılarak, bir yaklaşım ile hesaplanan ölüm oranıdır.

$1,000\hat{q}_x$  : 1 yıl içerisinde 1,000 kişiden ölen kişi sayısıdır.

DİE'den elde edilen nüfus istatistikleri 85 yaşa kadar, ölüm istatistikleri ise 98 yaşa kadar düzenlenmiş olup 84 ve 98 yaş sonrası ilişkin bilgi birikimli olarak verilmiştir. Bunun için, öncelikle 84 yaşa kadar yaşa özel ölüm oranları ( $M_x$ ) ve buradan da ham ölüm oranları için bir yaklaşım olan  $\hat{q}_x$  değerleri hesaplanmıştır. Nüfus istatistiklerini 98 yaşına taşıyabilmek için, bir başka ölüm tablosuna ihtiyaç duyulmuştur. Bazı ülkelerin ölüm oranları incelendiğinde, Finlandiya 80 yaş sonrası ölüm oranlarının Türkiye'nin 80'li yaşlardaki ölüm oranları ile genel seyir itibariyle iyi bir uyum gösterdiği grafiksel çizimler ile gözlemlenmiştir. Bu bilgi yardımıyla, 85 ve 100 yaşları arası  $\hat{q}_x$  değerleri için Finlandiya 1990 ve 2000 Erkek ve Kadın için ölüm oranları kullanılarak hesaplanmıştır. Verinin yapısını bozduğu düşüncesiyle elde edilen ham ölüm oranlarının belirli yaşlarında düzeltme yapılmıştır. 100 yaş değerleri ise bu yaş için genel olan değerler göz önünde bulundurularak belirlenmiştir. Elde edilen ham ölüm oranlarına, 1 - 99 yaşları için düzeltme uygulanmıştır, 0 yaş için oranlar çok yüksek olduğu ve düzeltilmiş eğrinin yapısını bozduğu için işleme katılmamış bu yaş için ham ölüm oranları son ölüm oranları olarak kabul edilmiştir.

Gerçekte genel nüfus için hesaplanan ölüm tablosundaki son yaş ile ilgili – bu çalışmada 100- bir yıl içinde ölme olasılığı 1,0'e kadar gitmemektedir. Bu sigorta şirketleri tarafından ölüm tabloları için hesaplama kolaylığı olsun diye yapılmaktadır. 95'li yaşlarda nüfustaki ölüm oranları 0,25'den 0,35'e, olgun yaşlarda ise üstel bir şekilde artış gösterir, ancak büyük yaşlarda azalma eğilimine girer ve belirli bir düzeyde kalır. Bu bir tahmin değil, kuramsal bir sonuçtur (Shiro ve Wilmoth, 1998). Aktüerler



Birliği'nin (Society of Actuaries - www.soa.org) derlediği tablolar benzer örüntü göstermektedir (Lowrie, 2004).

### 3.2. Whittaker Excel Minimizasyon Programı

Yöntemin uygulaması Microsoft Excel'de Temel Whittaker Formülasyonuna göre hazırlanmış olan bir programda yapılmıştır. Bu program Jerome Lamontagne (Canada, Mercer Company) tarafından hazırlanmıştır. Düzeltme sonrası, eğrideki düzensiz yapılar daha düzgün bir eğri haline dönüştürülmekte ve düzeltilmiş değerlerin ilk değerlerle olan uyumu sağlanmaktadır. Programda  $\hat{q}_x$  (gözlemlenmiş değerler) değerleri girildiği zaman,  $\tilde{q}_x$  (düzeltilmiş değerler) değerleri elde edilebilmektedir. Ayrıca farklı  $z$  ve  $h$  parametreleri denenerek uygun olan parametre değerlerinin seçimi, araştırmacının önceki deneyimlerine de dayanarak grafik çizimleri ile kolayca yapılabilmektedir. Aynı şekilde  $w_x$  değerlerinin hesaplanabilmesi için yaş bazında nüfus sayılarının da programda ilgili yere aktarılması gerekmektedir.

Whittaker düzeltme formülüne göre hazırlanmış olan programda düzgünlük parametresi  $h=0,05$ , fark operatörünün derecesi ise 3 alınmıştır. Diğer parametreler için de özellikle  $z = 4$  ve  $z = 5$  için bu yöntem tekrar uygulanmıştır. Grafik çizimleri incelendiğinde en iyi sonucu  $z = 3$  ve  $z = 4$ 'ün verdiği gözlenmiştir. Programda ağırlıklar ( $w_x$ ),  $n_x / (\sum_{x=1}^n n_x)$  formülüne göre hesaplanmaktadır. Yani  $x$  yaş için hesaplanan ağırlık, ilgili yaştaki nüfusun toplam nüfus içindeki payı kadardır. Programdaki hesaplamalar, Bölüm 2.3'de verilen matris ve vektörler kullanılarak yapılmış ve düzeltilmiş değerler elde edilmiştir.

0 ve 100 yaş için geçerli olan  $1,000\hat{q}_x$  ham ölüm oranları düzeltme işlemine katılmamıştır. 0 yaşa ilişkin ham ölüm oranı, ilk birkaç yaş değerini olumsuz etkilediği için düzeltme işlemine katılmamıştır. 100 yaşa ilişkin ham ölüm oranı ise verinin seyrine bakılarak belirlenmiş olduğu için düzeltme işlemine katılmamıştır.

Excel işlemlerde en çok  $52 \times 52$  boyutlu matrise izin verdiği için programın boyutu  $52 \times 52$  olarak hazırlanmıştır. Boyutun 52'yi aşması durumunda program Excel'de çalışmamaktadır. (<http://support.microsoft.com/default.aspx?scid=kb;en-us;166342&Product=xlw>, 2004).

Lowrie (1993), eğer büyük bir veri kümesi düzeltilmek isteniyorsa düzeltmenin, parça parça yapıp, sonra bu parçalı grupların birleştirilmesi ile yapılabileceğinden söz etmektedir. Makalesinde zaman ve yaşı dikkate alan iki boyutlu düzeltme için bir çözüm önermektedir. Bu çözüm tek boyut olması durumunda da kullanılabilir.

Bir büyük gruba düzeltme uygulamak küçük gruplara uygulamaktan daha çok zaman alabilir ya da bilgisayarın limit hacmini aşması nedeniyle çalışmayabilir. Bilgisayar donanımı ve yazılımı için, küçük grupları düzeltip birleştirme yöntemini uygulamak daha uygundur. Bu birleştirme yöntemi, araştırmacılar için bir ileri araştırma konusudur.



100 yaş için elde edilen  $1,000 \hat{q}_x$  değerlerine Excel minimizasyon programında uygulayabilmek için  $52 \times 52$  ve  $51 \times 51$  Excel minimizasyon programları hazırlanmış ve ayrı ayrı uygulama sonuçlarına ulaşılmıştır. Bu iki uygulama sonuçları 49, 50, 51 ve 52 yaşları için ağırlıklar verilerek Lowrie (1993)'nin önerdiği gibi büyük boyutlu verilere bir çözüm olan birleştirme işlemi ile 1-99 yaşları dizisi için düzeltilmiş değerler elde edilmiştir. Bu ağırlıklar  $w_x$  ağırlıkları ile karıştırılmamalıdır. Bu ağırlıklar, iki düzeltme sonuçlarının birleştirilmesi esnasında sadece bu dört yaşa verilen katsayılarıdır.

