



İstatistik Araştırma

Journal of Statistical Research Dergisi

Cilt 02 Volume

No 03 Number

Aralık 2003 December



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 İNAL, 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

Olca ARSLAN, Çukurova Üniv., Adana

Refik SOYER, George Washington University, USA

Soner GÖNEN, Gazi Üniv., Ankara

Zehra MULLUK, 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 BICYAP

Journal of Statistics Research Secretary Address

State Institute of Statistics

Department of

Research-Planing and Coordinating

Request and Subscription Address

State Institute of Statistics

Revolving Fund Management

Necatibey Street No: 114

06100 Yücetepe / ANKARA

Tel: +90 312 417 64 40 /702 - 732

Faks:+90 312 425 35 85

e-mail: dergi@die.gov.tr

URL: http://www.die.gov.tr

Tel: +90 312 417 64 40 / 323 - 319

Fax: +90 312 417 58 86

EDİTÖRDEN

Değerli Okuyucular ve Meslekdaşlarım,

Dergimizin bu sayısını da baskıdan çıkartarak ellerinize ulaşmasını, yayın kurulunda yer alan meslekdaşlarımızın ve DİE 'nin ilgili yönetici ve birimlerinin de katkılarıyla, sağlamış olmanın kıvancını ve mutluluğunu yaşıyoruz. Dergimizin bundan sonraki sayısından itibaren dergimizde yayınlanmış makalelerle ilgili olarak yazar veya yazarların kişiliklerine zarar vermeyeceğine kanaat getirdiğimiz okuyuculardan elimize ulaştırılan eleştiri yazılarını ve bunlara ilgili yazar veya yazarların cevaplarını içeren yazıların bulunduğu bir bölümümüzün olacağını sizlere duyurmak istiyoruz. Öte yandan istatistik ve ilgili alanlarda piyasaya sunulmuş bilim insanlarımızın-kendilerinin talep etmeleri halinde- kitaplarıyla ilgili belirleyeceğimiz uzmanlarca yapılmış eleştirilerinin yer alacağı bir bölümü, ve istatistik ve yakından ilgili alanlarda elimize ulaşan ileriki bir tarihte gerçekleştirilecek yurtiçi ve yurtdışı bilimsel toplantılarla ilgili temel bilgilerin yer alacağı bir diğer bölümü de yine bundan sonraki sayılarımızda görebilmeniz mümkün olabilecektir. Yukarıda duyurduğum ve kararlaştırdığımız yeniliklere ilave olarak gerek üniversitelerimize ziyaretçi öğretim üyesi olarak gelmiş kişilerle ilgili ve gerekse de Devlet İstatistik Enstitüsü, Devlet Planlama Teşkilatı, Hazine ve Dış Ticaret Müsteşarlıkları ve benzeri kamu kuruluşlarına yurtdışından gelmiş veya gelmesi beklenen ziyaretçi bilim insanları ile ilgili haberlere de dergimizde yer vererek sizlerle paylaşmak istiyoruz. Bundan böyle büyük özverileri ile dergimizin çıkmasında emeği geçen hakemlerimizin tamamının listesini o yıla ait son sayıda verebileceğiz. Ayrıca bu sayımızla birlikte, sizlerin de dikkatini çektiğini sandığım bir değişiklik de başlangıçtan buyana Dergimizin çeşitli sayılarında yayınlanmış makalelerin başlıklarına ve yazar veya yazarlardan birincisinin soyadına göre iki dizini (indeksi) sunuyoruz. Daha sonraki sayılarımızda çıkan makalelerle ilgili olanları ise ait olduğu cildin son sayısında yayınlayacağız.

Dergimizin bu sayısı dahil-Sempozyumlarla ilgili özel sayılar için olanlar hariç-değerlendirilmek üzere gönderilen toplam makale sayısı 62 olup bunlardan 26 sının basılması uygun görülüş 18 makale uygun görülmemiş veya reddedilmiş 18 tanesi ise halen hakem değerlendirmesi sürecinde bulunmaktadır. Aralık 2003 sayımızda yayınlanmak üzere 6 adet makale baskıya verilmiştir. Bu sayımızda da olduğu gibi Dergimizin bazı sayılarında yayınlanan makale sayısının düşük olmasının nedenlerinden biri bazı hakemlerimizde ilgili makalenin uzun süre elde bekletilmesi nedeniyle değerlendirme sürecinin uzamasından kaynaklanmaktadır. Derginin genel editörü olarak, dergimizde yayınlanmak amacıyla gönderilmiş makalelerin değerlendirilmelerini yapmaları için seçilmiş hakemlerimizin bu görevlerini büyük bir özveriyle yaptıklarından eminim. Derginin düzenli ve gecikmeden çıkabilmesi için, sizlerin adına, özellikle, kendilerince doldurulmuş değerlendirme raporlarını süresi içinde göndermelerini bilhassa rica ediyorum. Editörler Kurulu üyelerini ve hakemlerimizi en çok uğraştıran olumsuzluklardan birisi yazar veya yazarlarımızın Dergimizin yazım kurallarına yeterince özen göstermemeleridir. Bununla ilgili olarak web-sayfamızda yakın bir gelecekte bazı uyarılara dikkat çekilecektir.

Derginin bu sayısının ıkmasında emeđi geen bařta Dergimizde yayınlanması amacıyla makalelerini gnderen (yayınlanmasa bile) tm arařtırmacılara, bu sayımızda katkılarıyla hibir karřılık beklemezsiniz bizlere yardımcı olan tm hakemlerimize, Dergi Sekreteryasına ve Devlet İstatistik Enstitsnn deđerli alıřanlarına hepimizin adına teřekkr etmeyi bir bor biliyorum. İeriđi ve kalitesi daha zengin Dergimizin yeni sayılarıyla sizlere ulařmak dileđiyle saygılarımı sunuyorum.

Prof. Dr. Fetih YILDIRIM
Genel Editr

BU SAYIDA EMEĐİ GEÇEN HAKEMLER

Prof.Dr. Aşen APAYDIN

Prof.Dr. Hamza GAMGAM

Prof.Dr. Zehra MULUK

Prof.Dr. Semra Oral ERBAŞ

Prof.Dr. Soner GÖNEN

Prof.Dr. Hüseyin TATLIDİL

Prof.Dr. Fikri AKDENİZ

Prof.Dr.Serdar KURT

Prof.Dr.İhsan SABUNCUOĐLU

Prof.Dr.Öniz TOKTAMIŞ

Doç.Dr. Sadullah SAKALLIOĐLU

Doç.Dr. Fulya ALTIPARMAK

Doç.Dr. Hülya BAYRAK

Yrd.Doç.Dr. Aydın ERAR

Yrd. Doç.Dr. İlknur ÖZMEN

Genelleştirilmiş T (Gt) Dağılımına Dayalı Regresyon Analizi

Ali İhsan GENÇ

Olca ARSLAN*

ÖZET

Bu çalışmada $y = X\beta + u$ çoklu lineer regresyon modelindeki u hata teriminin 0 ortalamalı ve σ ölçek parametrelili genelleştirilmiş t (GT) dağılımından geldiği kabul edilmiştir. GT dağılımının şekil parametrelerinin bilindiği varsayımı altında regresyon modelinin parametreleri ve σ ölçek parametresi tahmin edilmiştir. Önerilen kestirim yöntemi bir takım problemleri veri kümelerine uygulanmış ve alınan sonuçlar diğer dayanıklı (robust) kestirimlerle karşılaştırılmıştır.

Anahtar Kelimeler: Maksimum Olabilirlik; Sapan Değer; Regresyon; Dayanıklı Kestirim; Etki Fonksiyonu.

1. GİRİŞ

Regresyon analizinin hem teorisinde hem de uygulamasında modeldeki hataların çoğunlukla normal (Gauss), bağımsız ve ortak bir varyansla dağıldığı varsayılır. Fakat bazı veri kümeleri için hataların dağılımı normalden daha kalın kuyruklu olabilir ve bu durumda normal dağılıma dayalı yapılan analizler doğru sonuç vermeyebilir. Literatürde kuyruk çeşitliliği bakımından zengin pek çok dağılım regresyonda hataların modellenmesinde normal dağılıma alternatif olarak kullanılmıştır. Örneğin, Laplace dağılımı (Bloomfield ve Steiger, 1983), t dağılımı (Zellner, 1976, Lange vd., 1989, Arslan, 1992), BT (Box-Tiao) dağılımı (Klein ve Spady, 1984), genelleştirilmiş üstel dağılımlar (Lye ve Martin, 1993). Çalışmada hataların modellenmesi için önerilen GT dağılımı ise başta t dağılımı ve BT dağılımı olmak üzere bir çok dağılım parametrelerinin özel durumlarında veya limit durumlarında içerdiğinden hataların modellenmesinde daha genel bir dağılım sınıfı oluşturmaktadır.

2. GENELLEŞTİRİLMİŞ T (GT) DAĞILIMI

McDonald ve Newey (1988) GT dağılımını normal ve t dağılımının bir alternatifi olarak regresyonda hataları modelleyip bir kısmen adapte kestirim (partially adaptive estimation) yöntemi geliştirmek için tanımlanmıştır. Kullanılan yöntem en küçük kareler (EKK), en küçük mutlak sapma (LAD-Least Absolute Deviation), L_p gibi kestiricileri özel durumları olarak içermektedir. Bu yöntem bir çok ekonomik modeli kestirmek için kullanılmıştır. Örneğin, market modelinde (McDonald ve Nelson, 1989, Butler vd., 1990), ARMA zaman serileri modellerinde (McDonald, 1989).

* Çukurova Üniversitesi Fen-Edebiyat Fakültesi Matematik Bölümü 01330 Balcalı-Adana

İstatistiksel ekonomideki kullanımına rağmen *GT* dağılımının parametrik özellikleri literatürde yeterince ele alınmamıştır. Arslan ve Genç (2002) tek değişkenli bir veri kümesini *GT* dağılımıyla modelleyip konum ve ölçek parametrelerinin maksimum olabilirlik kestiricilerini buldular. Ayrıca konum ve ölçek kestiricilerinin parametre uzayında tek olarak mevcut olabilmesi için $p \geq 2$, $pq \geq 1$ yeter koşulunu elde ettiler.

GT dağılımı aşağıda (1) denklemiyle verilen simetrik tek modlu bir yoğunluğa sahiptir:

$$f(x; \mu, \sigma, p, q) = \frac{pq^q}{2\sigma B(1/p, q)\sigma} \left(q + \frac{|x - \mu|^p}{\sigma^p}\right)^{-q-1/p}. \quad (1)$$

Yoğunluk fonksiyonunda $B(\cdot)$ beta fonksiyonu, $\sigma \in (0, \infty)$ ölçek parametresi, $-\infty < \mu < \infty$ konum parametresi ve $p > 0$ ve $q > 0$ da şekil parametreleridir. p ve q nun büyük değerleri yoğunluk fonksiyonunun kuyruklarını inceltirken küçük değerleri daha kalın kuyruklu yoğunluklara neden olur. Ayrıca bu parametrelerin çeşitli değerlerinde ve limit durumlarında çeşitli dağılımları elde ederiz. Örneğin, $p=2$ için t dağılımı, $q \rightarrow \infty$ için BT dağılımını ve $p \rightarrow \infty$ için de $(-\sigma, \sigma)$ üzerinde düzgün dağılımı elde ederiz (McDonald ve Newey, 1988).

3. PARAMETRELERİN KESTİRİMİ

y bağımlı değişken üzerindeki bir $n \times 1$ vektör, X , k regresör değişkeninin değerlerinin $n \times p$ matrisi olmak üzere

$$y = X\beta + u, \quad (2)$$

çoklu lineer regresyonu ele alalım. Burada u gözlenemez rasgele hataların $n \times 1$ vektörü, β ise modeldeki bilinmeyen parametrelerin $p \times 1$ vektörüdür.

(2) denklemindeki u_i hatalarının bağımsız ve özdeş olarak 0 ortalamalı, bilinmeyen σ ölçek parametrelili *GT* dağılımına sahip olduğunu varsayalım. Şekil parametreleri p ve q nun bilindiğini kabul edelim. O zaman x_i , X matrisinin i . satırı olmak üzere y_i nin yoğunluk fonksiyonu

$$f(y_i) = \frac{pq^q}{2\sigma B(1/p, q)} \left(q + \frac{|y_i - x_i\beta|^p}{\sigma^p}\right)^{-q-1/p}, \quad -\infty < y_i < \infty, \quad i=1,2,\dots,n \quad (3)$$

ve log-olabilirlik fonksiyonu

$$l(\beta, \sigma) = -n \log \sigma - (q + \frac{1}{p}) \sum_{i=1}^n \log \left(q + \frac{|y_i - x_i\beta|^p}{\sigma^p}\right), \quad (4)$$

dir. (4) denkleminin β ve σ ya göre kısmi türevlerinin alınıp 0 a eşitlenmesiyle aşağıda verilen olabilirlik kestirim denklemleri elde edilir.

$$\frac{\partial l}{\partial \beta} = (q + \frac{1}{p}) \sum_{i=1}^n \frac{\sigma^{-p} p x_i |y_i - x_i \beta|^{p-1} \text{sign}(y_i - x_i \beta)}{q + \frac{|y_i - x_i \beta|^p}{\sigma^p}} = 0, p > 1 \quad (5)$$

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + (q + \frac{1}{p}) \sum_{i=1}^n \frac{p \sigma^{-p-1} |y_i - x_i \beta|^p}{q + \frac{|y_i - x_i \beta|^p}{\sigma^p}} = 0. \quad (6)$$

(5) ve (6) da verilen kestirim denklemlerinin yeniden düzenlenmesiyle aşağıdaki denklemleri elde ederiz.

$$\sum_{i=1}^n w_i x_i (x_i \hat{\beta} - y_i) = 0, \quad (7)$$

$$\sum_{i=1}^n w_i (y_i - x_i \hat{\beta})^2 - n \hat{\sigma}^2 = 0, \quad (8)$$

ağırlıklar $w_i = \frac{(pq+1) |y_i - x_i \hat{\beta}|^{p-2} \hat{\sigma}^{2-p}}{q + |y_i - x_i \hat{\beta}|^p \hat{\sigma}^p}$ dir. Bu denklemler sıfıra azalan M-

kestirim denklemleridir. (7) ve (8) denklemlerinden regresyon parametrelerinin kestirimi aşağıdaki gibi yinelemeli tekrar ağırlıklandırılan en küçük kareler kestirimi şeklinde yazılabilir.

$$\hat{\beta} = (\sum_{i=1}^n w_i x_i x_i)^{-1} (\sum_{i=1}^n w_i y_i x_i), \quad (9)$$

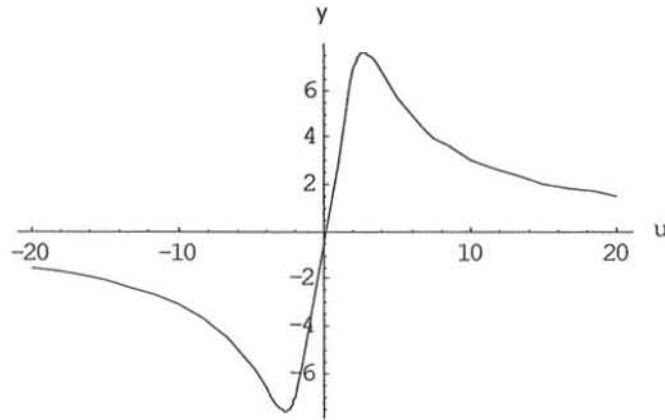
$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n w_i (y_i - x_i \hat{\beta})^2. \quad (10)$$

Hatalar GT dağılımlı olduğunda kestiricilerin hesaplanmasında kullanılacak yinelemeli yöntemin ağırlıklı en küçük kareler yönteminden farkı ağırlıkların sabit kalmayıp her iterasyonda değişecek olmasıdır. Üstelik w ağırlık fonksiyonu rezidülerin azalan bir fonksiyonu olduğundan büyük rezidülere küçük ağırlıklar verilecektir. Böylece veri kümemizdeki sapan değerlere küçük ağırlıklar verileceğinden elde edeceğimiz kestirimler sapan değerlerden çok az etkilenecektir.

$u_i = y_i - x_i \beta$ hatalarının ψ -fonksiyonu $p > 1$ için

$$\psi(u) = \frac{(pq+1) |u|^{p-1} \text{sign}(u)}{q \sigma^p + |u|^p} \quad (11)$$

olarak elde edilir (McDonald ve Newey, 1988) (Şekil 1). ψ -fonksiyonunun rezidülerin ağırlık fonksiyonuyla arasındaki ilişkisi $\psi(u) = w(u)u\sigma^{p-2}$ bağıntısıyla verilir. $u > 0$ için ψ -fonksiyonu $(0, [(p-1)q\sigma^p]^{1/p})$ aralığında artan, $([(p-1)q\sigma^p]^{1/p}, \infty)$ aralığında ise azalmandır. Üstelik $\lim_{|u| \rightarrow \infty} \psi(u) = 0$ olduğundan regresyon katsayılarının etki fonksiyonu sifıra azalan (redescending) tipindedir. ψ -fonksiyonu p ve q şekil parametrelerine bağlı olduğundan küçük ağırlıklar tahsis edilecek olan sapan değer sayısı p ve q nun değerlerine göre azalacak veya artacaktır. Bu yüzden p ve q şekil parametrelerine ayar sabitleri (tuning constants) olarak bakılabilir.