Düzeltilme sonrası elde edilen ölüm oranları ile 1990 ve 2000 yılları için kadın ve erkek yaşam beklentileri hesaplanmıştır. Sonuçlar Tablo 1. ve Tablo 2.'de verilmiştir. Bu çalışmada, Türkiye için 80 sonrası yaşlarda Finlandiya ölüm oranlarının alındığından söz edilmiştir. Aslında Finlandiya ülkesi gelişmiş bir ülkedir. Ancak ileri yaşlarda uyumun iyi olduğunun gözlenmesi sonucu, bu ülkenin ölüm oranlarının kullanılmasında bir sakınca görülmemiştir.

**Tablo 1.** 1990 yılı Türkiye için kadın ve erkek ortalama yaşam beklentisi

Yaş	Türkiye kadın ortalama yaşam beklentisi	Finlandiya kadın ortalama yaşam beklentisi	Türkiye erkek ortalama yaşam beklentisi	Finlandiya erkek ortalama yaşam beklentisi	Türkiye toplam ortalama yaşam beklentisi
0	73,3	78,8	67,8	70,9	70,5
1	74,7	78,3	69,4	70,3	72,0
5	71,1	74,3	65,9	66,4	68,5
10	66,2	69,4	61,0	61,5	63,6
20	56,4	59,6	51,2	51,8	53,8
30	46,6	49,9	41,6	42,6	44,1
40	37,0	40,2	32,2	33,5	34,6
50	27,6	30,8	23,4	24,9	25,6
60	19,0	21,9	15,7	17,1	17,4
70	11,6	13,8	9,7	10,7	10,8
80	6,6	7,4	5,9	6,0	6,4
90	3,7	3,6	3,2	3,2	3,6
100	0,0		0,0		0,0

**Tablo 2.** 2000 yılı Türkiye için kadın ve erkek ortalama yaşam beklentisi

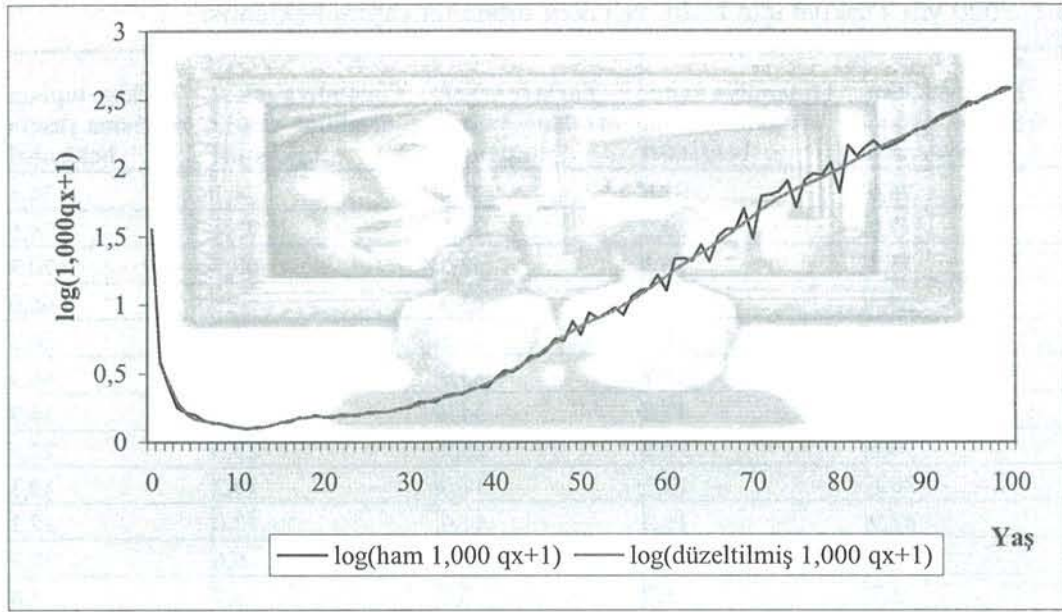
Yaş	Türkiye kadın ortalama yaşam beklentisi	Finlandiya kadın ortalama yaşam beklentisi	Türkiye erkek ortalama yaşam beklentisi	Finlandiya erkek ortalama yaşam beklentisi	Türkiye toplam ortalama yaşam beklentisi
0	76,6	81,0	71,4	74,1	74,1
1	77,0	80,2	71,9	73,4	74,6
5	73,3	76,3	68,2	69,5	70,9
10	68,3	71,3	63,3	64,5	66,0
20	58,5	61,5	53,5	54,8	56,2
30	48,7	51,7	43,9	45,4	46,5
40	39,0	42,0	34,3	36,1	36,8
50	29,5	32,7	25,2	27,3	27,6
60	20,6	23,7	17,3	19,2	19,3
70	12,9	15,2	11,0	12,0	12,3
80	7,3	8,2	6,3	6,6	7,2
90	3,7	3,7	3,3	3,3	3,6
100	0,0		0,0		0,0

Yukarıdaki Tablolarda sunulan yaşam beklentileri incelendiğinde 1990 ve 2000 yıllarında Finlandiya kadın ve erkek ortalama yaşam beklentilerinin Türkiye ortalama yaşam beklentilerinden daha yüksek olduğu gözlenmektedir.

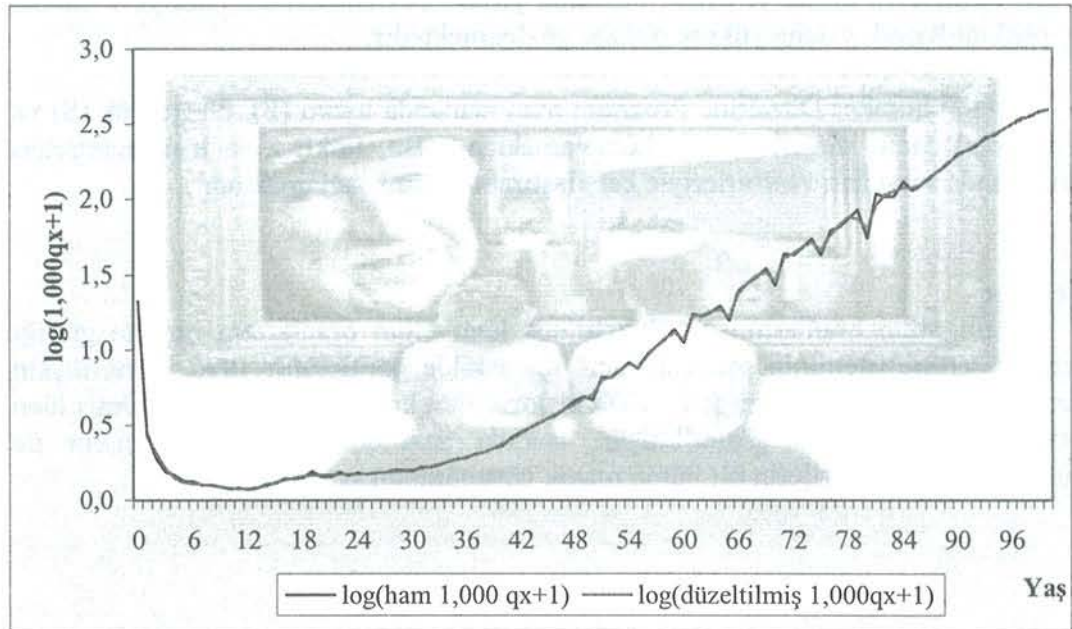
Excel Whittaker Düzeltme Programı aynı zamanda uyum (F), düzgünlük (S) ve M fonksiyonu sonuç değerlerini de hesaplamaktadır. Bu, farklı z ve h parametreleri kullanıldığında sonuçların birbirleriyle karşılaştırabilmesini sağlamaktadır.

### 3.3. Şekiller

Ham ölüm oranlarının ve düzeltilmiş ham ölüm oranlarının birlikte grafiği çizilerek düzeltme işleminin sonuçları açık bir şekilde görülebilir. 1990 yılına ilişkin ulaşılan kadın ve erkek toplamı için 1,000 kişiden ölen kişi sayıları ve düzeltilmiş ölen kişi sayıları Şekil 1.'de gösterilmiştir. Burada şekillerin Logaritmik çizim ile sunulmasının nedeni, eğrilerin bir bütün olarak görünümünü sağlamaktır.

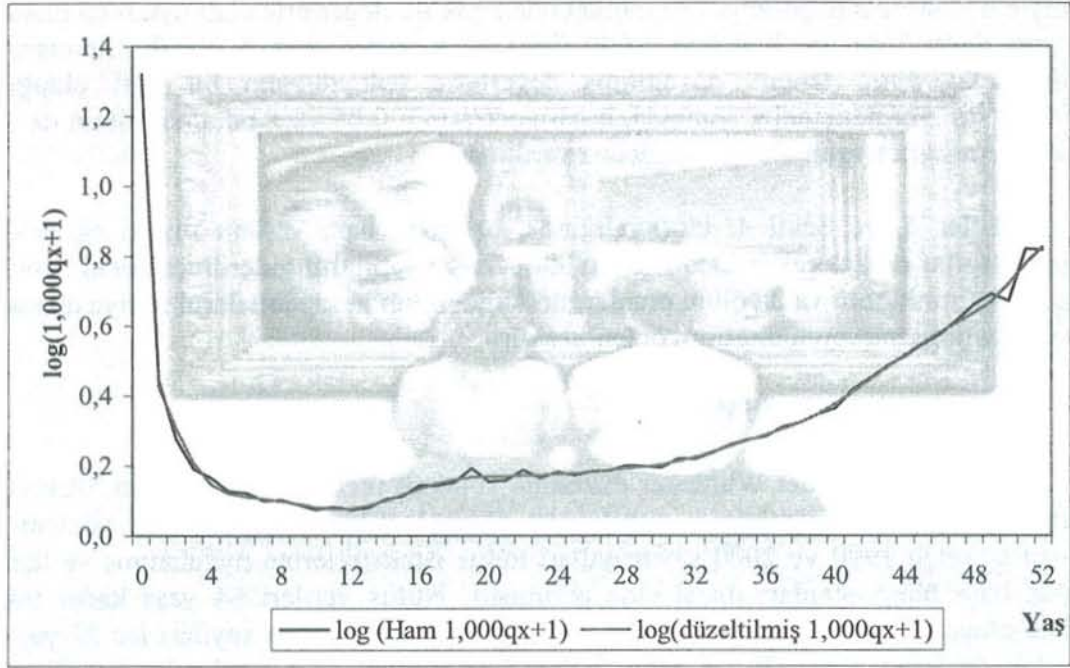


Şekil 1. 1990 yılı kadın ve erkek toplamı için log ham ve düzeltilmiş  $(1,000 \hat{q}_x + 1)$  değerleri ( $h=0,05, z=3$ )

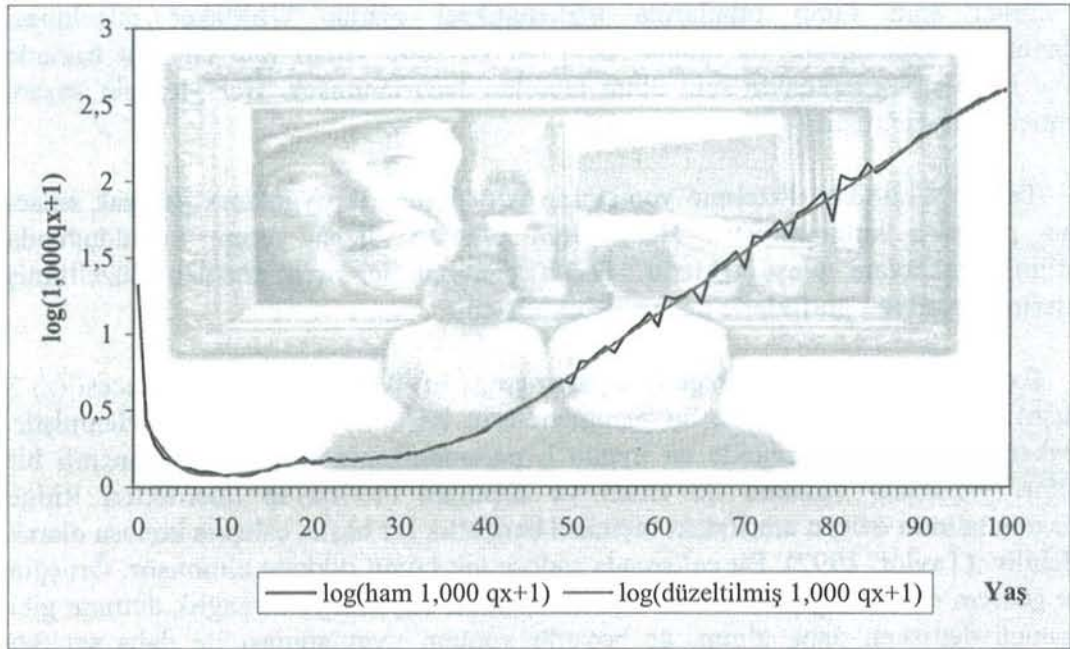


Şekil 2. 2000 yılı kadın ve erkek toplamı için log ham ve düzeltilmiş  $(1,000 \hat{q}_x + 1)$  değerleri ( $h=0,0001, z=4$ )





Şekil 3. 2000 yılı kadın ve erkek toplamı için log ham ve düzeltilmiş  $(1,000 \hat{q}_x + 1)$  değerleri ( $h=1.000, z=5$ )



Şekil 4. 2000 yılı kadın ve erkek toplamı için log ham ve düzeltilmiş  $(1,000 \hat{q}_x + 1)$  değerleri ( $h=0,05, z=3$ )

Birbirleriyle karşılaştırılabilmesi için, Şekil 2 ve Şekil 3’de farklı  $h$  ve  $z$  parametreleri kullanılmıştır. Şekil 2’de  $h$  parametresi 0,0001 ve  $z$  parametresi 4 olarak alınmıştır. Şekil 2. dikkatlice incelendiğinde, düzeltilmiş değerlerin ham ölüm oranlarına çok yakın değerler olduğu görülmektedir.  $h$  değerinin sıfıra çok yakın olması

nedeniyle düzeltilmiş değerlerin düzgünlüğünden çok ilk değerlerle olan uyum ön plana çıkmıştır. Şekil 3’de ise,  $h$  değeri 1,000 (bin) ve  $z$  parametresi 5 olarak alınmıştır. Burada Şekil 2’nin tersine düzeltilmiş değerlerin çok düzgün bir eğri olduğu görülmektedir. Bu denemeler sonunda,  $h$  parametresinin 0,05 ve  $z$  parametresinin de 3 olarak alınmasının uygun olduğu sonucuna varılmıştır.

Şekiller 1. ve Şekil 4. incelendiğinde, özellikle bazı yaşlara ilişkin eğrideki düzensiz yapılar dikkat çekmektedir. Bu gibi düzensiz yapıların giderilmesindeki temel amaç, ölüm oranlarının ya da ölüm oranlarının kullanıldığı hesaplamaların ileriye dönük tutarsız tahminlerinin bulunmasının önlenmesidir.