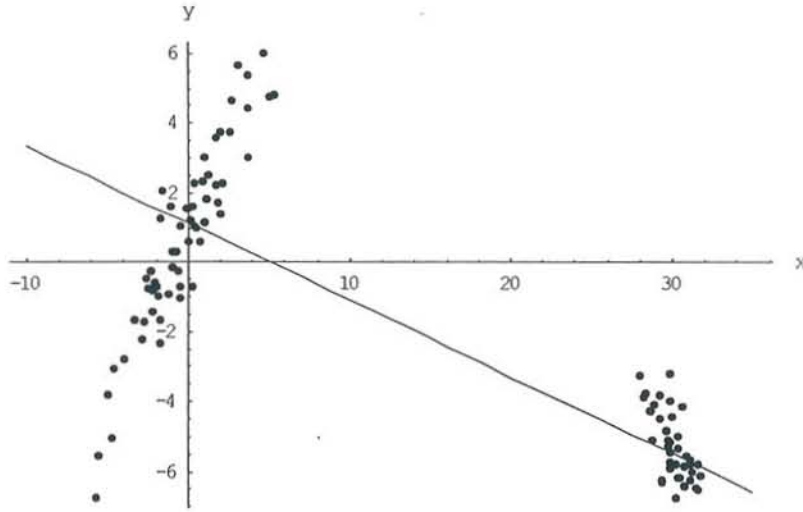
Şekil 1. ψ -fonksiyonu ($p=3, q=10$)

4. ÖRNEKLER

Örnek 1. Bu örnekte Arslan (2002) tarafından üretilen suni veriyi analiz edeceğiz. Veri kümesi içerisinde iki altgrubu barındıran 100 noktadan oluşmaktadır. Serpme diyagramından da görüldüğü gibi 40 noktalık sapan değerlerden oluşan bir alt grup x-uzayı yönünde bulunmaktadır (Şekil 2). Parametre kestirimleri aşağıdaki Tablo 1'de verilmiştir. İterasyona başlangıç değeri olarak en küçük kareler (EKK) kestirimleri alınmıştır. Tablo 1'den çeşitli yöntemler için benzer kestirimlerin elde edildiğini ve regresyon doğrusunun veri kümesindeki sapan değerlerden oluşan alt grubun arasından geçtiğini görürüz. (Şekil 2).

Tablo 1. Örnek 1' e ait regresyon kestirimleri (LMS: Least Median of Squares)

Yöntem	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}$	Log-lik
EKK	.34	-.17	2.50	-233.38
$GT(p=1.2, q=.01)$.95	-.21	1.03	-516.52
$GT(p=1.8, q=.01)$	1.12	-.22	.46	-476.78
$GT(p=1.3, q=.1)$	1.13	-.22	.52	-300.61
$GT(p=1.01, q=1)$.95	-.21	1.03	-69.31
t_1	.44	-.19	1.15	-234.94
t_2	.36	-.19	1.46	-228.46
LMS	1.82	-.24	1.68	
Huber	.39	-.18		
Tukey	.47	-.19		



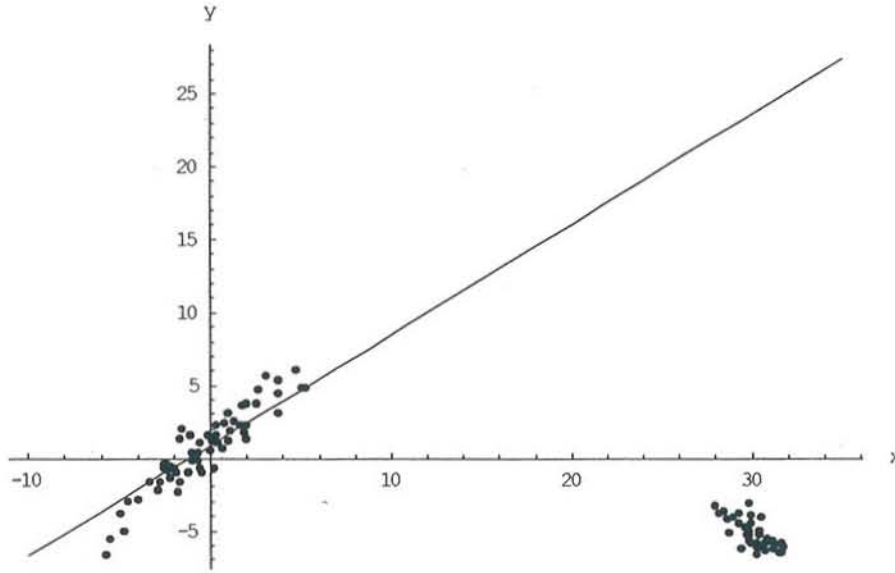
Şekil 2. Sapan değerlerin arasından geçen bir doğru.

Tablo 2. Örnek 1 için regresyon kestirimleri (*: sapan değersiz veri kümesi için)

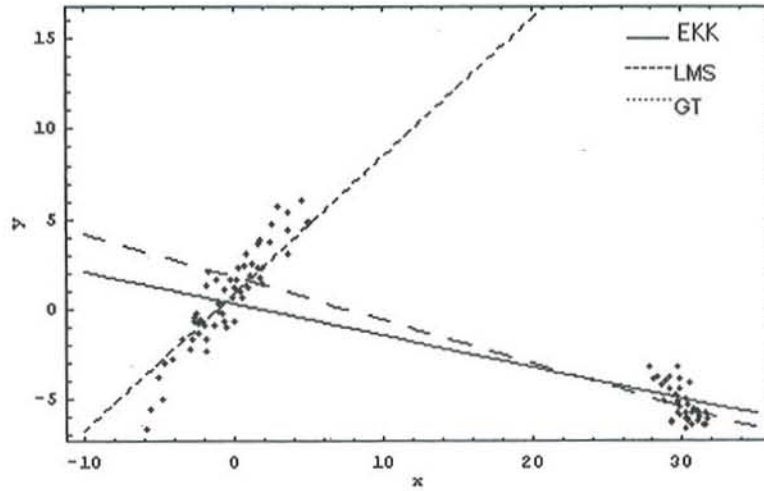
Yöntem	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}$	Log-lik
EKK*	.98	.97	.99	-84.43
$GT(p=1.2, q=.01)$.77	.90	.62	-516.52
$GT(p=1.8, q=.01)$.87	.76	.38	-476.78
$GT(p=1.3, q=.1)$.69	.63	.74	-300.61
$GT(p=1.01, q=1)$	-.65	-.16	1.04	-69.31
t_5	1.02	-.22	.76	-250.68
t_1	.45	-.19	1.15	-234.94
t_2	.36	-.19	1.46	-228.46
Huber	.39	-.18		
Tukey	.97	.97		

Eğer iterasyona başlangıç değeri olarak EKK* doğrusunu alırsak *GT* kestiricilerinin yeteri kadar küçük p and q değerleri için iyi gözlemlerin arasından geçecek şekilde doğrular ürettiğini buna karşılık bir başka sığara azalan *M*-kestiricisi Tukey'in dışındakilerinin sapan değerlerin etkisi altında kaldığını görürüz (Tablo 2 ve Şekil 3.)

Yüksek bozulma noktasına (breakdown point) sahip kestiriciler sınıfına dahil olan *LMS* kestiricisinin EKK gibi sapan değeri alt grubundan etkilendiğini ve iyi veri grubunu modelleyemediğini görmekteyiz (Tablo 1 ve Şekil 4.)



Şekil 3. İyi veri kümesi arasından geçen doğru.



Şekil 4. EKK, LMS ve GT regresyon doğrularının karşılaştırılması.

Örnek 2. Stackloss veri kümesi.

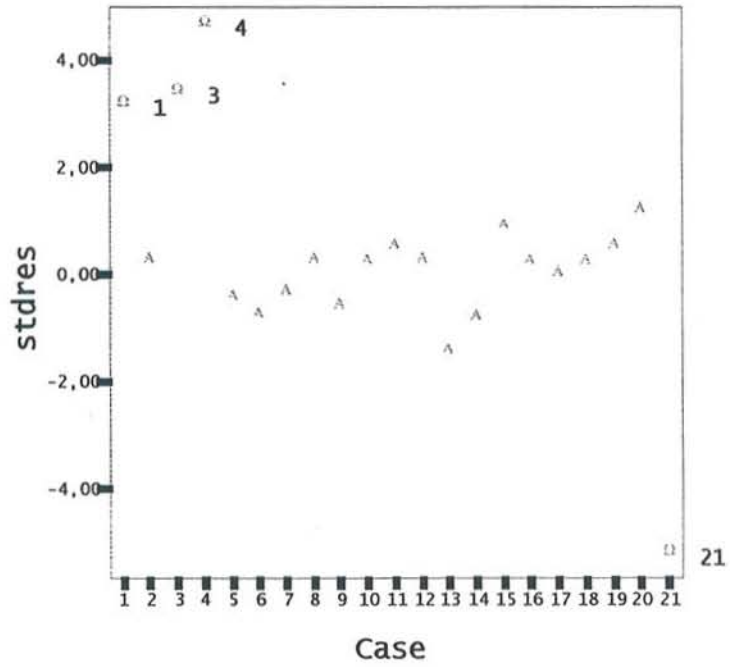
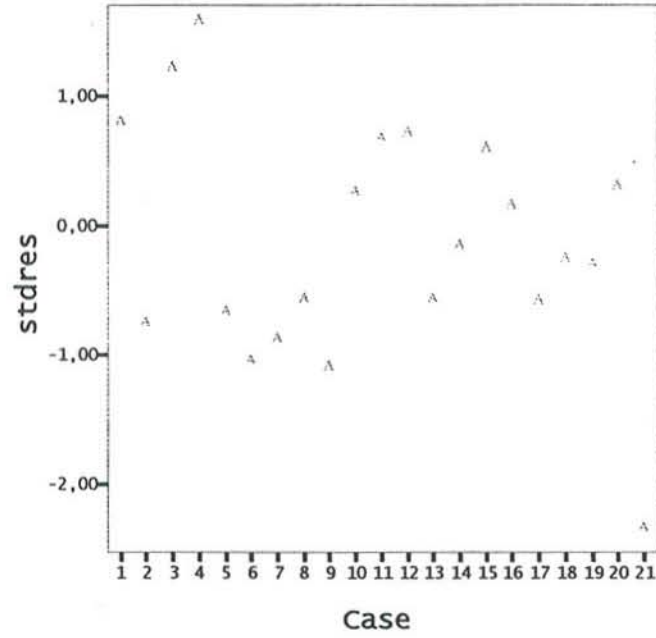
Bu veri kümesi bitkiler üzerine yapılan kimyasal çalışmalardan elde edilen gözlemlere dayanır. Her biri 21 gözlemlik 3 regresörden oluşan veri kümesine dayanlı kestirimle alakalı olarak literatürde sık rastlıyoruz (Örneğin, Lange vd., 1989, Andrews, 1974).

Model kurulup parametrelerin kestirimleri hesaplandığında EKK doğrusunu $\hat{y} = -39,92 + 0,72\hat{\beta}_1 + 1,30\hat{\beta}_2 - 0,15\hat{\beta}_3$ olarak buluruz. Bu doğruya ait endeks grafiği Şekil 5'te verilmiştir. Bu grafikte dikey eksendeki stdres standartlaştırılmış rezidüleri karşılık gelir. Grafikten EKK standartlaştırılmış rezidülerinin tümünün (-2.5; 2.5) aralığında bulunduğunu görürüz. Yani EKK yöntemine göre veri kümesinde sapan değer bulunmamaktadır.

Eğer iterasyona EKK doğrusu ile başlarsak Tablo 3'te verilen kestirimlere ulaşırız. $p=2.5$ ve $q=1$ için GT kestiricisi 4. ve 21. gözleme iterasyonun son adımında en küçük ağırlıklar olarak sırasıyla 0.4 ve 0.2 vermektedir. Dağılımın şekil parametrelerini biraz daha küçük seçtiğimizde nispeten daha küçük ağırlıklar alan gözlemlerin sayısı artar. Örneğin, $p=1.5$ ve $q=0.2$ için GT kestiricisi 1., 3., 4. ve 21. gözlemlerin hepsine son iterasyonda sıfır ağırlık vermektedir. Bu gözlemlerin birer sapan değer olduğunu $GT(p=1.5, q=.2)$ kestiricisine ait endeks grafiğinden de anlayabiliriz. Bu grafiğe göre 1., 3., 4. ve 21. gözlemler (-2.5;2.5) yatay bandının dışında kalmaktadır (Şekil.6.)

Tablo 3. Stackloss verisi için regresyon kestirimleri.

Yöntem	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\sigma}$	Loglik
EKK	-39.92	.72	1.30	-.15	2.97	-52.65
LMS	-34.25	.75	.50	0.00	1.21	
Huber	-40.75	.76	1.17	-.14		
Tukey	-41.33	.83	.95	-.13		
t_1	-38.63	.85	.49	-.07	.76	-49.58
t_2	-38.12	.85	.56	-.09	1.80	-50.31
$GT(p=2.5, q=1)$	-39.96	.86	.69	-.11	2.52	-14.56
$GT(p=1.5, q=.5)$	-40.47	.84	.55	-.05	.80	-33.28
$GT(p=1.5, q=.2)$	-39.94	.83	.57	-.06	.09	-49.23
$GT(p=1.3, q=.5)$	-40.09	.83	.56	-.06	.58	-34.34



Şekil 5. EKK endeks grafiği.

Şekil 6. $GT(p=1.5, q=.2)$ kestiricisine ait endeks grafiği.

Eğer bu 4 sapan değeri veri kümesinden atıp kalan gözlemler üzerine modeli kurarsak EKK* doğrusu $\hat{y} = -37.65 + .80x_1 + .58x_2 - .07x_3$ olarak elde edilir. Bu

doğrunun yeteri kadar küçük p ve q için bulunan ve Tablo 3'te verilen GT doğrularıyla benzer olduğunu görmekteyiz.

5. SONUÇ

Regresyonda hata teriminin normal dağılımlı olduğu yaygın olarak varsayılır. Fakat gerçekte hata dağılımı, özellikle de sapan değerlerin etkisiyle, normalden daha kalın kuyruklu olabilir. Bu durumda EKK analizi doğru sonuç vermeyecektir.

Biz bu çalışmada normal dağılıma alternatif olarak regresyonda hataları GT dağılımıyla modelledik. Dağılıma ait şekil parametrelerinin bilindiği varsayımı altında regresyon parametrelerini kestirdik. Elde ettiğimiz kestiriciler sıfıra azalan M-kestiricileri olup bunlar yinelemeli tekrar ağırlıklandırılmalı en küçük kareler formunda bulunmuştur. Dağılımın şekil parametreleri aynı zamanda dayanıklılık ayar sabitleri olup bunlar ağırlıkların sapan değerlere karşı hassaslığını kontrol etmektedir. Bulduğumuz GT kestiricilerinin performansını görmek ve diğer dayanıklı kestiricilerle karşılaştırmak için GT kestiricilerini problemleri veri kümelerinde kullandık. p ve q nun küçük olması durumunda ve iterasyona uygun bir başlangıç değerini seçilmesiyle GT kestiricilerinin iyi bir alternatif olabileceğini gördük. Ayrıca Örnek 1'de sapan değerlerin bir alt grup oluşturması durumunda da GT kestiricilerinin bozulmayacağını, dayanıklı kalabileceğini gördük.

KAYNAKLAR

- ANDREWS, D. F. (1974). *A robust method for multiple linear regression*. Technometrics, 16, 4, 523-531.
- ARSLAN, O. (1992). *Multivariate robust analysis based on the t distribution and the EM algorithm*. Unpublished PhD thesis, Leeds University, Leeds, U.K.
- ARSLAN, O., GENÇ, A. İ. (2002). *Robust location and scale estimation based on the univariate generalized t (GT) distribution*. (Submitted.)
- ARSLAN (2002). *A simple test to identify good solutions to redescending M-estimating equations for regression*. In Development in Robust Statistics, Proceedings of ICORS 2001. Edited by R. Dutter, U. Gather, P.J. Rousseeuw and P. Filzmoser, pp. 50-61.
- BLOOMFIELD, P., STEIGER, W. L. (1983). *Least Absolute Deviations Theory: Applications and Algorithms*. Boston: Birkhauser.
- BUTLER, R. J., McDONALD, J. B., NELSON, R. D., WHITE, S. B. (1990). *Robust and partially adaptive estimation of regression models*. The Review of Economics and Statistics, 72, 321-327.
- KLEIN, R., SPADY, R. (1984). *Quasi-maximum likelihood as a parametric approach to robust estimation*. working paper, Bell Communication Research.
- LANGE, K. L., LITTLE, J. A., TAYLOR, J. M. G. (1989). *Robust statistical modeling using the t distribution*. Journal of the American Statistical Association, 84, 881-896.