#### 4. SONUÇ VE TARTIŞMA

Bu çalışmada temel Whittaker düzeltme yöntemi incelenmiş ve bunun Türkiye verileri üzerinde bir uygulaması yapılmıştır. Temel Whittaker düzeltme yöntemi, DİE’den edinilen 1990 ve 2000 sayım yılları nüfus istatistiklerine uygulanmış ve tam yaşlarda ham ölüm oranları dizisi elde edilmiştir. Nüfus verileri 84 yaşa kadar tek yaşlarda olup, 85 ve sonrası için toplam sayı verilmektedir. Ölüm sayıları ise 97 yaşa kadar tek yaşlarda olup, 98 ve daha fazlası için toplam sayı verilmektedir. Ölüm istatistikleri sadece il ve ilçe merkezlerini içerdiği için nüfus istatistikleri de il ve ilçe merkezlerini içerecek şekilde elde edilmiştir. Yıllık ölüm oranları 84 yaşa kadar hesaplanmıştır. 84 sonrası yaşların hesabında ise Türkiye’nin 80’li yaşlarına iyi bir uyum sağladığı gözlenen Finlandiya ülkesinin oranları kullanılmıştır. 0-100 yaşları için elde edilen ham ölüm oranlarına matematiksel olarak Whittaker düzeltmesi uygulanmıştır. Düzeltilmiş bu oranlar ile 1990 ve 2000 yılları için cinsiyet bazında Türkiye il ve ilçe merkezleri için ölüm tabloları hazırlanmıştır. Her yaş için yaşam beklentisi hesaplanmıştır.

Temel Whittaker düzeltme yöntemine göre Excel’de uygulama yapmak amacı ile bir program kullanılmıştır. Ham ölüm oranları dizisi yerine konulduğunda düzeltilmiş sonuçlara ulaşılmaktadır. Farklı parametreler için ulaşılan düzeltilmiş değerlerin uygunluğu grafiklerle de görülebilmektedir.

Excel programlarında düzgünlük parametresi ( $h$ ) 0,05 ve polinom derecesi ( $z$ ) 3 alınmıştır. Düzeltme sonucunda bu parametrelerin iyi sonuçlar verdiği gözlenmiştir. Whittaker düzeltme yönteminde en uygun  $h$  parametresinin belirlenmesi önemli bir sorundur. Whittaker çözümü ile Hoerl ve Kennard (1970a)’ın önerdikleri Ridge Regression tahmin edicisi arasındaki biçimsel benzerlik bir başka çalışma konusu olarak önerilebilir (Taylor, 1992). Bu çalışmada sadece tek boyut dikkate alınmıştır. Örneğin yaş ve gözlem süresi değişkenleri dikkate alınarak iki boyutlu ya da sağlık durumu gibi bir üçüncü değişken daha alınıp, üç boyutlu yöntem uygulanması ile daha sağlıklı sonuçlara ulaşılabilir. Bir diğer araştırma konusu, büyük grupların düzeltilmek istenmesi durumunda, bu grubun parçalı küçük gruplara parçalanıp yapılması üzerinedir. Bu çalışmada Excel programında en fazla 52 boyutluk düzeltme yapılabilmesi nedeniyle 1-99 yaş gruplarının düzeltilmesi 52-52 ve 51-51 boyutluk düzeltmeler uygulanmış ve bu iki düzeltme sonuçları birleştirilmiştir. Bu hesaplamalar MATLAB kullanılarak da yapılabilir.



Ölüm oranları çalışmalarında göç, ölüm ve nüfus istatistiklerinin derlenip derlenmediği ve nasıl derlendiği çok büyük önem taşımaktadır. Bu çalışmada 1990 ve 2000 yılı nüfus ve ölüm istatistikleri dikkate alınmıştır. Ölüm istatistikleri sadece il ve ilçe merkezlerini kapsadığı için nüfus istatistikleri de il ve ilçe merkezlerini kapsayacak şekilde temin edilmiştir. Köy ve bucaklara ilişkin ölüm kayıtları bulunmadığından, Türkiye geneli için bir ölüm tablosu oluşturmak çok zordur. Bu kayıtların derlenmesine yönelik çalışmaların olabildiğince çabuk tamamlanması, çok yararlı olacaktır. 1990 yılı Türkiye il ve ilçe merkezleri için 0 yaşta yaşam beklentisi kadın için 73,3, erkek için 67,8, kadın ve erkek toplamı için 70,5 yine 2000 yılı 0 yaşta kadın yaşam beklentisi 76,6, erkek için 71,4, kadın ve erkek toplamı için 74,1 sonuçlarına ulaşılmıştır. Sonuçlar incelendiğinde, 10 yılda kadın ve erkek yaşam beklentisinde olumlu yönde bir gelişme olduğu, yani yaşam süresinin uzadığı kolayca görülmektedir.

### TEŞEKKÜR

Çalışmamızın uygulama aşamasındaki yardımları için Assoc. Prof. Sayın Walter B. Lowrie'ye ve Aktüer Hugo Berkouwer'a, verilere ulaşmamızdaki yardımları için de DİE Başkanlığına teşekkürlerimizi sunarız.

### KAYNAKLAR

- BARCLAY, G. W., 1958, Techniques of Population Analysis, Princeton University.
- BROFFITT, D. J., 1992, Refinements Of Whittaker Graduation Through The Minimization Of Bayes Risk, University Of Iowa, Statistics & Actuarial Science.
- CARLIN, P. BRADLEY AND KLUGMAN, A. STUART., 1993, Hierarchical Bayesian Whittaker Graduation, Scandinavian Actuarial Journal. 2: 183 – 196.
- JORDAN, CHESTER WALLACE, 1967, Society Of Actuaries' Text Book On Life Contingencies, Mathematics Williams College.
- GIESECKE, LEE., 1984, Piecewise Regression As A Measure Of Smoothness In Whittaker Graduation. Vol 2, 115 – 128.
- GUERRERO, M. V., JUAREZ R., PONCELA P., 2001, Data Graduation Based on Statistical Time Series Methods, Statistics & Probability Letters, 52, 169 – 175.
- LONDON, D., 1985, Graduation: The Revision Of Estimates. ACTEX, Winsted And Abington, CT.
- LOWRIE, W. B., 1993, Multidimensional Whittaker - Henderson Graduation With Constraints And Mixed Differences, Transactions Of The Society Of Actuaries XLV, 215-255.
- LOWRIE, W. B., 2004, E - Posta Yazışmaları
- SHIRO, H. AND WILMOTH R. J., 1998, Deceleration In The Age Pattern Of Mortality At Older Ages, 35(4), 391-412. [http://www.soa.org/sections/farm/farm\\_horiuchi.html](http://www.soa.org/sections/farm/farm_horiuchi.html) (2004)



SHRYOCK, S. H., JACOB S. S., 1971, The Methods And Materials Of Demography, Volume 1, 2, A United States Department Of Commerce Publication.

TAYLOR, GREG, 1992, A Bayesian Interpretation Of Whittaker-Henderson Graduation, Insurance: Mathematics And Economics 11, 7-16.

VERRALL, R. J., 1993, A State Space Formulation Of Whittaker Graduation, With Extensions, Insurance : Mathematics And Economics, 13, 7-14.

VERRALL, R. J., 2004, Information For The Survival Models Course (Subject 104), Department Of Actuarial Science And Statistics City University.  
<http://staff.city.ac.uk/r.j.verrall/dip104.html>

## **Whittaker Graduation Method and An Application to Turkish Mortality Rates**

### **ABSTRACT**

*The aim of this study is to discuss the non-parametric Whittaker graduation method as it is applied to remove statistical deviations from crude Turkish mortality rates. The whittaker method is popular since it has given good results in the past.*

*The Turkish mortality tables for 1990 and 2000 are based on crude mortality rates calculated from data from provincial and district centers of Turkey. The crude death rates are smoothed using the Whittaker method. A Microsoft Excel Program is used in order to utilize the Whittaker graduation method. Mortality tables are prepared for 1990 and 2000 year by graduated mortality rates by gender. Also, expectations of life are calculated using these mortality tables.*

*The results in the application reached are interpreted.*

**Key Words:** *Mortality rate, graduation, Whittaker*

## İNDEKS

MAKALE İSMİ*	YAZAR ADI	CİLT	NO	SAYFA	YIL
Üç Yönlü Tablolarda İstatistiğinin Kullanılması	AKYOL Mehmet, GÜRBÜZ Fikret	01	01	23	2002
Tiroit Bezi Verilerinin Bayes ve En Yakın K-Komşu Gibi Eğitici Yöntemlerle Sınıflanması	ALBAYRAK Songül	01	01	131	2002
Çok Değişkenli Kalite Kontrol Yaklaşımlarının Bir Değerlendirmesi	BAKIR M. Akif, KARACA Sezar	01	01	67	2002
Ortogonal Düzenler	BAYRAK Hülya, ALHAN Aslıhan	01	03	25	2002
Basılı Türkçe'nin Önemli Bazı İstatistiksel Özellikleri	DALKILIÇ Mehmet E., DALKILIÇ Gökhan	01	01	113	2002
Bernoulli Sayıları ve Lojistik Dağılım Fonksiyonu Üzerine	DÜNDAR Samim	01	01	1	2002
Ardışık Örneklem Planında Cusum Kontrol Kartlarının Kullanımı	HAMURKAROĞLU Canan, BACANLI Sevil	01	01	87	2002
Oransal ve Çarpımsal Tahmin Ediciler	KADILAR Cem, Hülya ÇINGİ	01	01	101	2002
Deney Tasarımında Kesirli Çoketkenli Tasarımlar ve Uygulamaları	KADILAR Cem, MULUK Zehra	01	01	11	2002
Farklı Ortalama Vektörü ve Farklı Kovaryans Matrisi Koşullarında Dört Değişkenli Lojistik Regresyon Modeli ve Diskriminant Analizine Ait Doğru Sınıflandırma Olasılıklarının Simülasyon Tekniği Yardımıyla Karşılaştırılması	KANIK E. Arzu, ÇAMDEVİREN Handan, GÜRBÜZ Fikret	01	03	13	2002
Hotelling T <sup>2</sup> Testinde Kovaryans Matrislerinin Heterojenliğinin I. Tip Hata Üzerine Etkisi: Bir Simülasyon Çalışması	Kanık E. Arzu, ÇAMDEVİREN Handan, GÜRBÜZ Fikret	01	03	1	2002
Varyansı Bilinmeyen Normal Dağılımlı Kitlelerin Ortalamasının Ardışık Tahmini	KOÇBERBER Güler, Hülya ÇINGİ	01	03	35	2002



MAKALE İSMİ*	YAZAR ADI	CİLT	NO	SAYFA	YIL
Matematiksel Finans	ÖNALAN Ömer	01	01	53	2002
Ankara İlinde Gelir Farklılıklarını Belirleyen Etmenler	SELİM Sibel	01	03	65	2002
Küçük Alan Tahminlerinde Sentetik Kestiriciler ve Özellikleri	SEVİNÇ Volkan	01	03	81	2002
Kardiyolojik Verilerin Doğrusal Olmayan Kanonik Korelasyon Analizi ile İncelenmesi	TÜRE Mevlüt, SÜT Necdet , KÜRÜM 01 Turhan , ÖZBAY Gültaç	01	03	55	2002
Bir Yatırım Problemine Analitik Hiyerarşi Prosesi Yönteminin Uygulanması	UYAR Yavuz, DİZDAR Ercüment N., KURT Mustafa	01	01	39	2002
Üçgensel Olumsuzluk Tablolarında Yarı-Bağımsızlık Modeli için Güç Analizi	AKTAŞ Serpil,SARAÇBAŞI Tülay	02	03	21	2003
İki ve Üç Yönlü Tabloların Gözelerinde 5'den Küçük Beklenen Frekans Olması Durumunda, I. Tip Hata Olasılığının Durumu	AKYOL Mehmet, GÜRBÜZ Fikret	02	01	47	2003
Eksik Blok Düzenlerinin Dual Yapıları	BAYRAK Hülya	02	01	1	2003
Koşullu Gauss Dağılımı ve Etkileşimleri	BAYRAK Hülya, GÖKPINAR Fikri	02	03	11	2003
Sürekli Değişken İçeren Grafikselleştirilmiş Modeller	BAYRAK Hülya, GÖKPINAR Fikri , ÖZKAYA Berrin	02	01	79	2003
Çok Değişkenli Normal Dağılıma Sahip Örneklerdeki Aykırı Gözlemlerin Belirlenmesi İçin Bayesgil Bir Yaklaşım	EKİZ Ufuk	02	01	11	2003
Genelleştirilmiş T (Gt) Dağılımına Dayalı Regresyon Analizi	GENÇ Ali İhsan, ARSLAN Olcay	02	03	1	2003

MAKALE İSMİ*	YAZAR ADI	CİLT	NO	SAYFA	YIL
Kanonik Korelasyon ve Açıklanmış Varyans Oranı İstatistiklerinin Başarı Düzeylerinin Tahminlenmesinde Kullanılması üzerine bir Araştırma	İŞÇİ Öznur	02	03	49	2003
Lineer Modellerde Yarı Uzay DerinliğineDayalı Dengeli Bootstrap Güven Bölgeleri	KARABULUT İhsan, ÖZTÜRK Fikri	02	03	63	2003
Bayes Ağlarda Koşullu Bağımsızlıkların İncelenmesi Üzerine Bir Çalışma	OLMUŞ Hülya, ERBAŞ Semra Oral	02	01	89	2003
Yapay Sinir Ağı Performansına Etki Eden Faktörlerin Analizinde Taguchi Yöntemi:Hisse Senedi Fiyat Tahmini Uygulaması	ÖZALP Alperen, ANAGÜN A. Sermet	02	01	29	2003
Yaşam Çözümlemesinde Yarışan Riskler ve Bir Uygulama	SERTKAYA Durdu, SÖZER M. Tekin	02	01	63	2003
Damar Darlık Derecesi ve Risk Faktörlerinin Homojenite Analizi ile İncelenmesi	SÜT Necdet, TÜRE Mevlüt	02	01	21	2003
Benzetimde Girdi Analizi Yapan Otomatik bir Sistem	TANIL Halil,BATMAZ İnci	02	03	31	2003
Türkiye'nin Sosyo-Ekonomik Yapısının Kanonik Korelasyon Analizi ile İncelenmesi	ARICIGİL ÇILAN Çiğdem	03	01	51	2004
Şehirleşme Seviyelerinin Projeksiyonu Üzerine bir Araştırma	İKİZ Fikret , KAYA Ahmet	03	01	61	2004
Bölünmüş Parsellerin Endüstriyel Deney Tasarımında Kullanımı ve Dört Düzeyli Etkenlerin bu Tasarımlara Yerleştirilmesi	MUTER İbrahim	03	01	21	2004
Ekstrem Değer Teorisi ile Riskin Değeri (Var)'ın Tahmini	ÖNALAN Ömer	03	01	35	2004