Koşullu Gauss Dağılımı ve Etkileşimleri

Hülya BAYRAK*

Fikri GÖKPINAR*

ÖZET

Grafiksel zincir modeller diye adlandırılan istatistiksel modeller, değişkenler arasında hem simetrik hem de nedensel ilişki içeren grafiklerin özel bir tipidir. Koşullu Gauss Dağılımı(CG) kesikli değişkenler verilmişken sürekli değişkenlerin bileşik Gauss dağılımı ve kesikli değişkenlerin her bir seviye kombinasyonunun pozitif olasılıkları ile tanımlanır. Bir CG dağılımında bir değişken çiftinin geri kalan değişkenler verilmişken koşullu bağımsız olabilmesi için gerek ve yeter koşul bu değişken çiftini ifade eden tüm etkileşim terimlerinin sıfır olmasıdır. Bu çalışmada koşullu Gauss etkileşimleri ve koşullu Gauss Zincir model tanıtılacaktır.

Anahtar Kelimeler:Koşullu Gauss dağılımı, Koşullu Gauss etkileşimleri, Zincir Grafik, Koşullu Gauss regresyonu

1. GİRİŞ

1.1 . Bağımlılık Zincirleri

Grafikler, koşullu bağımsızlık ilişkilerini formüle etmek için kullanılır. Grafikte; köşeler, rassal değişkenleri ve kenarlar bu değişkenler arası ilişkiyi gösterir. Grafik; köşeler kümesi ve kenarlar kümesinin terimleriyle belirlenir. Farklı köşelerin her kümesi en çok bir kenara sahiptir. Grafikte iki tip köşe vardır, bunlar çemberle gösterilen sürekli değişkenler ve noktayla gösterilen kesikli değişkenlerdir. (1)

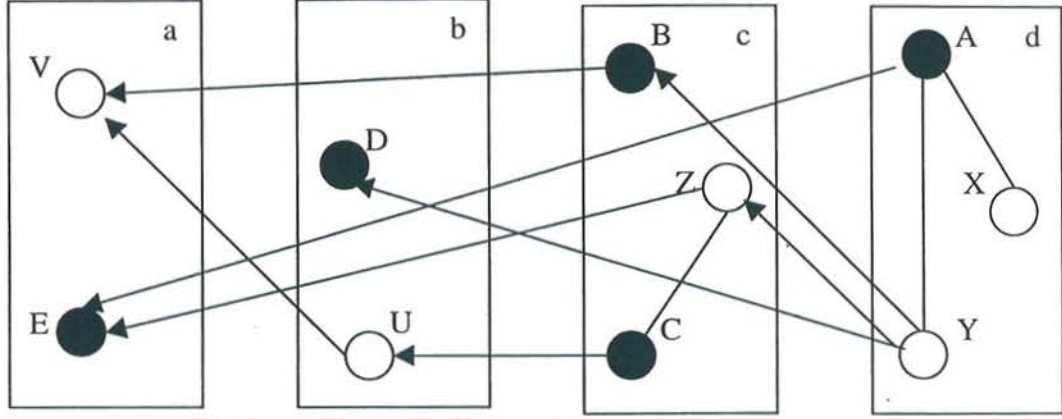
Değişkenler arasında iki çeşit ilişkiden bahsedilebilir; değişkenlerden biri, açıklanan değişken ve diğeri açıklayıcı değişken olduğunda, değişken çiftinin aralarında yönlü ilişki, değişkenlerin ikisi de açıklayıcı yada açıklanan değişkenler ise aralarında simetrik ilişki olduğu söylenir.

Tüm değişkenlerin kümesini gösteren grafik, alt kümelere ayrılabilir. Eğer alt kümeler için aşağıdaki koşullar sağlanıyorsa, bu grafiğe, bağımlılık zinciriyle ilişkilidir denir (2).

*Gazi Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü Ankara, hbayrak@gazi.edu.tr

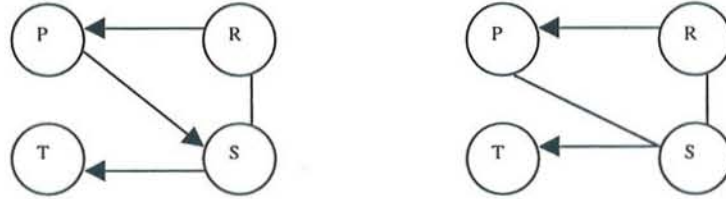
*Gazi Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü Ankara, fikri@gazi.edu.tr

- a) Alt kümeler yatay bir düzeyde sıralanır.
 b) Alt kümeler arasında sadece bir yönde oklar işaretlenir.
 c) Altkümeler içinde sadece çizgiler vardır.



Şekil 1.1 Zincir grafiklere ilişkin örnek

Bir grafik bir bağımlılık zinciri ile ilişkili ise, bu grafiğe zincir grafik denir. Tesadüfi (concurrent) değişkenler kümesi, bağımlılık zincirinde açıklanan değişken kümelerinin birer birer silinmesiyle elde edilir. Şekil 1.1'de tesadüfi değişkenlerin dört kümesi $a \cup b \cup c \cup d$, $b \cup c \cup d$, $c \cup d$ ile d 'dir. Grafiklerde aralarında çizgi yada ok olmayan değişkenler, geri kalan tesadüfi değişkenler hakkındaki bilgi verildiğinde, bu çiftler arasında koşullu bağımsızlık ilişkisi vardır şeklinde yorumlanır.



Şekil 1.2. Zincir grafiklere karşılık gelmeyen ilişki yapıları.

Yukarıdaki grafikler bloklar arası ok ve bloklar içi çizgi olacak şekilde düzenlenemediğinden zincir grafik değildir.

1.2. Koşullu Bağımsızlık Yapıları

Grafiğin tüm köşe çiftleri arasında kenar varsa grafik tamdır. Tam grafikte hiçbir koşullu bağımsızlık belirlenemez. Tüm değişkenler kümesi K ; Δ kesikli ve Γ sürekli değişkenler alt kümelerini belirtmek üzere $K = \Delta \cup \Gamma$ şeklinde yazılabilir.

Bir zincir grafikte bağımlılık zincirinden açıklanan değişkenler kümesi, tesadüfi değişkenler kümesi ve her kayıp kenarın anlamı yorumlanabilir.

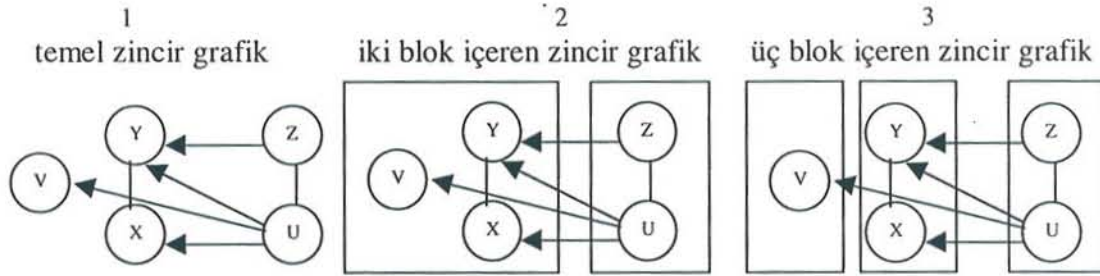
Şekil 1.1'de bağımlılık zinciri $\varphi = (a, b, c, d)$, köşe kümesi $K = a \cup b \cup c \cup d$, açıklanan değişken kümesi a, b, c, d ve 4 tesadüfi kümesi $a \cup b \cup c \cup d, b \cup c \cup d, c \cup d$ ve d 'dir.

Kayıp bir kenarı, geri kalan tüm tesadüfi değişkenler verilmişken o çiftler arasında koşullu bağımsızlık vardır şeklinde yorumlanabilir.

Örneğin, aşağıdaki seçilmiş çiftler için;

$$\begin{aligned} (V,E) &: V \perp E / (a \cup b \cup c \cup d \setminus \{V,E\}) \text{ yada } V \perp E / (A,B,C,D,X,Y,Z,U) \\ (D,B) &: D \perp B / (b \cup c \cup d \setminus \{D,B\}) \text{ yada } D \perp B / (A,C,X,Y,Z,U) \\ (X,Y) &: X \perp Y / (d \setminus \{X,Y\}) \text{ yada } X \perp Y / A \end{aligned}$$

Değişik bağımlılık zincirleri, aynı *temel zincir grafiğe* sahip olabilir. Örnek olarak, şekil 1.3'teki grafikler verilebilir. Araştırma hipotezini anlamak için (her kayıp kenarın belirli bir anlam belirttiği sürece) bağımlılık zincirini bilmek gerekir.



Şekil 1.3. İkinci grafikte (X,Z) kayıp kenarının anlamı $X \perp Z / (Y,U,V)$ ve üçüncü grafikteki (X,Z) kayıp kenarının anlamı $X \perp Z / (Y,U)$ 'dur.

Koşullu bağımsızlık grafiğinin üç sınıfı aşağıdaki gibidir: simetrik ortaklık grafiği, sadece bir açıklanan değişkeni olan ve çoklu açıklanan değişkene sahip olan bağımsızlık grafiği. Değişik durumlar arasındaki farkı ayırt etmek için, genel geri dönüşlü koşullu bağımsızlık grafiği G ile gösterilsin. Bir koşullu bağımsızlık grafiğinde;

- i) hiçbir açıklanan değişkene kümesi yoksa simetrik ortaklık grafiği (G^a)
- ii) en az bir açıklanan değişken kümesi birden fazla değişken içeriyorsa çoklu açıklanan grafiği (G^{mr})
- iii) tüm açıklanan değişken kümesi bir değişken içeriyorsa tek açıklanan grafiği (G^{sr})

şeklinde sınıflanabilir.

2. GRAFİKSEL KOŞULLU GAUSS (CG) ZİNCİR MODEL

Bileşik dağılım, tesadüfi değişkenlerin farklı kümelerini gerektiren dağılımlar cinsinden belirlenebiliyorsa, grafik, bir grafiksel zincir modele karşılık gelir. Zincir grafik bileşik dağılım üzerinde koşullu bağımsızlık kısıtlarını gösterir.

2.1 . Zincir Modelde Bileşik Dağılım

Bir bağımlılık zinciri, istatistiksel modelde iki farklı rol oynar: 1) Sistemdeki tüm değişkenlerin bileşik dağılımını elde etmenin bir yolunu belirtir. 2) Grafikteki kayıp kenara sahip olan her değişken çifti için koşullu bağımsızlık sınırlamasını tanımlar.

Bağımlılık zinciri T tane eleman içeriyorsa sistemdeki tüm değişkenlerin bileşik dağılımı f_v , T-1 tane koşullu yoğunluk ve bir tane marjinal yoğunluğun bileşiminden elde edilir(3). Örneğin şekil 1.1 için

$$f_v = f_{a|bcd} f_{b|cd} f_{c|d} f_d$$

yazılır. CG zincir modellerde, T-1 tane koşullu yoğunluk ve 1 tane marjinal yoğunluk CG tipindedir. g sembolü CG tipindeki yoğunlukları ayırt etmek için kullanılır. Şekil 1.1'de bir CG zincir modelin bileşik dağılımı aşağıdaki gibi yazılabilir.

$$g_v = g_{abcd} g_{bcd} g_{cd} g_d$$

2.2 .Koşullu Gauss Dağılımı

Değişkenler kümesi, $(K = \Gamma \cup \Delta)$ q tane sürekli değişken içersin. K için CG dağılımı kesikli değişkenler verilmişken sürekli değişkenlerin koşullu bileşik Gauss dağılımı ve kesikli değişkenler her seviyedeki kombinasyonuna ilişkin olasılıklar tarafından tanımlanır. Tüm değişkenlerin bileşik yoğunluğu moment özelliklerinin yardımıyla ifade edilebilir. Bunlar Π_l olasılıkları, μ_l ortalamaları ve kovaryans matrisleri Σ_l 'dir. Burada $l=1,2,\dots,L$ kesikli değişkenlerin seviye kombinasyonunu gösterir. Koşullu kovaryans matrisi kesikli değişkenleri seviye kombinasyonlarına bağlı olmadığı zaman, (örn. $\Sigma_l = \Sigma$) CG dağılımına, homojen denir. Bileşik yoğunluk, koşullu Gauss yoğunlukları $g_{\Gamma/\Delta}$ 'in ve $g_\Delta = \Pi_l$ marjinal olasılık fonksiyonunun bir ürünüdür.

$$g_v = g_{\Gamma/\Delta} = \left[\left\{ \frac{1}{2\pi} \right\}^{q/2} |\Sigma_l|^{-1/2} \exp \left\{ -\frac{1}{2} (x - \mu_l)^T \Sigma_l^{-1} (x - \mu_l) \right\} \right] \Pi_l$$

Eşdeğer olarak yoğunluğun logaritması kanoniksel özellikler cinsinden aşağıdaki gibi yazılabilir.

$$\ln g_v = d_l + h_l' x - \frac{1}{2} x' K_l x$$

Burada kanonik özellikler olan kesikli lineer ve kuadratik özellikler sırasıyla d_l , h_l ve K_l ile gösterilir.

İki özellik kümesi (d_l, h_l, K_l) ve (Π_l, μ_l, Σ_l) arasındaki ilişki şöyledir(4):

$$d_l = \log \Pi_l - \frac{1}{2} \left[q \ln(2\pi) + \ln |\Sigma_l| + \mu_l^T \Sigma_l^{-1} \mu_l \right], \quad h_l = \Sigma_l^{-1} \mu_l, \quad K_l = \Sigma_l^{-1} \quad (*)$$

(* ifadesinin ispatı için Eklere bakınız.)

2.3. Koşullu Gauss Dağılımlarının Etkileşimleri

Dolaylı ilişkiler ile ilgili araştırma hipotezlerinin değerlendirilmesinde CG dağılımının önemli bir özelliği etkileşimlerle araştırma hipotezlerinin parametrelendirilmesidir.

CG dağılımlarında bir değişken çiftinin geri kalan değişkenler üzerinden koşullu bağımsız olması sadece ve sadece bu değişken çiftlerini içeren tüm etkileşim terimlerinin sıfır olmasına bağlıdır (5).

Etkileşimlerle CG dağılımının parametrelenmesi, $K=\{A,B,X,Y\}$ olmak üzere, $\varphi=(K)$ için örneklenmiştir (burada $\Delta=\{A,B\}$ A, B'in sırasıyla $i=1,2,\dots,I$ ve $j=1,2,\dots,J$ kategorilerine sahip olsun ve sürekli değişkenler kümesi $\Gamma=\{X,Y\}$ şeklinde verilsin). Bu durumda, CG yoğunluğu kanoniksel özellikler terimleriyle aşağıdaki gibi yazılabilir.

$$\ln g(i, j, x, y) = d_{ij} + h_{ij}^x x + h_{ij}^y y - \frac{1}{2} k_{ij}^x x^2 - \frac{1}{2} k_{ij}^y y^2 - k_{ij}^{xy} xy \quad (2.1)$$

ve etkileşim terimleriyle şöyle ifade edilebilir.

$$\begin{aligned} \ln g(i, j, x, y) = & \lambda + (\lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}) + (\eta^X + \eta_i^{AX} - \eta_j^{BX} + \eta_{ij}^{ABX})x + (\eta^Y + \eta_i^{AY} + \eta_j^{BY} + \eta_{ij}^{ABY})y - \\ & - \frac{1}{2}(\Psi^X + \Psi_i^{AX} + \Psi_j^{BX} + \Psi_{ij}^{ABX})x^2 - \frac{1}{2}(\Psi^Y + \Psi_i^{AY} + \Psi_j^{BY} + \Psi_{ij}^{ABY})y^2 - \\ & - (\Psi^{XY} + \Psi_i^{AXY} + \Psi_j^{BXY} + \Psi_{ij}^{ABXY})xy \quad (**) \end{aligned}$$

(** ifadesinin ispatı için Eklere bakınız)

$d \subseteq \Delta$ olmak üzere buradaki λ , η , ψ etkileşim terimlerine aşağıdaki gibi yorumlanabilir.

λ^d : $|d|=0$ ise sabit, $|d|=1$ ise kesikli değişkenin ana etkisi, $|d|>1$ ise d'deki kesikli değişkenlerinin birbirleriyle etkileşimlerini gösterir.