MAKALE İSMİ*	YAZAR ADI	CİLT	NO	SAYFA	YIL
Karmaşık Örnekleme Planlarında Çeşitli Varyans Tahmin Yöntemleri ve Uygulama	TERCAN Funda , ÇINGİ Hülya	03	01	85	2004
Sağdan Sansürlü Veriler İçin Parametrik Regresyon Modeli ve Kemik İliği Naklinde Kullanımı	TERZİ Yüksel , BEK Yüksel , CENGİZ Mehmet Ali	03	01	9	2004
Hipertansiyonun Tahmini için Çoklu Tahmin Modellerinin Karşılaştırılması	TÜRE Mevlüt ,KURT İmran , YAVUZ Ebru, KÜRÜM Turhan	03	01	73	2004
Rasgele Sıkıştırma Yoluyla Weibull Dağılımının Yeni Bir Karakterizasyonu	YURT ÖNCEL Sevgi , ALİEV ALIOĞLU Fazıl , AYGÜN Funda	03	01	1	2004
Üniversitelerarası Kurul Yabancı Dil Sınavı(ÜDS) Sonuçlarının Loglineer Analiz İle İncelenmesi	ALTAŞ Dilek , YILDIRIM Esen Zeren	03	03	43	2005
Türkiye'de 1980-1985 Dönemi Bölgelerin İl Merkezi, İlçe Merkezi İle Bucak Ve Köyler Bakımından Daimi İkametgaha Göre İç Göçler	ÇELEBİOĞLU Salih	03	03	67	2005
Kümelenmiş Verilerde Bağımlı İki Oranın Karşılaştırılması	GÖKMEN Derya , GENÇ Yasemin , ATAKURT Yıldız	03	03	21	2005
Çok Değişkenli Veri Kümeleri Üzerinde Tanımlı Aykırı Değer Belirleme Tekniklerinin Simülasyon Çalışması ile Karşılaştırılması	KIRAL Gülsen , BİLLOR Nedret	03	03	55	2005
Lojistik ve Cox Regresyon Modellerinin İncelenmesi ve Karşılaştırılması	SERTKAYA Durdu	03	03	31	2005
Whittaker Düzeltme Yöntemi ve Türkiye Ölüm Oranlarına Bir Uygulaması	TUZGÖL Hatice , GÜNAY Süleyman	03	03	85	2005
Andrew Fonksiyon Çizim Tekniği Yaklaşımı ile Çokdeğişkenli Kalite Kontrol Eğrileri ve Bir Uygulama	UYAR Yavuz	03	03	1	2005
Leslie Modeli Kullanılarak Türkiye'deki Kadın Nüfusu İle İlgili Bazı Parametrelerin Tahmini	ZIRHLIOĞLU Gürol	03	03	9	2005

\* İstatistik Araştırma Sempozyumu bildirilerinin yer aldığı Özel Sayılar dahil değildir.



## METİN HAZIRLAMA KALIBI

1. Araştırma, yazılar, kaynaklar, tablo ve şekiller ile birlikte en az 2 en çok 15 sayfa olmalıdır.
2. Gönderilecek araştırma PC ortamında Word 7.0 veya daha yukarı versiyonları ile Times New Roman font ortamında yazılmalıdır.
3. Araştırma A4 normundaki beyaz kağıda sol ve üstten 3,5 cm, sağ ve alttan 2,5 cm boşluk bırakılarak yazılmalıdır.
4. Araştırmanın türkçe ve ingilizce başlıkları metne uygun olmalıdır. Araştırmanın başlıkları büyük harflerle yazılmalı ve Özet büyük harflerle ortalı, 12 punto harf büyüklüğünde koyu olarak yazılmalıdır.
5. Yazarın adı ve soyadı, ünvan belirtilmeden başlığın iki satır altından ortalı olarak ad küçük, soyad büyük harfli olarak yazılmalıdır. İki veya daha fazla yazar olması durumunda, yan yana kolon (sütun) açılarak yazılmalıdır.
6. Yazarın adresi dip not şeklinde verilerek yıldız(\*) ile gösterilmelidir. Birden fazla yazar söz konusu olduğunda, yazışmaların hangi yazar ve adresle yapılacağını ise parantez içinde (haberleşme adresi) yazılarak verilmelidir. Dip not vermek gerektiğinde de yıldız(\*) kullanılmalıdır. Yazar(lar)ın adresi ve dip not ilgili sayfanın altına Times New Roman font ve 10 punto harf büyüklüğü kullanılarak yazılmalıdır.
7. Çalışma herhangi bir kurumun desteği ile gerçekleştirilmişse, kurumun adı ilk sayfa altında dip not olarak yazılmalıdır.
8. Araştırma bölümleri; Türkçe özet, Araştırma metni, Kaynaklar ve İngilizce özet (Abstract) şeklinde olmalıdır.
  - Türkçe özet, yazar isminden sonra üç satır boşluk bırakılarak yazılır. 200 kelimeyi geçmeyecek şekilde soldan 5,5 cm ve sağdan 4,5 cm boşluk bırakılarak 11 punto harf büyüklüğü kullanılarak, italik olarak yazılmalıdır.
  - Araştırma metni 12 punto harf büyüklüğü kullanılarak bir satır aralığında ve paragraflar arasında bir satır boşluk bırakılmalıdır. Paragraflar ve formüller bir tab içeriden yazılmalıdır. Birinci derece bölüm başlıkları büyük harfle, ikinci derece alt bölüm başlıklarında her sözcüğün ilk harfi büyük, diğerleri küçük harfle, üçüncü ve daha alt derece alt bölüm başlıklarının yalnız ilk harfi büyük, diğerleri küçük harfle yazılmalıdır. Bütün bölüm başlıkları koyu olarak yazılmalıdır, tablo ve şekillere başlık ve sıra numarası bölüm numarası içermeksizin verilir. Tablo ve şekil başlık ve sıra numaraları yarım satır aralıklı tablolarda üstte, şekillerde altta yer almalıdır.
  - Kaynaklara göndermeler metin içinde açılan ayrıçlarla yapılmalıdır. Ayrıç içindeki sıra şöyledir: Yazar(lar)'ın soyadı ve kaynağın yılı. Örneğin; ...kanıtlanmıştır (Rao, 1974)., ...(Grossman ve Weiss, 1983)., ...(Baumal, 1952; Tobin, 1956)., ... (Winebrake vd, 1995)., ...Rao (1974) kanıtlamıştır. vb. şeklinde gösterilmelidir.

Çalışmada gönderme yapılan bütün kaynaklar, kaynaklar listesinde belirtmeli; çalışmada yararlanılmayan kaynaklar, kaynaklar listesinde yer almamalıdır. Kaynaklar araştırma metninin sonunda yazarının soyadına göre alfabetik sırada ve 11 puntoda

kaynaklar arasında bir satır boşluk bırakılarak yazılmalıdır. Bunların yazım şekli aşağıda gösterildiği gibi standart formda olmalıdır:

**Örnekler:**

Kitap

BRUBAKER, S. (1967), *Trends in the World Aluminium Industry*, Baltimore, Maryland: John Hopkins Press.

Araştırma

RAO, J.N.K. (1994), *Estimating Totals and Distribution Function Using Auxiliary Information at the Estimation Stage*, Journal of Official Statistic, 10, 153 – 165.

Derleme

ARTHUR, W.B. (1988), *Competing Technologies: An Overview*, G.Dosi, C. Freeman, R. Nelson, G. Silverberg ve L. Soete (der.), *Technical Change and Economic Theory* içinde Londra:Pinter, 590-607.

İnternet

SUTCLIFFE, M.J., Wo, Z.G. and OSWALD, R.E. (1996). *Three-dimensional models of non-NMDAglutamaterceptors*, Erişim: [[http://neon.chem.le.ac.uk/cornell/Sutcliffe\\_BJ/Sutcliffe\\_BJ.html](http://neon.chem.le.ac.uk/cornell/Sutcliffe_BJ/Sutcliffe_BJ.html)].Erişim Tarihi: 22.12.1996

- Araştırmanın İngilizce dilde özeti araştırmanın sonunda verilmelidir. Araştırmanın İngilizce adı üstten 2 satır boşluk bırakılarak ortalı, her sözcüğün ilk harfi büyük, 14 punto harf büyüklüğünde, Abstract büyük harflerle ortalı, 12 punto harf büyüklüğünde koyu olarak yazılmalıdır. İngilizce özet soldan 5,5 cm ve sağdan 4,5 cm boşluk bırakılarak 200 kelimeyi geçmeyecek şekilde 11 punto harf büyüklüğünde italik olarak araştırmanın İngilizce adından sonra 3 satır boşluk bırakılarak yazılmalıdır.
  - Anahtar kelimeler (Key words) her iki özeti bir satır altına, anahtar kelimeler ve key words koyu italik olarak yazılmalıdır.
9. Matematik simge ve formüllerin yazımında aşağıdaki hususlara dikkat edilir:
- Simgelerin ayırt edilmesi önemlidir. Özellikle büyük ve küçük harfler, düz ve koyu harfler, Klasik Yunan ve Latin harfleri, alt ve üst indisler, sıfır (0) rakamı ve O harfi, Bir (1) rakamı ve l (l) harfi ayırt edilebilmelidir. Çoklu indislerden sakınılmalıdır.
  - Denklemler word, standart (default) ölçülerde 1 tab (1,27 cm) içerden ve numara vermek gerekliyse bölüm numarasını içermeksizin en sağına parantez içinde yazılmalıdır. Uzun formüller metin içinde yer almamalıdır.
  - Kesirler, metin içinde (/) işareti ile gösterilmelidir.
  - Karmaşık ifadeler içeren denklemler olabildiğince kısaltma simgeleri kullanılarak yazılmalıdır.
  - İç içe çoklu ayrıçlar aynı formülde yer aldığında, sıra düzeni örneğin  $\{[( 0 )]\}$  biçiminde olmalıdır.
10. Araştırmanın Türkçe yazım kurallarına uygun olması yazarın sorumluluğunda altındadır.

## DUYURU PANOSU \*

### 5. SINIFLANDIRMA VE VERİ ANALİZİ KONFERANSI (CLADAG)

06 – 08 Haziran 2005, Parma / İtalya

Ayrıntılı bilgi için: Andrea Cerioli and Isabella Morlini, Dept. Of Economics,  
University of Parma, Kennedy 6, 43100 Parma / İtalya

e-mail: [info@cladag2005.unipr.it](mailto:info@cladag2005.unipr.it)

Website : <http://www.cladag2005.unipr.it>

### “EKONOMİ VE SOSYAL AÇIDAN İŞGÜCÜ PİYASASINA İLİŞKİN GÖÇ İSTATİSTİKLERİ” CEIES SEMİNERİ

09 – 10 Haziran 2005, Riga / Latvia

Ayrıntılı bilgi için bakınız: <http://forum.europa.eu.int/Public/irc/dsis/ceies/library>

### 5. ULUSLARARASI ROBUST İSTATİSTİK KONFERANSI

12 – 17 Haziran 2005, Jyväskylä / Finlandiya

Ayrıntılı bilgi için:

e-mail: [icors@maths.jyu.fi](mailto:icors@maths.jyu.fi)

Website : [www.stat.jyu.fi/icors2005](http://www.stat.jyu.fi/icors2005)

### 27. ULUSLARARASI BİLGİ TEKNOLOJİSİ ARAYÜZLERİ KONFERANSI (ITI 2005)

20 – 23 Haziran 2005, Cavtat-Dubrovnik / Hırvatistan

Ayrıntılı bilgi için: SRCE – University Computing Centre, University of Zagreb, J.  
Marohnica bb, HR-10000 Zagreb, Hırvatistan

e-mail: [iti@srce.hr](mailto:iti@srce.hr)

Website: <http://iti.srce.hr/>

### PROF. KAI TAI FANG'İN 65. YAŞI ONURUNA ULUSLARARASI İSTATİSTİK KONFERANSI

20-24 Haziran 2005, Hong Kong

Ayrıntılı bilgi için: e-mail: [fang65@math.hkbu.edu.hk](mailto:fang65@math.hkbu.edu.hk)

Website : <http://www.math.hkbu.edu.hk/Fang65>

### 30. STOKASTİK SÜREÇLER VE UYGULAMALARI KONFERANSI

26 Haziran - 1 Temmuz 2005, California / ABD

Ayrıntılı bilgi için: Raya Feldman

e-mail: [feldman@pstat.ucsb.edu](mailto:feldman@pstat.ucsb.edu)

Website : <http://www.pstat.ucsb.edu/projects/spa05>



#### 4. ULUSLARARASI ÇOKLU KARŞILAŞTIRMA SÜREÇLERİ KONFERANSI – MCP 2005

17 - 19 Ağustos 2005, Shanghai / Çin  
Ayrıntılı bilgi için : Jason Hsu  
e-mail: [Hsu.1@osu.edu](mailto:Hsu.1@osu.edu)  
Website : <http://www.stat.ohio-state.edu/~mcp2005>

#### 14. AVRUPA GENÇ İSTATİSTİKÇİLER TOPLANTISI

22 - 27 Ağustos 2005, Debrecen / Macaristan  
Ayrıntılı bilgi için : Katalin Varga  
e-mail: [varga@szit.bme](mailto:varga@szit.bme), [vkati@hu.inter.net](mailto:vkati@hu.inter.net)

#### ULUSLARARASI UYGULAMALI İSTATİSTİK 2005 KONFERANSI

18 - 21 Eylül 2005, Ribno / Slovenya  
Ayrıntılı bilgi için : Andrej Blejec  
e-mail: [info.AS@nib.si](mailto:info.AS@nib.si)  
Website : <http://ablejec.nib.si/AS2005>

#### STOKASTİK GEOMETRİ VE UYGULAMALARI

03 - 07 Ekim 2005, Bern / İsviçre  
Ayrıntılı bilgi için : Prof. I. Molchanov  
e-mail: [ilya@stat.unibe.ch](mailto:ilya@stat.unibe.ch)  
Website : <http://www.cx.unibe.ch/~ilya/wbec>

#### 12. ULUSLARARASI İSTATİSTİK, KOMBİNATORİK, MATEMATİK VE UYGULAMALAR KONFERANSI

02 - 04 Aralık 2005, Alabama / ABD  
Ayrıntılı bilgi için : Mark Carpenter [carpediem@auburn.edu](mailto:carpediem@auburn.edu)  
Satya Mishra [mishra@jaguar1.usouthal.edu](mailto:mishra@jaguar1.usouthal.edu)  
Website : <http://www.stat.auburn.edu/samc2005>

#### BİLİM VE MÜHENDİSLİK TEKNOLOJİSİNDE ARAŞTIRMA METODOLOJİLERİ ULUSLARARASI KONFERANSI - ICRMSET

05 - 07 Aralık 2005, Tehran / İran  
Ayrıntılı bilgi için : Dr. Alireza Bayat  
e-mail : [info@icmrse.ir](mailto:info@icmrse.ir)  
Website : [www.icmrse.ir](http://www.icmrse.ir)

## **SRTL-4: 4. ULUSLARARASI İSTATİSTİKİ TÜMEVARIM, DÜŞÜNME VE OKURYAZARLIK ARAŞTIRMA FORUMU**

02 – 07 Temmuz 2005, Yeni Zelanda  
Ayrıntılı bilgi için: Maxine Pfannkuch  
e-mail: [m.pfannkuch@auckland.ac.nz](mailto:m.pfannkuch@auckland.ac.nz)  
Website : <http://www.stat.auckland.ac.nz/srtl4/>

## **13. INFORMS - UYGULAMALI OLASILIK KONFERANSI**

06 - 08 Temmuz 2005, Ottawa / Kanada  
Ayrıntılı bilgi için: Co-chair Doug Down  
e-mail: [down@univmail.cis.mcmaster.ca](mailto:down@univmail.cis.mcmaster.ca)  
Website: <http://www.appliedprob.society.informs.org/ottawaconf.html>