η^d : $|d|=0$ ise sürekli değişkenin ana etkisi, $|d|\neq 0$ ise sürekli değişken ile d'deki kesikli değişkenlerin karma lineer etkileşimleri gösterir.

ψ^d : $|d|=0$ ise sürekli değişkenlerin saf kuadratik etkisi, $|d|\neq 0$ sürekli değişken çifti ile d'deki kesikli değişkenler arası kuadratik etkileşimi tanımlar. (5)

Modellerin etkileşimleri için aşağıdaki sonuçlar çıkarılabilir.

Sonuç 2.1:

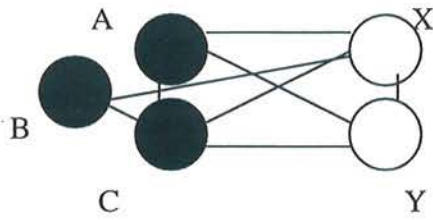
Simetrik ortaklık grafik G^a 'da, değişkenler kümesi $\Delta \cup \Gamma$ 'den her çift için, aşağıdaki ifadelere karşılık gelir.

- i) Değişken çifti, diğer tüm değişkenler verilmişken koşullu olarak bağımsızdır.
- ii) Değişken çiftini içeren tüm etkileşimler sıfıra eşit olur.
- iii) Değişken çiftinin kenarı simetrik ortaklık grafiğinde yoktur.

Aşağıdaki tabloda bazı koşullu bağımsızlıklar ve etkileşimler verilmiştir.

Tablo 2.1
Sonuç 2.1 için örnekler

Simetrik Ortaklık Grafiği	Karşılık Gelen Bağımsızlıklar	Sıfır Olan Etkileşimlerin Kümesi
<p>Şekil 2.1</p>	$A \perp B / (B, C, Y)$	$S_1 = \{ \lambda_{ij}^{AB}, \lambda_{ijk}^{ABC} \}$ $\cup \{ \eta_{ij}^{ABX}, \eta_{ijk}^{ABCX} \}$ $\cup \{ \eta_{ij}^{ABY}, \eta_{ij}^{ABY} \}$ $\cup \{ \Psi_{ij}^{ABX}, \Psi_{ijk}^{ABCX} \}$ $\cup \{ \Psi_{ij}^{ABY}, \Psi_{ijk}^{ABY} \}$ $\cup \{ \Psi_{ij}^{ABXY}, \Psi_{ijk}^{ABCXY} \}$



Şekil 2.2

$$A \perp B / (B, C, Y) \quad S_2 = S_1 \cup \{ \eta_j^{BY}, \eta_{ij}^{ABY}, \eta_{ijk}^{ABCY} \}$$

$$\text{ve} \quad \cup \{ \Psi_j^{BY}, \Psi_{ij}^{ABY}, \Psi_{ijk}^{ABCY} \}$$

$$Y \perp B / (A, C, X) \quad \cup \{ \Psi_j^{BXY}, \Psi_{ij}^{ABXY}, \Psi_{ijk}^{ABCXY} \}$$

Tablo 2.1'deki Şekil 2.1'de A ile B arasındaki kayıp kenarı $A \perp B / (B, C, Y)$ koşullu bağımsızlığını ifade eder. Bu koşullu bağımsızlığın anlamı tabloda verildiği gibi A ve B'yi içeren tüm etkileşimlerin 0 olmasıdır. Benzer olarak Şekil 2.2'de (A, B) ile (B, Y) içeren tüm etkileşimler sonuç 2.1'den dolayı 0 olur.

2.4 . Koşullu Gauss Regresyonu

Bir CG zincir modelinde koşullu yoğunluklara CG regresyonu denir. CG regresyon bir koşullu dağılımdır (açıklanan ve açıklayıcı değişkenler kümesindeki değişkenlerin bileşik dağılımını bileşik CG dağılımını belirtir). Bu koşullu dağılımın yoğunluğu kesikli, lineer ve kuadratik kanoniksel ifadelerin cinsinden açıklanabildiği sürece, CG tipindedir.

Bir tek değişkenli CG regresyonu, eğer açıklanan değişken sürekli ise, bir *lineer regresyondur*, ve eğer açıklanan değişken kesikli ise (etkileyen değişkenler üzerinden lineer ve kuadratik bağımlılıkla birlikte) *lojistik regresyondur*. Bir homojen CG regresyonu, homojen CG dağılımından türetilir. Bu durumda, lineer regresyonlar

paraleldir. Örneğin eşitlik (2.1)'den A,B ve X verilmiş iken Y'nin lineer regresyonu aşağıdaki gibi verilir (6)

$$E(Y/x) = \alpha_{ij} + \beta_{ij}x \quad \text{Var}(Y/x) = 1/k_{ij}^y \quad \alpha_{ij} = h_{ij}^y / k_{ij}^y \quad \beta_{ij} = -k_{ij}^{xy} / k_{ij}^y \quad (***)$$

(* ** ifadesinin ispatı için Eklere bakınız)

2.5 . Koşullu Gauss Zincir Modelinde Bileşik Dağılım

Bir CG modelinin varsayımları aşağıdaki gibidir.

- (a) Bileşik dağılımdaki T-1 tane koşullu yoğunlukların tümü, CG regresyonlardır.
- (b) Açıklanan değişken içermeyen değişkenler kümesinin marjinal dağılımı, bir CG dağılımıdır.

Sadece bir açıklanan değişkene sahip olan CG zincir modele G^{sr} grafiği tarafından verilen *tek değişkenli geri dönüşlü regresyon modeli* denir. Çoklu açıklanan değişkene sahip olan modele *blok geri dönüşlü regresyon modeli* (G^{mr} tarafından verilen) denir. Her grafiksel zincir model T, simetrik ortaklık grafiğinin yardımıyla yorumlanabilir: grafiksel zincir model tamamıyla bağımlılık zincirine bağlıdır.

3. SONUÇ

Dolaylı ilişkiler hakkındaki araştırma hipotezlerinin değerlendirilmesinde CG dağılımının önemli bir özelliği etkileşimleriyle birlikte araştırma hipotezlerinin parametrelenmesidir. CG dağılımındaki etkileşim fikri, varyans analizini modelleri kavramında kullanılan etkileşim fikriyle ilişkili fakat farklıdır (7).

Eğer bileşik dağılım tesadüfi değişkenlerin değişik kümelerini gerektiren dağılımlar cinsinden belirlenebiliyorsa, şekil 1.1 gibi bir grafik, grafiksel zincir modele karşılık gelir ve zincir grafik bileşik dağılım üzerinde koşullu bağımsızlık kısıtlarını gösterir. Genel olarak değişik bağımlılık zincirleriyle tamamlanmış grafikler, değişkenlerin aynı koleksiyonu için değişik doymuş model tanımlar. Bir CG zincir modelinin tüm dağılımları CG dağılımları olmasına rağmen, bunun bileşik dağılımının kendisinin CG dağılımı olmasına gerek yoktur.

EKLER

(*)'in ispatı

$$(*) \quad g_v = \left[\left\{ \frac{1}{2\pi} \right\}^{q/2} |\Sigma_t|^{-1/2} \exp \left\{ -\frac{1}{2} (x - \mu_t)^T \Sigma_t^{-1} (x - \mu_t) \right\} \right] \Pi_t$$

denklemin logaritması alındığında

$$\ln g_v = \log \Pi_t - \frac{1}{2} \left\{ q \log(2\pi) + \log |\Sigma_t| + \mu_t^T \Sigma_t^{-1} \mu_t \right\} + \Sigma_t^{-1} \mu_t x - \frac{1}{2} x^T \Sigma_t^{-1} x$$

olur.

$$\log g_v = d_t + h_t' x - \frac{1}{2} x^T K_t x \text{ olduğu bilindiğinden;}$$

kanonik özellikler ile moment özellikleri arasındaki ilişkiler

$d_i = \log \Pi_i - \frac{1}{2} \{ q \log(2\pi) + \log |\Sigma_i| + \mu_i^T \Sigma_i^{-1} \mu_i \}$, $h_i = \Sigma_i^{-1} \mu_i$, $K_i = \Sigma_i^{-1}$
biçimindedir.

(**)’ın ispatı

Burada ilk önce Möbius inversion tanımlamak faydalı olur.
H ve J sonlu A kümesinin bir alt kümesi a üzerinde tanımlanan fonksiyonlar olsun. O zaman aşağıdaki eşitlikler birbirine denktir.

$$(i) \quad \forall a \subseteq A: H(a) = \sum_{b \subseteq a} J(b)$$

$$(ii) \quad \forall a \subseteq A: J(a) = \sum_{b \subseteq a} (-1)^{|a/b|} H(b)$$

öncelikle $d \subseteq \Delta$ olmak üzere d’nin alabileceği değerlerden oluşan bir ζ kümesi tanımlansın. Sabit fakat keyfi bir $i^* \in \zeta$ seçilsin. $i \in \zeta$ için $i(d) \in \zeta$ ifadesine değiştirme elemanı densin.

$$i(d)_{\delta} = \begin{cases} i_{\delta} & \delta \in d \\ i_{\delta}^* & \delta \notin d \end{cases}$$

örneğin $i(\Delta) = i$ ve $i(\emptyset) = i^*$ şeklinde verilebilir.

CG dağılımındaki kanoniksel özellikleri aşağıdaki gibi yazılabilir.

$$p_d(i) = d(i(d)), \quad \xi_d(i) = h(i(d)), \quad \Phi_d(i) = K(i(d)) \quad (E1)$$

$a \subseteq \Delta$ iken

$$\lambda_a(i) = \sum_{d \subseteq a} (-1)^{|a/d|} p_d(i)$$

$$\eta_a(i) = \sum_{d \subseteq a} (-1)^{|a/d|} \xi_d(i)$$

$$\Psi_a(i) = \sum_{d \subseteq a} (-1)^{|a/d|} \Phi_d(i) \quad (E2)$$

yukarıdaki denklemde özel olarak $d = \Delta$ ve $i(\Delta) = i$ alındığında

$$p_d(i) = d(i) = \sum_{b \subseteq \Delta} \lambda_b(i)$$

$$\xi_d(i) = h(i) = \sum_{b \subseteq \Delta} \eta_b(i)$$

$$\Phi_d(i) = K(i) = \sum_{b \subseteq \Delta} \Psi_b(i),$$

elde edilir. Burada $d_{ij} = d(i)$, $h_{ij} = h(i)$, $K_{ij} = K(i)$ dir. Burada sadece λ ’lar bulundu. Benzer şekilde η, ξ değerleri de bulunabilir.

$$p_{AB}(i, j) = d(i, j) = \lambda_{\phi}(i^*, j^*) + \lambda_A(i, j^*) + \lambda_B(i, j^*) + \lambda_{AB}(i, j)$$

$$\lambda_{\phi}(i^*, j^*) = p_{\phi}(i^*, j^*)$$

$$\begin{aligned}\lambda_A(i, j^*) &= -p_\phi(i^*, j^*) + p_A(i, j^*) \\ \lambda_B(i^*, j) &= -p_\phi(i^*, j^*) + p_B(i^*, j) \\ \lambda_{AB}(i, j) &= p_\phi(i^*, j^*) - p_A(i, j^*) - p_B(i^*, j) + p_{AB}(i, j)\end{aligned}$$

(***) ispatı

Bir regresyon denklemi ve katsayıları aşağıdaki gibidir.

$$E(Y/x) = \alpha(i, j) + \beta(i, j)x, \quad \beta(i, j) = \frac{\sigma_{xy}(i, j)}{\sigma_{xx}(i, j)}, \quad \alpha(i, j) = \mu_{ij}^y - \mu_{ij}^x \beta(i, j), \quad \text{Var}(Y/x) = \sigma_{yy}(i, j)$$

$$h(i, j) = (h_{ij}^x, h_{ij}^y) = \Sigma(i, j)^{-1} \mu(i, j) = (k_{ij}^x \mu_{ij}^x + k_{ij}^{xy} \mu_{ij}^y, k_{ij}^{xy} \mu_{ij}^x + k_{ij}^y \mu_{ij}^y) \quad (\text{E3})$$

$$\Sigma_{ij} = \begin{pmatrix} \sigma_{xx}(i, j) & \sigma_{xy}(i, j) \\ \sigma_{xy}(i, j) & \sigma_{yy}(i, j) \end{pmatrix}$$

olduğundan konsantrasyon matrisi aşağıdaki gibi yazılabilir.

$$K(i, j) = \Sigma(i, j)^{-1} = \begin{pmatrix} k_{ij}^x & k_{ij}^{xy} \\ k_{ij}^{xy} & k_{ij}^y \end{pmatrix} = \frac{1}{\sigma_{xx}(i, j)\sigma_{yy}(i, j) - \sigma_{xy}(i, j)^2} \begin{pmatrix} \sigma_{yy}(i, j) & -\sigma_{xy}(i, j) \\ -\sigma_{xy}(i, j) & \sigma_{xx}(i, j) \end{pmatrix} \quad (\text{E4})$$

$$\text{(E4)'den } \beta(i, j) = \beta_{ij} = \frac{-k_{ij}^{xy}}{k_{ij}^y} \text{ bulunur.}$$

$$\text{(E3) ve (E4)'den } \alpha(i, j) = \frac{h_{ij}^y}{k_{ij}^y} \text{ bulunur.}$$

KAYNAKLAR

Lauritzen S.L. *Graphical Models*. Clarendon press, New York, (1996).

Erbaş S.O., Bayrak H. . *Grafiksel Zincir Modeller*, İstatistik Sempozyumu 2000, 273-284,(2000).

Lauritzen S.L and Richardson T.S., *Chain graph models and their casual interpretation*, Department of Mathematical Science, Aalborg university, Research report R-01-2003, (2001).

Lauritzen S.L. and Jensen F., *Stable local computation with conditional gaussian distribution*, Department of Mathematical Science, Aalborg university, Research report R-99-2014, (1999).

Lauritzen , S.L and Wermuth, N. *Graphical models for associations between variables*, some of which are qualitative and some quantative. *Annals of Statistics*, **17**, 31-57, (1989).

Wermuth, N. And Lauritzen, S.L. On substantive research hypotheses, conditional independence graphs and graphical chain models (with discussion). *Journal of Royal Statistical Society, Series B* , 52, 21-72, (1990).

Cox, D.R. "Interaction" *Int. Statist. Rev.*, 52, 1-31, (1984).

Conditional Gaussian Distribution and Interactions

ABSTRACT

Statistical models called graphical chain models correspond to special types of graphs which included both symmetric and casual relation between variables. A conditional Gaussian (CG) distribution is defined by a joint Gaussian distribution of the continuous variables given discrete variables and by positive probabilities for each level combination of the discrete variables. In CG distribution, a variable pair is conditionally independent given remaining variables if and only if all interaction terms containing this variable pair are zero. In this study, we present CG chain models and CG interactions.

Key Words: *Conditional Gaussian distributions, Conditional Gaussian interactions, Chain graph, Conditional Gaussian regression*

Üçgensel Olumsuzluk Tablolarında Yarı-Bağımsızlık Modeli için Güç Analizi

Serpil AKTAŞ*

Tülay SARAÇBAŞI

ÖZET

Üçgensel olumsuzluk tabloları tamamlanmamış olumsuzluk tablolarının özel bir durumudur. Üçgensel tablolarda ana köşegen elemanlarının alt ya da üst üçgen kısımlarında yapısal sıfırlı elemanlar bulunur. Bu tür tablolarda satır ve sütun değişkenleri arasındaki bağıntı yarı-bağımsızlık modeli ile araştırılır. Yarı-bağımsızlık modeli test edilirken ki-kare ya da olabilirlik oran istatistiklerinden yararlanılır. H_0 hipotezi doğru olmadığında bu test istatistikleri yaklaşık olarak merkezsiz olmayan ki-kare dağılımı göstermektedir. Bu çalışmada iki değişkenli normal dağılımdan yararlanarak türetilen 30.000 üçgensel tabloda yarı-bağımsızlık modeli test edilmiştir. Bu modele uyan ve uymayan tablo yüzdeleri verilmiştir. Yarı -bağımsızlık modelinden elde edilen olabilirlik oran istatistiklerinin testleri için güç değerleri hesaplanmıştır. Yarı-bağımsızlık modelinin reddedildiği tablo yüzdeleri ile güç değerleri karşılaştırılmıştır.

Anahtar Kelimeler : Olumsuzluk tablosu , Yarı-bağımsızlık modeli , Güç analizi

1.GİRİŞ

Tamamlanmamış olumsuzluk tablolarını ilk defa Bishop-Fienberg (1969) ve Mantel (1970) tanıtmışlardır. Üçgensel(triangular) olumsuzluk tabloları tamamlanmamış olumsuzluk tablolarının özel bir durumudur.Üçgensel olumsuzluk tablolarında sıklıklar ana köşegen elemanlarının alt ya da üst üçgen kısımlarında yer alır. Sarkar(1989) aşağıdaki koşulları sağlayan üst-sağ, üst-sol, alt-sağ ve alt-sol olmak üzere dört çeşit üçgensel olumsuzluk tablo tanımı yapmıştır.