## **ULUSLARARASI PARAMETRİK OLMAYAN ÇIKARSAMA SEMİNERİ**

13 - 15 Temmuz 2005, Coruna / İspanya  
Ayrıntılı bilgi için: Ricardo Cao, Universidade de Coruna  
Website : <http://www.orzancongres.com/eventos/2005/isni/>

## **IUSSP ULUSLARARASI NÜFUS KONFERANSI**

18 - 23 Temmuz 2005, Tours / Fransa  
Ayrıntılı bilgi için: e-mail: [France2005-Booth@iussp.org](mailto:France2005-Booth@iussp.org),  
[France2005-SideMeeting@iussp.org](mailto:France2005-SideMeeting@iussp.org)  
Website : <http://www.iussp.org/France2005/>

## **ISIPTA '05: 4. ULUSLARARASI ETKİLEYİCİ OLASILIKLAR VE UYGULAMALARI SEMPOZYUMU**

20 - 24 Temmuz 2005, Pennsylvania / ABD  
Ayrıntılı bilgi için bakınız: <http://www.sipta.org/isipta05>

## **25. AVRUPA İSTATİSTİKÇİLER TOPLANTISI (EMS)**

24 - 28 Temmuz 2005, Oslo / Norveç.  
Ayrıntılı bilgi için : EMS 2005,P.O. Box 114 Blindern, N-0314 Oslo, Norveç  
e-mail: [frigessi@nr.no](mailto:frigessi@nr.no), [ems2005@nr.no](mailto:ems2005@nr.no)  
Website : [www.ems2005.no](http://www.ems2005.no)

**ULUSLARARASI BİLGİSAYAR VE ENFORMASYON BİLİMLERİ  
KONFERANSI – ICCIS 2005**

15 - 17 Aralık 2005, Kolhapur / Hindistan  
Ayrıntılı bilgi için : Prof. Rajendra Akerkar  
e-mail : [kpr\\_tmrf@sancharnet.in](mailto:kpr_tmrf@sancharnet.in)  
Website : [http://pune.sancharnet.in/kpr\\_tmrf/iccis05.html](http://pune.sancharnet.in/kpr_tmrf/iccis05.html)

**ULUSLARARASI İSTATİSTİK KONFERANSI  
“TEKNOLOJİ ÇAĞINDA İSTATİSTİK”**

27 - 31 Aralık 2005, Kuala Lumpur / Malezya  
Ayrıntılı bilgi için : Uluslararası İstatistik Konferansı Sekreterliği  
e-mail : [isc\\_malaysia@um.edu.my](mailto:isc_malaysia@um.edu.my)  
Website : <http://iscm.math.um.edu.my>

**5. ULUSLARARASI İŞ VE ENDÜSTRİYEL İSTATİSTİK SEMPOZYUMU**

11 - 13 Ocak 2006, Lima / Peru  
Ayrıntılı bilgi için : Geoff Vining  
e-mail: [vining@vt.edu](mailto:vining@vt.edu)  
Website : <http://kitchen.stat.vt.edu/~vining/isbis5/>

**5. ULUSLARARASI OLASILIK VE İSTATİSTİK KONFERANSI  
PROBASTAT 2006**

05 - 09 Haziran 2006, Slovak Cumhuriyeti  
Ayrıntılı bilgi için : Dr. Viktor Witkovsky, Institute of Measurement Science,  
Slovak Academy of Sciences, Dubravska cesta 9, 84104 Bratislava, Slovakiya  
e-mail: [probastat@savba.sk](mailto:probastat@savba.sk)  
Website : <http://aiolos.um.savba.sk/~viktor/probastat.html>

**6. ULUSLARARASI S4G KONFERANSI**

26 - 29 Haziran 2006, Prag / Çek Cumhuriyeti  
Ayrıntılı bilgi için : Viktor Benes ([benesv@karlin.mff.cuni.cz](mailto:benesv@karlin.mff.cuni.cz))  
Radka Juzkova ([radka.juzkova@svses.cz](mailto:radka.juzkova@svses.cz))  
Website : <http://www.karlin.mff.cuni.cz/s4g/>

**26. AVRUPA İSTATİSTİKÇİLER TOPLANTISI**

24 - 28 Temmuz 2005, Thorun / Polonya  
Ayrıntılı bilgi için : Adam Jakubowski (Yerel Organizasyon Komitesi Başkanı)  
Website : [www-m4.mathematik.tu-muenchen.de/m4/erc/](http://www-m4.mathematik.tu-muenchen.de/m4/erc/)  
<http://www.ems2006.umk.pl>



## COMPSTAT 2006

28 Ağustos - 01 Eylül 2006, Roma / İtalya

Ayrıntılı bilgi için : e-mail: [compstat2006@uniroma1.it](mailto:compstat2006@uniroma1.it)

Website : <http://w3.uniroma1.it/compstat2006>

## ISI 56. DÖNEM TOPLANTISI

22 - 29 Ağustos 2007, Lizbon / Portekiz

Ayrıntılı bilgi için : ISI Permanent Office, 428 Prinses Beatrixlaan, P.O. Box 950,  
2270 AZ Voorburg, Holanda.

e-mail: [isi@cbs.nl](mailto:isi@cbs.nl)

\*Kaynak: International Statistical Institute (ISI) Web sitesi

<http://www.cbs.nl/isi/calendar.htm#TOP>

**Andrew Fonksiyon Çizim Tekniği Yaklaşımı ile Çokdeğişkenli Kalite Kontrol Eğrileri ve bir Uygulama**  
*Construct Multivariate Quality Control Curve by the Approach of Andrew's Function Plot Technique and an Application*

Yavuz UYAR..... 1

**Leslie Modeli Kullanılarak Türkiye'deki Kadın Nüfusu ile İlgili Bazı Parametrelerin Tahmini**  
*Estimation of Some Parameters for Woman Population in Turkey: A Leslie Model Approach*

Gürol ZIRHLIOĞLU..... 9

**Kümelenmiş Verilerde Bağımlı İki Oranın Karşılaştırılması**  
*Comparison of Two Correlated Proportions for Clustered Data*

Derya GÖKMEN  
Yasemin GENÇ  
Yıldır ATAKURT  
Banu YAĞMURLU..... 21

**Lojistik ve Cox Regresyon Modellerinin İncelenmesi ve Karşılaştırılması**  
*Review of Logistic and Cox Regression Models*

Durdu SERTKAYA..... 31

**Üniversitelerarası Kurul Yabancı Dil Sınavı(ÜDS) Sonuçlarının Loglineer Analiz ile İncelenmesi**  
*Evaluation of the Results of the UDS by Loglinear Models*

Dilek ALTAŞ..... 43

**Çok Değişkenli Veri Kümeleri Üzerinde Tanımlı Aykırı Değer Belirleme Tekniklerinin Simülasyon Çalışması ile Karşılaştırılması**

*A Comparison of the Multiple Outlier Detection Method for Multivariate Data by Simulation Study*

Gülsen KIRAL  
Nedret BİLLOR..... 55

**Türkiye'de 1980-1985 Dönemi Bölgelerin İl Merkezi, İlçe Merkezi ile Bucak ve Köyler Bakımından Daimi İkametgaha Göre İç Göçler**

*The Internal Migration by the Permanent Residence Among the Regions in Turkey in View of Centers of Province, District and Subdistrict and Villages, in the Period 1980-1985*

Salih ÇELEBİOĞLU..... 67

**Whittaker Düzeltme Yöntemi ve Türkiye Ölüm Oranlarına Bir Uygulaması**  
*Whittaker Graduation Method and an Application to Turkish Mortality Rates*

Hatice TUZGÖL  
Süleyman GÜNAY..... 85