Üst-sağ (sol) üçgensel olumsuzluk tablosu π_{ij} , i. satır ve j. kolon'a karşılık gelen gözlenen sıklık olmak üzere, $\pi_{ij} = 0 \quad i > j$, ve alt-sol(sağ) üçgensel olumsuzluk tablosu ise $\pi_{ij} = 0 \quad i < j$ olarak tanımlanır. Üst-sağ üçgensel olumsuzluk tablosu Tablo 1'de verilmiştir.

* Hacettepe Üniversitesi, Fen Fakültesi İstatistik Bölümü Beytepe-Ankara, e-mail: spxl@hacettepe.edu.tr

Tablo 1. Üst-Sağ Üçgensel Olumsuzluk Tablo Yapısı

i/j	1	2	3	...	R
1	π_{11}	π_{12}	π_{13}	...	π_{1R}
2		π_{22}	π_{23}	...	π_{2R}
3			π_{33}	...	π_{3R}
...			
R				...	π_{RR}

Üçgensel olumsuzluk tablolarında tablonun tamamlanmamış yapısı nedeniyle satır ve kolon arasındaki bağımsızlık ilk kez Goodman(1968) tarafından tanımlanan yarı-bağımsızlık(YB) modeli ile incelenir. YB modeli üst-sağ üçgensel olumsuzluk tablosu,

$$\begin{aligned} \pi_{ij} &= \alpha_i \beta_j & i \leq j \\ &= 0 & i > j \\ i &= 1, \dots, R, \text{ ve } j = 1, \dots, R \end{aligned} \quad (1)$$

olarak tanımlanmıştır.

Burada $\alpha_i > 0$ ve $\beta_j > 0$ pozitif sabitlerdir.

Sarkar(1989) yarı-bağımsızlık modelini sıralanabilir ilişki cinsinden uyumlu(concordant) C_π ve uyumsuz(discordant) D_π gözlem çiftleri arasındaki farka dayanan ve Eşitlik 2'de verilen η_π ile

$$\begin{aligned} C_\pi &= 2 \sum_{i < k} \sum_{j < l} \pi_{ij} \pi_{kl} \quad , \quad D_\pi = 2 \sum_{i > k} \sum_{j > l} \pi_{ij} \pi_{kl} \\ \eta_\pi &= 2 \sum_{i < k} \sum_{j < l} (\pi_{ij} \pi_{kl} - \pi_{il} \pi_{kj}) \end{aligned} \quad (2)$$

tanımlanmıştır.

Burada $\eta_\pi = 0$ ise yarı-bağımsızlık modelinin varlığından söz edilir. Dolayısıyla yarı-bağımsızlık modeli için $H_0: \eta_\pi = 0$ hipotezine karşılık $H_s: \eta_\pi > 0$ ($\eta_\pi < 0$) seçenek hipotezi ile test edilebilir. H_0 hipotezini H_s 'e karşı test etmek için uygun test istatistiği Eşitlik (3) ile,

$$z = \frac{\sqrt{n} \hat{\eta}_\pi}{\hat{\sigma}_\pi} \quad (3)$$

tanımlanır.

z 'nin dağılımı $N(0,1)$ 'dir (Sarkar, 1989). η_π 'nin dağılımı ile ilgili teorem:

Teorem: Çokterimli(multinomial) örneklem altında $\sqrt{n}(\hat{\eta}_{\pi} - \eta_{\pi})$ 'nin dağılımı $N(0, \sigma_{\pi}^2)$ dir.

$$\sigma_{\pi}^2 = \sum_{i \leq j} \phi_{\pi,ij}^2 \pi_{ij} - 4\eta_{\pi}^2 \quad (4)$$

Eşitlik (4) deki $\phi_{\pi,ij}$ $i \leq j$ için

$\phi_{\pi,ij} = 2[P(X < i, Y < j) + P(X > i, Y > j) - P(X < i, Y > j) - P(X > i, Y < j) - P(X > j) - P(Y < i)]$ biçiminde ifade edilir H_0 doğru iken σ_{π}^2 nin en çok olabilirlik tahmini $\hat{\sigma}_{\pi}^2$ 'dir Sarkar(1989).

2. OLUMSALLIK TABLOLARINDA GÜÇ ANALİZİ

Testin gücü bilindiği gibi seçenек hipotezi doğru olduğunda yokluk hipotezini reddetme olasılığıdır ve $Güç=1-P(II.Tip Hata)$ olarak tanımlanır. Yokluk hipotezi doğru olmadığında χ^2 ve G^2 gibi test istatistiklerinin dağılımı büyük örneklemelerde yaklaşık olarak merkezselsel olmayan ki-kare dağılımı gösterir. Merkezi olmayan ki-kare dağılımının olasılık yoğunluk fonksiyonu χ'^2 aşağıdaki şekilde tanımlanır.

X_1, X_2, \dots, X_v raslantı değişkenlerinin herbiri normal dağılım gösterebilir ve eğer a_1, a_2, \dots, a_v sabitleri $\lambda = \sum_{i=1}^v a_i^2$ ve $\chi'^2 = \sum_{i=1}^v (X_i - a_i)^2$ ise yoğunluk fonksiyonu

$$f(\chi'^2) = \frac{\exp\left\{-\frac{1}{2}(\chi'^2 + \lambda)\right\}}{2^{\frac{1}{2}v}} \sum_{j=0}^{\infty} \frac{(\chi'^2)^{\frac{1}{2}v+j-1} \lambda^j}{\Gamma\left(\frac{1}{2}v + j\right) 2^{2j} j!} \quad (5)$$

ile verilir. Burada v serbestlik derecesidir (Pearson and Hartley, 1976). Agresti (1990)'nin tanımladığı merkezselsel olmayış parametreleri λ , G^2 için Eşitlik (6)'da χ^2 için Eşitlik (7)'de vermiştir. Eşitlikler bağımsızlık modeli için tanımlanmıştır.

$$\lambda = 2n \sum_i \sum_j \pi_{ij} \log \frac{\pi_{ij}}{\pi_{ij(m)}} \quad (6)$$

$$\lambda = n \sum \frac{(\pi_{ij} - \pi_i \pi_j)^2}{\pi_i \pi_j} \quad (7)$$

Eşitlik (6)'da $\pi_{ij(m)}$, π_{ij} 'nin en çok olabilirlik tahminidir

$\lambda = 0$ olduğu durum merkezselsel ki-kare dağılımını verir. H_0 hipotezi doğru olduğunda $\pi_{ij} = \pi_i \pi_j$ olur.

Üçgensel olumsuzluk tablolarının çözümlenmesinde yararlanılan yarı-bağımsızlık modeli benzetim çalışması ile türetilen tablolarda çözümlenmiş ve yarı bağımsızlık testlerinin güçleri hesaplanmıştır.

3. BENZETİM ÇALIŞMASI

Yarı bağımsızlık modelini ilişki cinsinden incelemek ve güçlerini hesaplamak için benzetim çalışması yapılmıştır. Goodman (1981,1985) çalışmalarında ilişkisiz olumsuzluk tablolarını türetebilmek için korelasyon katsayısı sıfıra yakın iki değişkenli normal dağılımlı kitlelerden çekilen örneklemelerden yararlanmıştır. Üçgensel olumsuzluk tablolarında yarı bağımsızlık modeline uyumu belirleyen G^2 istatistiği ile Sarkar'ın önerdiği yöntemi karşılaştıran, test istatistiklerinin güç değerlerini hesaplayan benzetim çalışmasında bağımsız üçgensel tablolar oluşturabilmek için Goodman (1981), Saraçbaşı ve Muluk (1985) çalışmalarından yararlanılmıştır. Kitle korelasyon katsayısı (ρ), 0.0, 0.2, 0.4 olan iki değişkenli normal dağılımlı kitlelerden örneklem büyüklüğü (n), 100, 250, 500, 1000 ve tablo boyutu (R), 4,5,6,7 ve 8 olan üst-sağ üçgensel olumsuzluk tabloları türetilmiştir. Herbir kombinasyondan 500 tekrar yapılarak toplam 30.000 üst-sağ üçgensel olumsuzluk tablosu oluşturulmuştur (Aktaş, 1999). Türetilen 30.000 tablonun YB modeline uyumu G^2 ve Sarkar (1989)'ın önerdiği Z değerlerine göre incelenmiştir. Tabloların tümünde güç değerleri herbir örneklemde 500 tekrardan elde edilen veriler için S-Plus programından yararlanılarak hesaplanmıştır. Yarı-bağımsızlık modeli varsayımı altında olabilirlik oran istatistiği için güç değerleri Tablo 2'de verilmiştir. Güç değerlerinin benzetim parametrelerine göre değişimi ise Şekil 1-2 ve 3'de görülmektedir.

Tablo 3'de ise YB modeli için 500 tekrarda elde edilen reddedilen ve kabul edilen hipotezlerin oranları verilmiştir. Sarkar (1989) yöntemine göre hesaplanan $\hat{\eta}_\pi$ değerlerinin 500 tekrar için ortalama ve standart hata değerleri Tablo 4'de verilmiştir.

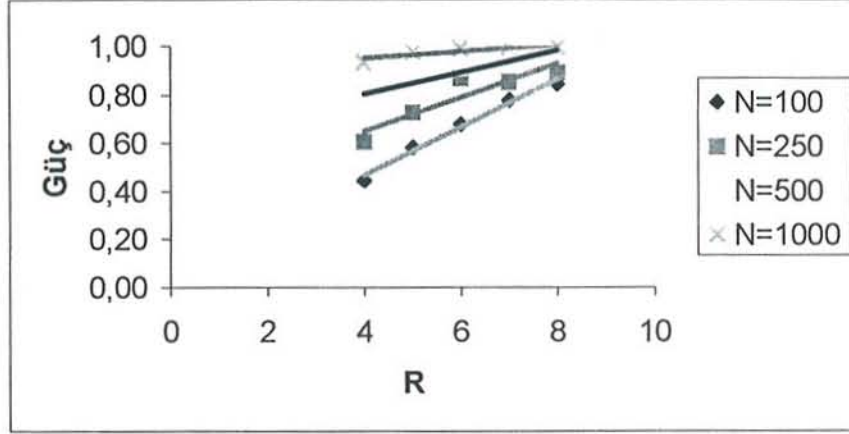
Tablo 2. YB Modeli İçin Olabilirlik Oran İstatistiğinin Güç Değerleri

n=100			
R	ρ		
	0.0	0.2	0.4
4	0.4464	0.4563	0.4359
5	0.5829	0.5830	0.5602
6	0.6797	0.6642	0.6720
7	0.7789	0.7679	0.7889
8	0.8454	0.8348	0.8368

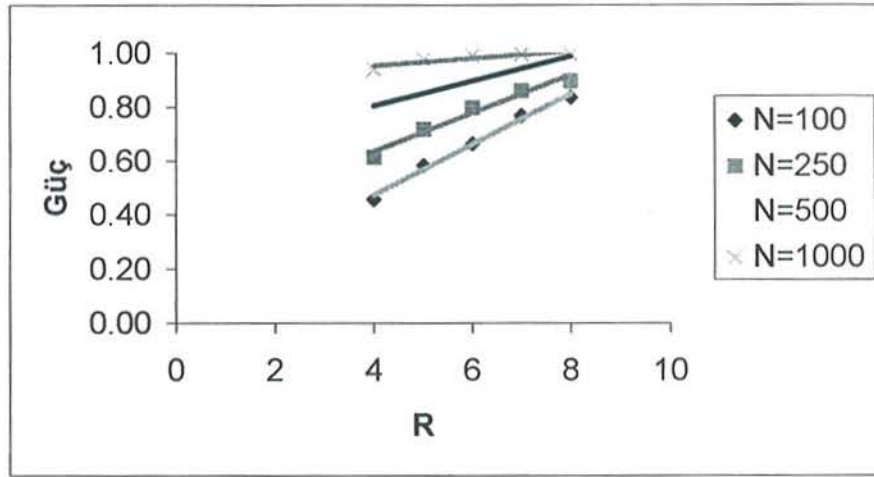
n=250			
R	ρ		
	0.0	0.2	0.4
4	0.6077	0.6131	0.6123
5	0.7292	0.7180	0.7277
6	0.8690	0.7967	0.7942
7	0.8533	0.8622	0.8503
8	0.8916	0.8969	0.8954

n=500			
R	ρ		
	0.0	0.2	0.4
4	0.7777	0.7693	0.7902
5	0.8731	0.8798	0.8960
6	0.9118	0.9171	0.9247
7	0.9398	0.9509	0.9366
8	0.9646	0.9630	0.9405

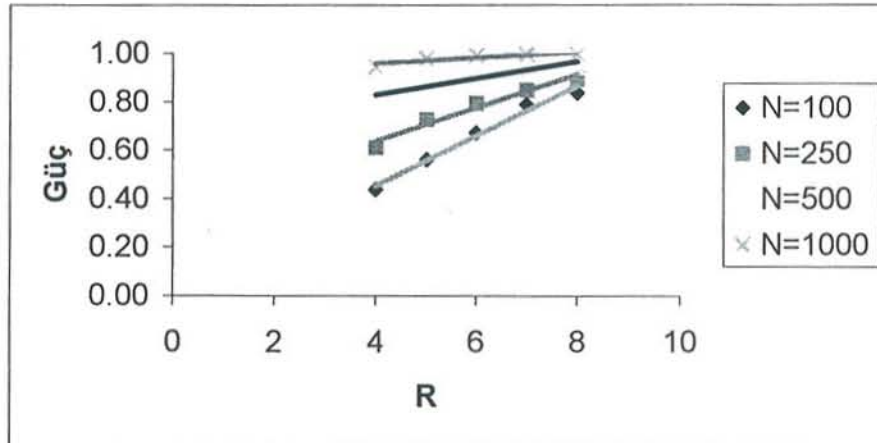
n=1000			
R	ρ		
	0.0	0.2	0.4
4	0.9337	0.9384	0.9453
5	0.9764	0.9776	0.9796
6	0.9912	0.9887	0.9914
7	0.9943	0.9961	0.9958
8	0.9964	0.9963	0.9954



Şekil 1. $\rho=0.0$ için Güç Değerleri



Şekil 2. $\rho=0.2$ için Güç Değerleri



Şekil 3. $\rho=0.4$ için Güç Değerleri

Tablo 2 incelendiğinde tablo boyutu ve örneklem büyüklüğü arttıkça güç değerlerinde de artış görülmektedir; ancak korelasyon katsayısı gücü etkilememektedir.

$n=100$, $\rho=0.4$ ve $R=4$ olduğu durumda güç minimum; $n=1000$, $\rho=0.0$ ve $R=8$ olduğu durumda ise güç maksimumdur. En yüksek güç değerleri $n=1000$ için elde edilmiştir. Bilindiği gibi gücün 1'e yakın olması beklenir. Bu değer, seçenek hipotezi doğru iken yokluk hipotezinin reddedilme olasılığını verir.

Şekil 1-2-3 güç değerlerinin boyut, örneklem büyüklüğü ve korelasyon katsayısına göre değişimini göstermektedir. Tüm şekillerde güç, örneklem büyüklüğü ve boyuta göre artış göstermektedir. Ancak, küçük örneklerde boyut büyüdükçe bu artışın daha hızlı olduğu gözlenmektedir. $n<250$ iken tablo boyutunun artırılması güçte hızlı bir artış yaratmaktadır.

Tablo 3. YB Modeli için 500 Tekrarda Reddedilen ve Kabul Edilen Hipotezlerin Oranları

<i>n</i>	ρ	<i>R</i>	<i>red</i>	<i>kabul</i>	<i>n</i>	ρ	<i>R</i>	<i>red</i>	<i>kabul</i>
100	0.0	4	0.21	0.79	500	0.0	4	0.76	0.24
100	0.0	5	0.26	0.74	500	0.0	5	0.80	0.20
100	0.0	6	0.21	0.79	500	0.0	6	0.77	0.23
100	0.0	7	0.19	0.81	500	0.0	7	0.71	0.29
100	0.0	8	0.13	0.87	500	0.0	8	0.70	0.30
100	0.2	4	0.23	0.77	500	0.2	4	0.75	0.25
100	0.2	5	0.24	0.76	500	0.2	5	0.82	0.18
100	0.2	6	0.18	0.82	500	0.2	6	0.78	0.22
100	0.2	7	0.17	0.83	500	0.2	7	0.78	0.22
100	0.2	8	0.09	0.91	500	0.2	8	0.66	0.34
100	0.4	4	0.21	0.79	500	0.4	4	0.78	0.22
100	0.4	5	0.18	0.82	500	0.4	5	0.84	0.16
100	0.4	6	0.18	0.82	500	0.4	6	0.81	0.19
100	0.4	7	0.19	0.81	500	0.4	7	0.70	0.30
100	0.4	8	0.09	0.91	500	0.4	8	0.48	0.52
250	0.0	4	0.47	0.53	1000	0.0	4	0.97	0.03
250	0.0	5	0.50	0.50	1000	0.0	5	0.98	0.02
250	0.0	6	0.42	0.58	1000	0.0	6	0.99	0.01
250	0.0	7	0.37	0.63	1000	0.0	7	0.99	0.01
250	0.0	8	0.30	0.70	1000	0.0	8	0.98	0.02
250	0.2	4	0.49	0.51	1000	0.2	4	0.97	0.03
250	0.2	5	0.47	0.53	1000	0.2	5	0.98	0.02
250	0.2	6	0.42	0.58	1000	0.2	6	0.99	0.01
250	0.2	7	0.40	0.60	1000	0.2	7	0.99	0.01
250	0.2	8	0.29	0.71	1000	0.2	8	0.98	0.02
250	0.4	4	0.48	0.52	1000	0.4	4	0.98	0.02
250	0.4	5	0.49	0.51	1000	0.4	5	0.99	0.01
250	0.4	6	0.44	0.56	1000	0.4	6	0.99	0.01
250	0.4	7	0.36	0.64	1000	0.4	7	0.99	0.01
250	0.4	8	0.27	0.73	1000	0.4	8	0.98	0.02

Tablo 3, 500 tekrarda YB modeli için reddedilen ve kabul edilen hipotez oranlarını vermektedir. Sonuçlara bakıldığında reddedilen hipotezlerin oranında örneklem

büyüklüğü arttıkça büyüme görülmektedir. Bu sonuç Tablo 2'deki güç değerlerinin verdiği sonuçlar ile çakışmaktadır. Örneklem büyüklüğü arttıkça kabul edilen hipotezlerin oranında azalmaktadır. Dolayısıyla YB için güç değerleri ve Tablo 3'deki sonuçlar birbirini desteklemektedir.

Tablo 4. Sarkar Yöntemine Göre Hesaplanan $\hat{\eta}_\pi$ Parametrelerinin 500 Tekrar İçin Ortalama ve Standart Hata Değerleri

n	ρ	R	$\hat{\eta}_\pi$	$S\hat{\eta}_\pi$	n	ρ	R	$\hat{\eta}_\pi$	$S\hat{\eta}_\pi$
100	0.0	4	0.0621	0.0018	500	0.0	4	0.0488	0.00057
100	0.0	5	0.0714	0.0022	500	0.0	5	0.0696	0.00078
100	0.0	6	0.0725	0.0022	500	0.0	6	0.0745	0.00074
100	0.0	7	0.0739	0.0024	500	0.0	7	0.0755	0.00078
100	0.0	8	0.0676	0.0021	500	0.0	8	0.0731	0.00093
100	0.2	4	0.0654	0.0018	500	0.2	4	0.0482	0.00062
100	0.2	5	0.1399	0.021	500	0.2	5	0.0692	0.0009
100	0.2	6	0.0729	0.020	500	0.2	6	0.0730	0.00084
100	0.2	7	0.0722	0.0021	500	0.2	7	0.0757	0.00092
100	0.2	8	0.0680	0.0022	500	0.2	8	0.0738	0.00086
100	0.4	4	0.0634	0.0018	500	0.4	4	0.0497	0.00068
100	0.4	5	0.0690	0.0021	500	0.4	5	0.0713	0.0009
100	0.4	6	0.0700	0.0021	500	0.4	6	0.0730	0.00087
100	0.4	7	0.0233	0.0024	500	0.4	7	0.0754	0.00073
100	0.4	8	0.0676	0.0020	500	0.4	8	0.0740	0.00073
250	0.0	4	0.0555	0.0011	1000	0.0	4	0.0425	0.00042
250	0.0	5	0.0698	0.0013	1000	0.0	5	0.0671	0.00063
250	0.0	6	0.0737	0.0014	1000	0.0	6	0.0725	0.00053
250	0.0	7	0.0724	0.0012	1000	0.0	7	0.0761	0.00051
250	0.0	8	0.0698	0.0014	1000	0.0	8	0.0763	0.00054
250	0.2	4	0.0551	0.0011	1000	0.2	4	0.0438	0.00042
250	0.2	5	0.0696	0.0013	1000	0.2	5	0.0675	0.00057
250	0.2	6	0.0749	0.0013	1000	0.2	6	0.0713	0.00054
250	0.2	7	0.0727	0.0014	1000	0.2	7	0.0749	0.00068
250	0.2	8	0.0730	0.0014	1000	0.2	8	0.0713	0.00059
250	0.4	4	0.0555	0.0011	1000	0.4	4	0.0435	0.00037
250	0.4	5	0.0702	0.0014	1000	0.4	5	0.0669	0.00053
250	0.4	6	0.0742	0.0013	1000	0.4	6	0.0714	0.00069
250	0.4	7	0.0748	0.0013	1000	0.4	7	0.0763	0.00069
250	0.4	8	0.0726	0.0012	1000	0.4	8	0.0743	0.00062

Tablo 4'de ortalama $\hat{\eta}_\pi$ değerleri verilen çözümlerinin tümü için Z değerlerinin güç değerleri tüm örneklerde "1" olarak bulunmuştur. Tablo 4' de yer alan $\hat{\eta}_\pi$ değerlerinin istatistiksel olarak anlamlı olması güç değerlerindeki desteklemektedir. Tamamlanmamış tablo yapısı gösteren üçgensel olumsuzluk tablolarının yarı bağımsızlık modeli ile çözümlenmesi iterasyona bağlı çözümlenmedir. Sarkar'ın önerdiği yöntem uygulaması daha kolay olan bir yöntemdir. Yapılan çalışmada Yarı- bağımsızlık ve

Sarkar Yöntemi'nin paralel sonuçlar verdiği görülmüştür. Bu nedenle üçgensel olumsuzluk tablo çözümlerinde Sarkar Yöntemi'nin uygulamadaki kolaylığı nedeniyle YB çözümlenmesi yerine kullanabileceği önerilebilir.

KAYNAKLAR

- AGRESTI, A., (1990), *Categorical Data Analysis*, John Wiley&Sons, New-York.
- AKTAŞ, S., (1999), *Üçgensel Olumsuzluk Tablolarında Bağımsızlık Çözümlemesi*, Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi, Ankara.
- BISHOP, Y.M.M.,FIENBERG, S.E., (1969), *Incomplete two-dimensional contingency tables*, *Biometrics* 28, 177-202.
- GOODMAN, L. A., (1968), *The analysis of cross-classified data: Independence, quasi-independence and interactions in contingency tables with or without missing entries*, *JASA* 63, 1091-1131.
- GOODMAN,L.A.,(1981), *Association models and the bivariate normal for contingency tables*, *Biometrics*, 35,651-655.
- GOODMAN, L.A., (1985), *The Analysis of cross-classified data having and/or unordered categories: Association models, correlation models, and asymmetry models or contingency tables with or without missing entries*, *The Annals of Statistics*, Vol.13, No.1, 10-69.
- MANTEL, N., (1970), *Incomplete contingency tables*, *Biometrics* 26, 291-304
- PEARSON, E.S. and HARTLEY H.O., (1976), *Biometrika Tables for Statisticians Volume II*, Cambridge University Press.
- SARAÇBAŞI, T. ve MULUK, F.Z.(1985) *İki değişkenli normal dağılımdan oluşturulan olumsuzluk çizelgelerinde ilişki katsayılarının incelenmesi*, H. Ü. Fen ve Mühendislik Bilimleri Dergisi, 10, 9-33.
- SARKAR, S. K., (1989), *Quasi-independence in ordinal triangular contingency tables.*, *JASA* 84,592-597.

Power Analysis of Quasi-Independence Model for Triangular Contingency Tables

ABSTRACT

Triangular contingency tables are the special class of incomplete contingency tables. The quasi-independence model investigates usual independence model between row and column classifications. It is known that when the alternative hypothesis is true, chi-squared or

likelihood ratio statistics is distributed as noncentral chi-squared distribution. In this study, the powers of tests for triangular contingency tables were calculated by generating 30.000 random tables from the bivariate normal distribution and quasi-independence model was tested. The proportions of fitted and not fitted to this model were given. Power values were calculated for the likelihood ratios tests. The proportions of rejected hypothesis for the quasi-independence and the power values were compared.

Key Words: *Contingency tables, Quasi-independence model, Power Analysis*

Benzetimde Girdi Analizi Yapan Otomatik bir Sistem*

Halil TANIL**

İnci BATMAZ***

ÖZET

Benzetim yönteminin başarısı, sistemin gerçeği yansıtacak şekilde modellenmesine bağlıdır. Bu nedenle bu çalışmada, rasgele bir sistemi en iyi modelleyen teorik dağılımı otomatik olarak saptayan bir bilgisayar sistemi geliştirmek amaçlanmıştır. Geliştirilen GAYUS isimli bu sistem, bilgi-tabanında tek değişkenli teorik dağılımlara ilişkin bilgiler içermektedir. Sistem karar verme sürecinde çeşitli sezgiler, iyi-uyum testleri ve grafik görüntülerden yararlanmaktadır. Bunun yanı sıra, arayüzü istatistik ve bilgisayar kullanımı konusunda yeterli bilgi ve deneyimi olmayan kullanıcılara uygun bir şekilde yapılandırılmıştır. Bir "benzetim-destekli yazılım" olan GAYUS'un "Yapay Us yardımlı benzetim" gurubunda yer alan "bilen ön arayüz" türüne örnek olduğu söylenebilir. Ayrıca, geliştirilen sistemin başarımı Monte Carlo yöntemi ile türetilen rasgele örneklem üzerinde bir grup kullanıcı yardımı ile sınanmıştır. Bu kullanıcılar tarafından anlaşılır ve kolay kullanılabilir olarak değerlendirilen GAYUS'un, örneklem ölçümü büyüdükçe verinin türetildiği teorik dağılımı en iyi model olarak önerdiği gözlenmiştir. Sistem başarımının daha çok sayıda kıstas ve daha güçlü girdi analizi yöntemleri kullanılarak artırılabilirliği düşünülmektedir.

Anahtar Kelimeler: *Benzetim-destekli yazılım, Yapay us, Girdi analizi, İstatistiksel grafik yöntemler, Teorik dağılımlar*

1. GİRİŞ

Benzetim yöntemi, sistem analizi ve tasarımında yaygın ve güçlü bir problem çözme aracı olarak kullanılmaktadır (Law ve Kelton, 2000). Bu yöntem, matematiksel modelleme içerisinde yer alan dinamik ve sayısal bir modelleme yöntemi olup (Gordon, 1978), ardışık ve yinelemeli bir dizi adımdan oluşmaktadır (Maria, 1997; Shannon, 1998). Bu adımlar: sistemin tanımlanması ve modellenmesi; modelin onaylanması ve programlanması; bilgisayar programının doğrulanması ve sonuçların yorumlanmasıdır. Benzetim sürecinde rasgele bileşenler içeren bir sistem modellenirken, elde edilen verilere teorik bir dağılımın uyumu aranır. "Girdi analizi" olarak tanımlanan bu süreç, benzetim çalışmasının en önemli adımı olarak bilinmektedir (Selia, 1995). Çünkü, yöntemin başarısı, gerçek sistemin ne kadar iyi yansıtıldığına bağlıdır.

* Bu çalışma Ege Üniversitesi Araştırma Fonu'nun desteği ile gerçekleştirilmiştir.

** Ege Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 35100 Bornova, İzmir.

*** (Orta Doğu Teknik Üniversitesi, Fen ve Edebiyat Fakültesi, İstatistik Bölümü, 06531 Ankara).

Benzetim sürecinde kullanılacak modele ilişkin bilgisayar programının hazırlanmasını kolaylaştırmak amacı ile çeşitli yazılımlar geliştirilmiştir (Banks, 1998b). Benzetim çalışması yapan kişiyi ayrıntıdan kurtararak, çalışma süresini kısaltan bu yazılımlar işlevlerine göre altı grupta toplanmaktadır: genel amaçlı yazılımlar (GPSS, SIMAN, SIMSCRIPT, vb.), üretime yönelik yazılımlar (ProModel, ARENA, vb.), iş merkezli yazılımlar (ProcessModel, SimProcess, vb.), benzetim tabanlı çizelgelemeler (AutoSched, Factor, vb.), animatörler (Proof Animation, vb.) ve benzetim destekli yazılımlar. Bunların yanı sıra son yıllarda zamana bağımlı (veya zamana bağlı olmayan) ve çok değişkenli rasgele süreçleri modellemek amacı ile daha karmaşık yapıda olan “Poisson süreçleri” (Schmeiser, 1999), “Johnson ailesi”, “tek değişkenli Bezier dağılımları” ve “geliş-sayma süreçleri” (Nelson ve Yamnitsky, 1998) gibi girdi modellerini kullanan yazılımlar üzerinde de araştırmalar yapılmaktadır.

Yukarıda sözü edilen, benzetim çalışmalarında girdi modellemek amacı ile yaygın olarak kullanılan yazılımlardan biri ARENA/Input Analyzer'dır (Kelton ve diğ., 2002). Bu sistemde kullanıcı ya en uygun dağılımı kendisi belirler veya buna karar veremiyorsa verinin tüm dağılımlara uyumunun yapılmasını sistemden ister. Her iki durumda da sistem, parametreleri tahminleyerek, verinin teorik dağılım(lar)a uyumunu gösteren çeşitli değerler (hata kareler ortalaması, uyum-iyiliği testlerine ilişkin p-değerleri vb.) hesaplar. Bu değerleri yorumlayarak veriye en uygun dağılımı belirlemek kullanıcının sorumluluğundadır. Benzetim destekli yazılımlardan en çok bilinenler ise BestFit, ExpertFit (Law ve McComas, 1999) ve Stat::Fit'tir (Banks, 1998b). Bu yazılımlar tek değişkenli dağılım modellerinin (Binom, Poisson, Normal, Üstel, Gamma vb.) veriye uyumunu sağlamaya çalışırlar. ExpertFit bilgi tabanındaki dağılımlardan verinin modellenmesi için uygun bulduklarını göreceli puanlarla listelemektedir (Law ve Kelton, 2000). Benzer şekilde Stat::Fit yazılımı da teorik dağılımlar arasında birçok kritere dayalı göreceli karşılaştırmalar yaparak herbirinin kabul edilebilirliğinin ölçümünü vermektedir. Buna karşın BestFit yazılımının karar stratejisi sadece kullanıcının seçtiği uyum-iyiliği testinin sonuçlarına göre girdi verisine ilişkin hipotez edilen teorik dağılımların sıralanması düşüncesine dayanmaktadır.

Verilere en uygun rasgele modeli belirlemeye çalışan benzetim destekli yazılımlardan bazıları Yapay Us (YU) tekniklerinden yararlanılarak geliştirilmiştir (Yılmaz ve Sabuncuoğlu, 2000; Nielsen, 1991; Rothenberg, 1990). Bu yazılımlardan YU teknikleri kullanarak benzetilen davranış üretenler *YU-tabanlı benzetim*, YU tekniklerini yalnızca belirli adımlarda (girdi-çıkı analiz vb.) kullananlar ise *YU-yardımlı benzetim* olarak nitelendirilmektedir (Ören, 1994). YU-yardımlı benzetim sınıfına giren yazılımlar ayrıca, *bilen* (cognizant) *benzetim ortamları*, *iç içe bilgi-tabanlı sistemli bilen ortamlar* ve *yoğun bilen ortamlar* gibi çeşitlere sahiptirler.

İnsan usunun nasıl modelleneceği ile ilgilenen YU günümüzde robotlar, teorem ispatlama, doğal dil algılama, konuşma, görme, öğrenme, uzman sistemler (US) ler ve yapay sinir ağları gibi konularda uygulama bulmaktadır (Tanimoto, 1995). Bunlar içerisinden özellikle US geliştirmede YU teknikleri başarıyla kullanılabilirler (Hayes-Roth, 1983; Giarratano ve Riley, 1994). Herhangi bir uzmanlık alanına ilişkin problemleri insan uzman gibi çözebilmek amacıyla geliştirilen US'lar, genel olarak iki bileşenden oluşmaktadırlar (Elias, 1996): bilgi tabanı ve çıkarsama makinası. Bilgi-

tabanı, ilgili uzmanlık alanına ait bilgi, sezgi ve deneyimlerden meydana gelir. Bu sistemlerde bilgi gösterimi çeşitli şekillerde yapılabilmektedir. Çıkarsama ise bilgi-tabanı üzerinde yapılan bir arama işlemi olup, aramanın yönüne göre ileri-zincirleme veya geri-zincirleme şekilde uygulanabilir. Bu sistemlerde problemin çözüm süresini ve yolunu kısaltmak amacı ile sezgisel fonksiyonlar kullanılmaktadır (Rich, 1983).

Bu çalışmada benzetim sürecinde kullanıcı tarafından sağlanan rasgele bir sisteme ilişkin verilere en uygun teorik dağılım modelini otomatik olarak belirleyen bir bilgisayar sistemi geliştirmek amaçlanmıştır. GAYUS (Girdi Analizi Yapan Uzman Sistem) adı verilen bu sistemin başarımı Monte Carlo yöntemi kullanarak bilgi tabanında bulunan dağılım türlerinden çeşitli büyüklüklerde türetilen örneklemeler üzerinde bir grup kullanıcı yardımı ile ölçülmeye çalışılmıştır. Geliştirilen sistem, ExpertFit ve Stat::Fit yazılımları ile yapısal benzerlik göstermektedir. Ancak bu yazılımların geliştirilme süreçleri ve girdi analizinde kullandıkları tekniklere ilişkin bilgiler kaynaklarda yeterince yer almamaktadır. Çalışmamızla bu kara kutu sistemler açılarak, işleyişleri ve geliştirilmelerinde kullanılan yaklaşımlar ortaya konulmuştur. Ayrıca, GAYUS istatistik kullanımı konusunda yeterli bilgisi olmayan kullanıcılara uygun bir şekilde yapılandırılmıştır. Bu yönü ile sistemimiz ARENA/Input Analyzer'dan ayrılmaktadır.

Makalenin 2. Bölümde girdi analizi konusunda genel bilgilere yer verilmiştir. GAYUS'un genel yapısı, başarımlar değerlendirme sistemi, problem çözme stratejisi ve bir uygulama ile işleyişi 3. Bölümde tanıtılmıştır. 4. Bölümde sistemin başarımının sınanmasına ilişkin bulgular sunulmuştur. Son bölümde ise sonuçlar ve öneriler yer almaktadır.

2. RASGELE SİSTEMLERDE GİRDİ ANALİZİ

Gerçek sistemlerin çoğu rasgele süreçler içerirler. Girdi analizinde amaç, bu süreçleri gösteren rasgele değişkenlerin dağılımını belirlemektir (Vincent, 1998). Bu amaçla ilk önce verilerin ait olabileceği genel dağılım ailesi, verilerin elde edildikleri değişkenlerin zamana bağlı, tek değişkenli, kesikli ve durağan olup olmaması gibi çeşitli ölçütler göz önüne alınarak ayırt edilmeye çalışılır (Leemis, 1999). Bir sonraki adımda, seçilen bu dağılım aileleri, çeşitli sezgisel yöntemler kullanılarak daraltılır (Law ve Kelton, 2000). Bu sezgilerden bazıları verinin hangi araştırma alanından geldiğine, bazıları ise çeşitli nokta istatistikleri ve grafik yöntemlere dayandırılmaktadır. İstatistiksel yöntemlere göre bir veri kümesinin ait olabileceği genel dağılım ailesi, sürekli veriler için değişim katsayısı ($\delta = \sqrt{\text{Var}(X)}/E(X)$) ve kesikli veriler için varyansın ortalamaya oranı ($\tau = \text{Var}(X)/E(X)$) nokta istatistikleri kullanılarak sınıflandırılabilir. Burada X bir rasgele değişken, $\text{Var}(X)$ dağılımın varyansı ve $E(X)$ ise dağılımın ortalamasıdır.

Daha sonra, önceki adımda seçilen dağılımlara ilişkin parametreler tahminlenip, iyi-uyum testleri yapılır. Parametre tahminleme yöntemleri çeşitlidir (Freund, 1992). En yüksek olasılık ve momentler yöntemleri en yaygın kullanılan parametre tahminleme yöntemleridir. Burada iyi-uyum testlerinin amacı, verilerin hipotez edilen dağılım ailesi bireyinden gelip gelmediğinin istatistiksel olarak belirlenmesidir. Bu amaçla en çok Ki-kare, Kolmogorov-Smirnov (Daniel, 1978) ve Anderson-Darling (D'Agostin ve

Stephens, 1986) testlerinden yararlanılmaktadır. Burada Ki-kare iyi-uyum testi histogram aralıklarına; Kolmogorov-Simirnov ve Anderson-Darling testleri ise gözlemsel dağılım fonksiyonuna dayandırılmaktadır. Ki-kare iyi-uyum testinde örnekleme ilişkin gözlemsel sıklıklar ile hipotez edilen teorik dağılımın beklenen sıklıkları karşılaştırılmaktadır. Ki-kare testinde kullanılan aralık sayısı değişikçe çelişkili sonuçlar ortaya çıkabilmektedir. Diğer yandan Kolmogorov-Simirnov testi gözlemsel ve teorik dağılım fonksiyon değerleri arasındaki farklardan en büyüğünün mutlak değerinin önemini istatistiksel olarak değerlendirilmektedir. Basit bir formülle hesaplanan Anderson-Darling test istatistiği ise hipotez edilen teorik ve örnekleme ilişkin gözlemsel değerler arasındaki farkların karelerinin ağırlıklı ortalaması olarak tanımlanmıştır. Kolmogorov-Simirnov ve Anderson-Darling testlerinde verilerin sınıflandırılmasından dolayı bilgi kaybı söz konusu olmadığı için Ki-kare testinden daha güçlü olduğu söylenmektedir (Vincent, 1998).

Verinin bir teorik dağılım modeline uygunluğu, çeşitli grafik yöntemlerle de sınıranabilmektedir (Chambers ve diğ., 1983). Histogram, çizgi, bölen (quantile) ve fark grafikleri girdiyi analiz etmek amacıyla kullanılan grafik yöntemlerden bazılarıdır. Histogramlar sürekli verilerin dağılım şeklini görmek için yaygın olarak kullanılan bir grafik gösterim şeklidir. Histogramları, teorik olasılık (yoğunluk) fonksiyon grafikleriyle karşılaştırarak veriler için en uygun olan teorik dağılım modellerini belirlemek olasıdır. Diğer yandan, teorik bölen grafiklerinin yatay ekseninde sıralı gözlemler, dikey ekseninde ise sıralı gözlemlere karşılık gelen teorik dağılıma ilişkin bölen değerleri yer alır. Bu grafiklerde noktalar eğer $x = y$ doğrusu etrafında yer alıyorsa söz konusu teorik dağılımın verilerimizi modellemek için uygun olduğu düşünülebilir. Fark grafikleri ise, bir teorik dağılım modelinin veriden sapmasını gösterir. Eğer hipotez edilen teorik dağılım veriler için uygun bir model ise, fark grafiğinin $y = 0$ doğrusu etrafındaki değişiminin küçük olması beklenir.

3. SİSTEMİN TANITILMASI

3.1. Özellikleri ve İşleyişi

Bu çalışmada, bir benzetim sürecinde, rasgele bir sistemden kullanıcı tarafından sağlanan verilere en uygun teorik dağılım modelini otomatik olarak belirleyen bir yazılım geliştirilmeye çalışılmıştır. Bu yazılım, girdi analizi konusunda yeterli bilgi sahibi olmayan kişilere uygun ve kolay kullanılabilir bir yapıda tasarlanmıştır. Bu özellikleri ile GAYUS, Banks'ın (1998b) sınıflandırmasındaki "benzetim-destekli yazılımlar" grubuna dahil edilebilir. Bunun yanı sıra yapısı itibari ile de "bilen ön arayüz" türüne örnektir.

Genel olarak GAYUS şu şekilde işlemektedir. Sistem problem çözme sürecinde 2. Bölüm'de özetlenen girdi analizi yöntemlerini belirli bir strateji altında bazen kullanıcı bilgisine başvurarak, bazen de kendi bilgisini değerlendirerek uygulamaktadır. Bir yandan da, elde ettiği sonuçları puana dayalı bir sistem çerçevesinde değerlendirerek bilgi-tabanında yer alan teorik dağılım modellerinin girdi verilerine uygunluğunun derecesini gösteren değerlere dönüştürmektedir. Sonuçta, olası dağılım modelleri sistem tarafından ayrıntılı bir puan tablosu ile listelenmekte ve en büyük puana sahip olan dağılım (puanı 0,4'ten büyük olması koşulu ile); verilere en uygun

teorik dağılım modeli olarak önerilmektedir. Ayrıca GAYUS, karar sürecini ayrıntılı bir şekilde açıklamaktadır.

Bu bölümde ilk önce GAYUS'un birimsel yapısı tanıtılacaktır. Daha sonra, kullanılan başarımlar değerlendirme sistemi ve problem çözme stratejisi ayrıntılı olarak anlatılacaktır. Son olarak da sistemin işleyişi bilgisayarda türetilen rasgele bir örneklem üzerinde açıklanacaktır.

3.2. Birimsel Yapısı

GAYUS dört temel birimden oluşmaktadır (Şekil 1). Bunlar:

1. İstatistik ve Grafik Birimi: Bu birim, kullanıcı verileri girdikten sonra sistem tarafından hipotez edilen dağılımlara ilişkin parametreleri en yüksek olasılıkla tahminleyicileri ile tahmin eder. Sonra, Ki-kare ve Kolmogorov-Smirnov iyi-uyum testlerini uygulayarak teorik dağılımın veriye uyumunu istatistiksel olarak sınar. Ayrıca histogram, bölen ve fark grafikleri ile uyumu değerlendirir. Bu işlemler Excel'deki formüller, Visual Basic'te (Gray ve Strain, 1995) yazılan makrolar ve Pascal'da (Konvalinen ve Wileman, 1987) kodlanan programlar ile otomatik olarak yapılmaktadır.

Şekil 1. GAYUS'u oluşturan temel birimler ve ilişkileri.

2. Bilgi-tabanı ve Çıkarılma Birimi: Esta for Windows (PDC, 1993) yazılımı kullanılarak geliştirilen bilgi-tabanı, üretim kuralları ile kodlanmış tek değişkenli dağılımlara ilişkin bilgilerden oluşmaktadır. Bu dağılımlardan sürekli olanlar: Beta, Gamma, Üstel, Normal, LogNormal, Uniform ve Poisson (Johnson ve diğ., 1994, 1995); kesikli olanlar ise: Kesikli Uniform, Geometrik, Binom ve Negatif Binomdur (Johnson ve diğ., 1992). Bu birim önce, kurallar içerisinde yer alan ve *İstatistik ve Grafik birimi* tarafından belirlenen parametrelerin değerini ya Çevirici birimden okur yada kendisi kullanıcıdan edinir. Sonra, bilgi-tabanı dosyası içerisindeki bu kuralları ileri-zincirleme yöntemine göre atarlar.

3. Çevirici Birim: Bu birim, istatistik ve grafik birimi tarafından üretilen bilgilerin bilgi-tabanı ve çıkarılma birimince kullanılabilir bir biçimde ifade edilmesini sağlar.

4. Arayüz: Kullanıcı ile sistem arasında bilgi alışverişini sağlayan birimdir. Bu birim veri girişi ve istatistiksel hesaplamaların yapıldığı, grafik ve bulguların görüntülediği pencereler ve kullanıcı ile bilgi alışverişinin yapıldığı diyalog kutularından oluşan grafik bir arayüz olarak tasarlanmıştır.

3.3. Başarımlar Değerlendirme Sistemi

GAYUS, bilgi-tabanında yer alan teorik dağılım modelleri içinden girdi verilerine uyabilecek adayları bir sonraki bölümde ayrıntılı bir şekilde anlatılacak olan strateji ile belirler. Bu süreçte bir yandan da aday olan dağılımların veriyi temsil etme gücünü puan olarak ifade eder. GAYUS'un başarımlar değerlendirme sisteminde aday olan teorik dağılımlara ilişkin puan hesaplanırken, bileşenleri çeşitli parametrelerin

belirli ağırlıklarda katılımlarından oluşan toplamsal bir fonksiyon kullanılmaktadır. Puan fonksiyonunun genel şekli aşağıda verilmektedir:

$$puan_j = \sum_i^7 [ağırlık_i] [parametre_i]; 0,0 \leq puan_j \leq 1,0; 1 \leq j \leq 11 \text{ ve } \sum_i ağırlık_i = 1,0$$

(1)

Burada j, aday olabilecek dağılımları ve i'de ölçütleri göstermektedir. Puan fonksiyonunda yer alan parametreler, dağılımın çeşitli girdi analizi yöntemlerindeki başarımlarının sayısal değerleri veya bu değerlerin bir fonksiyonudur. Puan fonksiyonunda adı geçen parametrelerin anlamları kısaca şöyledir:

- Parametre 1:* Nokta istatistikleri ile belirlenen dağılım ailesine ilişkin parametre.
- Parametre 2, 3, 4,:* İstatistiksel grafiklere ilişkin parametreler.
- Parametre 5,6:* İyi-uyum testlerine ilişkin parametreler.
- Parametre 7:* Teorik dağılım modelinin veriden sapmasını gösteren hata kareler toplamlarına ilişkin parametre.

Başarım değerlendirme sistemine göre puanın sıfır olması verilere uygun olmayan bir dağılım modelini, bir olması ise verilere tam uygunluk gösteren bir teorik dağılım modelini ifade etmektedir. Eğer puan 0,40'ın altında ise "kötü uyum", 0,40 ile 0,50 arasında ise "orta uyum", 0,50 ile 0,75 arasında ise "iyi uyum" ve 0,75 ile 1,0 arasında ise "mükemmel uyum" yorumu yapılmaktadır. Yukarıda tanımlanan puan fonksiyonunda yer alan ağırlıklar ve puanın değerine göre modelin veriye uyumunun kalitesinin yorumlanması konunun uzmanı olan kişilerin görüşleri doğrultusunda sezgisel olarak belirlenmiştir.

3.4. Problem Çözme Stratejisi

GAYUS'un problem çözme stratejisi aşağıdaki adımlardan oluşmaktadır:

1. Kullanıcı verileri girdikten sonra veri tipini belirler (Sürekli, Kesikli):
2. Veriye ilişkin aşağıdaki işlemleri yapar:
 - Girdi verilerini sınıflandırır,
 - Sürekli veriler için δ veya kesikli veriler için τ değerini hesaplar,
 - Olası her bir teorik dağılım modeline ilişkin parametreleri tahminler,
 - Ki-kare ve Kolmogorov-Simirnov iyi-uyum test istatistiklerini hesaplar,
 - Hata kareler ortalamalarını (HKO) hesaplar,
 - HKO' sı en küçük olan üç dağılım için üst üste ilgili grafikleri hazırlar.
3. Çevirici dosyaları yaratır.
4. Bilgi-tabanı ve çıkarsama birimine geçer.
5. Olası teorik dağılımlara ait puanları sıfırlar ($puan_j = 0,0$)
6. İyi-uyum testlerini değerlendirmek için gerekli α hata payımı kullanıcıdan edinir.
7. Girdi verilerinin ait olabileceği dağılım ailesini nokta istatistiklerini kullanarak aşağıdaki şekilde belirler:

- 7.1. Sürekli veriler için δ değerini çevirici dosyalardan okur ve aşağıdaki kuralları atesler (Law ve Kelton, 2000, sf. 333):

- Eğer ($\delta < 0,8$ veya $\delta > 1,2$) ise dağılım ailesinin üyeleri Beta, Gamma, Normal, Lognormal veya Uniform tipinde olabilir.
 - Eğer ($\delta < 0,0$) ise dağılım ailesinin üyeleri Normal veya Uniform tipinde olabilir.
 - Eğer ($\delta \geq 0,8$ ve $\delta \leq 1,2$) ise dağılım ailesinin üyesi Üstel tipinde olabilir.
 - Eğer [$0,8 \leq (\text{Gamma dağılımının } \alpha \text{ parametresi}) \leq 1,2$ ve $0,8 \leq \delta \leq 1,2$] ise dağılım ailesinin üyeleri Üstel ve Gamma tipte olabilir.
- 7.2. Kesikli veriler için τ değerlerini çevirici dosyalardan okur ve aşağıdaki kuralları ateşler (Law ve Kelton, 2000, sf.334):
- Eğer ($\tau < 0,8$ veya $\tau > 1,2$) ise dağılım ailesinin üyeleri Binom, Negatif Binom, Geometrik, Kesikli Üniform tipinde olabilir.
 - Eğer ($\tau \geq 0,8$ ve $\tau \leq 1,2$) ise dağılım ailesinin üyesi Poisson tipinde olabilir.
 - Eğer ($\tau < 0,0$ veya $\tau > 1,0$) ise dağılım ailesinin üyeleri Poisson, Negatif Binom, Geometrik, Kesikli Üniform tipte olabilir.
 - Eğer ($\tau < 1,0$) ise dağılım ailesinin üyeleri Poisson, Binom, Kesikli Üniform tipte olabilir.
- 7.3. Parametre 1'in değerini belirler ve 7.1 veya 7.2 adımlarında saptanan olası dağılım ailesinin her bir üyesinin puanını güncelleştirir*.
8. Verilere ilişkin grafikleri çevirici dosyalardan alır ve kullanıcıya:
- 8.1. Üst üste çizilmiş histogram grafiklerini gösterip, histograma en uygun olan teorik dağılımı sorar; Parametre 2'nin değerini seçilen dağılışı için 1 ve diğerleri için 0 olarak belirler ve dağılımların puanlarını güncelleştirir.
- 8.2. Fark grafiklerini gösterip, $y=0$ doğrusu etrafında en az değişim gösteren teorik dağılımı sorar; Parametre 3'nin değerini seçilen dağılışı için 1 ve diğerleri için 0 olarak belirler ve dağılımların puanlarını güncelleştirir.
- 8.3. Teorik bölen grafiklerini gösterip, $x=y$ doğrusuna en iyi uyan teorik dağılımı sorar; Parametre 4'nin değerini seçilen dağılışı için 1 ve diğerleri için 0 olarak belirler ve dağılımların puanlarını güncelleştirir.
9. Dağılım ailesinde yer alan ve 8. adımda belirlenen teorik dağılım modellerinin 6. adımda belirlenen α hata payına ilişkin Ki-kare test istatistiği değerini çevirici dosyalardan okur; Parametre 5'in değerini belirler ve söz konusu dağılımların puanlarını güncelleştirir.
10. Dağılım ailesinde yer alan ve 8. adımda belirlenen teorik dağılım modellerinin 6. adımda belirlenen α hata payına ilişkin Kolmogorov-Simirnov test istatistiği değerini çevirici dosyalardan okur; Parametre 6'nın değerini belirler ve söz konusu teorik dağılımların puanlarını güncelleştirir.
11. Dağılım ailesinde yer alan ve kullanıcının grafiklere bakarak seçtiği teorik dağılım modellerine ilişkin fark değerlerini çevirici birimden okur; Parametre 7'nin değerini belirler ve söz konusu teorik dağılımların puanlarını güncelleştirir.

* Puan güncelleştirme: $\text{puan}_j = \text{puan}_j + [\text{ağırlık}_i][\text{parametre}_i]$ ($1 \leq j \leq 11$ ve $1 \leq i \leq 7$) şeklinde yapılmaktadır.

12. Sonuçları aşağıdaki şekilde görüntüler:

- 12.1 Dağılım ailesinde yer alan ve 8. adımda belirlenen tüm teorik dağılım modellerine ilişkin puanları listeler.
- 12.2 Eğer 0,4'ten yüksek puan yoksa herhangi bir teorik dağılım modeli önermez ve 12.5' inci adıma gider.
- 12.3 Ayrıntılı puan tablosunu görüntüler.
- 12.4 En büyük puana sahip dağılımı, verilere uygun bir teorik model olarak önerir ve nedenlerini kullanıcıya açıklar.
- 12.5 İyi-uyum testlerinin sonuçlarını görüntüler ve durur.

3. 5. Sistemin Bir Örnekle İşletilmesi

Bu bölümde GAYUS'un işleyişi $\lambda = 0,5$ parametrelili Üstel dağılımıtan $n = 60$ genişliğinde türetilen rasgele bir örneklem ile gösterilecektir. Adım adım analiz süreci ve bu süreçte görüntülenen ekranların kopyaları aşağıda yer almaktadır.

GAYUS' u çalıştırdığımızda karşımıza gelen açılış ekranı tıklanarak geçildikten sonra, verileri girebileceğimiz ortamla karşılaşırız (Şekil 2). Veriler A2 hücresinden başlayarak alt alta girilip, 'Veri Girişini Tamamladım' isimli bağlantı tıklandıktan sonra, verilerin sınıflandırılması görüntülenmektedir (Şekil 3). Burada, bazı sınıfların sıklık sayısının beşten küçük olması durumunda, bu sınıfların birleştirilmesi gerektiğini söyleyen bir diyalog kutusu yayınlanır. Böyle bir durumda, birleştirilmek istenen sınıfların sıklık sayılarının bulunduğu satırlara karşı gelen 'Ki-kare İçin Birleştirme' isimli sütunda yer alan hücrelere aynı rakamlar (1, 2, 3, ...) girilerek 'Birleştir' isimli bağlantı tıklanır. Aksi halde, 'İşleme Devam Etmek İçin' isimli bağlantı tıklanır. İstatistik hesaplamaların yapılması ve grafiklerin hazırlanması amacıyla gerekli bağlantı onaylandıktan sonra bilgi-tabanı ve çıkarsama birimine geçilmiş olunur.

Bu birimde 'Consult' menüsü altındaki 'Begin Consultation' menü elemanı tıkladığında kullanıcıdan bilgi edinme sürecine geçilir. Bu süreçte önce iyi-uyum testlerinde kullanılacak α yanılma payı belirlenir (Şekil 4). Ardından kullanıcıya hangi dağılıma ilişkin olasılık (yoğunluk) fonksiyon grafiğinin verilerin histogramına daha çok benzediği sorulur (Şekil 5). Daha sonra üst üste çizilmiş teorik bölen grafiklerinden hangisinin $x=y$ doğrusu etrafında yoğunlaştığı öğrenilir (Şekil 6). Son olarak ta hangi dağılıma ilişkin fark grafiğinin $y=0$ eksenine etrafında en az değişkenlik gösterdiğinin belirlenmesi istenir (Şekil 7). Bu sorgularda yer alan seçeneklerde dağılım tipleri A, B, C, vb. sembollerle gösterilerek, kullanıcının önyargılı karar vermesi önlenmeye çalışılmıştır. Grafiklerin yer aldığı pencerelerde bulunan 'Zoom' düğmesi grafik görüntüsünün büyütülmesini sağlamaktadır.

Bu aşamadan sonra girdi analizi sonuçları yayınlanır (Tablo 1). Bu sonuçlara dayanarak sistem, verilerimize en uygun teorik dağılım modeli olarak 0,9638 puanla ilk sırada yer alan $\lambda = 0,594$ parametrelili Üstel dağılımı önermektedir. Daha sonra görüntülenen tablolarda puanların nasıl hesaplandığı (Tablo 2), bu karara ilişkin gerekçeleri (Tablo 3) ve iyi-uyum test sonuçları (Tablo 4) ayrıntılı olarak gösterilmektedir.

Şekil 2. Veri giriş ortamı.

Şekil 3. Sınıflandırılmış verilerin yer aldığı pencere.

Şekil 4. Kullanıcıya yanılma payının sorulduğu pencere.

Şekil 5. Veriye ilişkin histogram-yoğunluk grafiği.

Şekil 6. Teorik bölen grafikleri.

Şekil 7. Fark grafikleri.

Tablo 1. Girdi analizi sonuçları.

Tablo 2. Ayrıntılı puan tablosu.

Tablo 3. Kararın gerekçelendirilmesi.

Tablo 4. İyi-uyum test sonuçları.

4. SİSTEMİN BAŞARIMININ SINANMASI

Çalışmanın son aşamasında, geliştirilen sistem sınınanarak başarımı ölçülmeye çalışılmıştır. Bu amaçla, önce Minitab for Windows (Ryan ve diğ., 1985) yazılımının rasgele sayı türeteçleri kullanılarak, bilgi tabanında bulunan dağılım türlerinden (farklı parametrelere sahip 14 dağılım), çeşitli büyüklüklerde ($n=10, 60, 100, 500$) ve her büyüklük için 30'ar (toplam 1680 tane) örneklem türetilmiştir. Daha sonra GAYUS, türetilen herbir örneklem için değişik kullanıcılar tarafından çalıştırılıp, aday olarak belirlenen olası dağılımlara ilişkin parametre tahminleri ve hesaplanan puanlar kaydedilmiştir. Bunun yanı sıra, sistemin arayüzü hakkında, kullanıcıların görüşleri alınmıştır.

Elde edilen sınav sonuçları değerlendirilmeden önce, verilerin türetildiği teorik dağılımlara ilişkin sistem tarafından hesaplanan puanların sırasına ve örneklem ölçümlerine bağlı olarak (örneğin, $n=10$ iken verilerin türetildiği teorik dağılımın en yüksek puanı alması gibi) sınıflandırılmıştır. Daha sonra her sınıfa ilişkin ortalama puan ve bu puanı alması olasılığı hesaplanmıştır (Tablo 5). Burada olasılık, analiz edilen örneklemin geldiği (türetildiği) dağılımın uygun dağılım modeli olarak i . ($i=1, \dots, 6$) en yüksek puanı alması olasılığını; ortalama puan ise, i . en yüksek puanı alan bu modelin puan ortalamasını göstermektedir.

Tablo 5. GAYUS'a ilişkin başarımların sınav sonuçları.

Başarımların sınav sonuçlarının yer aldığı Tablo 5. incelendiğinde, geliştirilen sistemin beklentilerimizi aşağıda ifade edildiği gibi sağladığı görülmektedir:

1. Örneklem ölçümü büyüdükçe, verilerin türetildiği teorik dağılımın en yüksek puanı alması olasılığı artmaktadır. Örneğin, örneklem ölçümü $n=10$ olduğunda, verilerin türetildiği teorik dağılımın en yüksek puan alması olasılığı 0.4429' dur. Örneklem ölçümü $n=60$ olduğunda ise olasılık değerinin 0.7881' e çıktığı görülmektedir.
2. Örneklem ölçümü büyüdükçe, verilerin türetildiği teorik dağılımın en yüksek puanı almaması olasılığı azalmaktadır. Örneğin, en yüksek 2. puanı alması durumunda $n=10$ için olasılık değeri 0.2881 iken, örneklem ölçümü 60' a çıktığında bu değer 0.1881' e düştüğü görülmektedir.

3. Olası tüm teorik dağılımlara ilişkin puanlar arasındaki büyükten küçüğe doğru sıralamada, verilerin türetildiği teorik dağılıma ait puanın daha yukarılarda olması olasılığı, daha aşağılarda olma olasılıklarından daha büyük olmaktadır.

5. SONUÇ VE ÖNERİLER

Bu çalışmada benzetim yönteminin önemli bir adımı olan, girdi verilerine en uygun dağılım modelinin belirlenmesi amacı ile Microsoft Excel, Esta for Windows uygulamaları, Visual Basic ve Pascal programlama dilleri kullanılarak otomatik bir bilgisayar yazılımı geliştirilmiştir. Geliştirilen sistem, bilgi-tabanında tek değişkenli ve zamandan bağımsız on bir kesikli ve sürekli dağılım modellerini içermektedir. Yapılanmasında kullanılan tekniklerle GAYUS'un *YU-Yardımlı Benzetim* gurubunda yer alan *Bilen Ön Arayüz* türüne örnek olduğu söylenebilir. Arayüzü kullanıcılar tarafından anlaşılır ve kolay kullanılır olduğunu ifade edilmiştir. Sistem zaman zaman karar verme sürecinde kullanıcı desteğini beklemektedir. Ancak, kullanıcının, hesaplanan puanlara katkısı çok az olup, hiçbir uzmanlık bilgisi gerektirmemektedir.

GAYUS piyasada aynı amaçla yaygın olarak kullanılan ve özellikleri giriş bölümünde açıklanan ARENA/Input Analyzer'dan temelde iki farklılık göstermektedir. GAYUS: 1. konuya ilişkin uzmanlık bilgisine sahip olmayan kullanıcı tarafından da güvenle kullanılabilir, 2. kullandığı sezgilerle çözüm uzayını daralttığından dolayı daha hızlı sonuç vermektedir. Buna karşın GAYUS benzetim destekli ExpertFit, BestFit ve StatFit yazılımları ile yapısal benzerlikler göstermektedir. Ancak, GAYUS karar sürecinde BestFit'e göre daha çok başarımlı ölçütünü değerlendirmektedir. Geliştirilen sistemin etkinliğinin mevcut diğer sistemlerle karşılaştırılması ilerki bir araştırma konusu olarak düşünülmektedir.

GAYUS'un geliştirilmesinde çeşitli sorunlarla karşılaşmıştır. Bunlardan en önemlisi değişik uygulama yazılımlarının bütünleştirilmesidir. Bu sorun genelde tüm büyük yazılımların oluşturulmasında ortaya çıkmaktadır. Bazı hazır yazılımlar, ancak belirli yazılımlar ve yüksek düzeyli diller ile bağlantı kurabilmektedir. Dolayısı ile sistemin yetenekleri, bütünleştirilen uygulamaların yetenekleri ile sınırlı kalmaktadır.

Sistemin başarımının sınanmasına ilişkin sonuçlar beklentilerimize uygundur. Yani, güçlü parametre tahmin yöntemleri (En yüksek olabilirlik tahminleme yöntemi gibi) kullanıldığında ve büyük örneklemelerin analizinde sistem, verinin geldiği dağılımın tipik özelliklerini kolayca yakalayarak, bu dağılımı girdi verileri için en uygun teorik dağılım olarak birinci sırada önerebilmektedir.

Geliştirilen sistemin başarımının yapılacak yeni çalışmalarla arttırılabileceği düşünülmektedir. Bunlardan bazıları aşağıda verilmiştir:

- Girdi verilerinin hangi dağılımdan gelmiş olabileceğini araştırırken kullanılan ölçütler çeşitlendirilebilir (Öztürk ve Dudewicz, 1992). Bunun yanı sıra, uygun dağılım seçimi yapma aşamasında, kullanacağımız ölçütlerin hangi ağırlıklarla puan fonksiyonunu etkileyeceği konusu araştırılmalıdır. Bu amaçla protokol analizleri ile istatistik uzmanlarından bilgi edinilebilir.

- Ki-kare iyi-uyum testi, bazı durumlarda, girdi verilerinin sınıf sayısına bağı olarak farklı sonuçlar üretebilmektedir. En uygun sınıf sayısının belirlenmesi konusunda yapılacak çalışmalar (Hoaglin ve diğ., 1983), iyi-uyum testlerinin gücünün ve güvenilirliğinin artmasına yardımcı olacaktır.
- Sistemde uyumun iyiliğini değerlendiren Ki-kare ve Kolmogorov-Smirnov testleri genel amaçlı testlerdir. Bunların yanı sıra, belirli tip dağılımlara uyumu değerlendiren özel testler de bulunmaktadır. Örneğin, Shapiro-Wilk testi (Shapiro ve Wilk, 1965) bir örneklemin Normal dağılımdan gelip gelmediğini değerlendiren güçlü bir iyi-uyum testidir. Genel amaçlı testler yerine daha güçlü özel testlerin kullanılması daha doğru sonuçların üretilmesini sağlayabilir.
- Küçük örneklemler için geliştirilmiş özel yöntemler uygun dağılımın belirlenmesinde daha başarılı olacaktır.

Bir benzetim çalışmasına destek olarak geliştirilen bu yazılım, gerçekte otomatik olarak istatistiksel modelleme yapan bir sistemdir. Bu nedenle GAYUS'un yalnızca benzetim çalışmalarında değil, tüm rasgele süreçler içeren sistemlerin analizinde kullanılması önerilebilir.

KAYNAKLAR

- BANKS, J. (1998a), *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, New York: John Wiley and Sons.
- BANKS, J. (1998b), *Software for Simulation*, J. Banks (der.), *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice* içinde, New York: John Wiley and Sons, 813-835.
- CHAMBERS, J., CLEVELAND, W., KLEINER, B. ve TUKEY, P. (1983), *Graphical Methods for Data Analysis*, California: Wadsworth Int. Group.
- D'AGOSTINO, R.B. ve STEPHENS, M.A. (1986), *Goodness-of-Fit Techniques*, New York: Marcel Dekker.
- DANIEL, W. (1978), *Applied Nonparametric Statistics*, Boston: Houghton Mifflin Com.
- ELIAS, M. (1996), *Building Expert Systems: Principles, Procedures and Applications*, Minneapolis: Wadsworth.
- FREUND, J.E. (1992), *Mathematical Statistics*, New Jersey: Prentice Hall.
- GIARRATANO, J. ve RILEY, G. (1994), *Expert Systems: Principles and Programming*, Boston: PWS Pub.
- GORDON, G. (1978), *System Simulation*, New Jersey: Prentice Hall Inc.
- GRAY, C. ve STRAIN, T. (1995), *Visual Basic 4 Nuts and Bolts: for Experienced Programmers*, California: McGraw Hill.

- HAYES-ROTH, F., WATERMAN, A.D. ve LENAT, B.D. (1983), *Building Expert Systems*, New-York: Addison-Wesley.
- HOAGLIN, D.C., MOSTELLER, F. ve TUKEY, J.W. (1983), *Understanding Robust and Exploratory Data Analysis*, Canada: John Wiley and Sons.
- JOHNSON, N.L., KOTZ, S. ve KEMP, A.W. (1994), *Continuous Univariate Distributions*, Vol 1, Boston: Houghton Mifflin.
- JOHNSON, N.L., KOTZ, S. ve KEMP, A.W. (1995), *Continuous Univariate Distributions*, Vol 2, Boston: Houghton Mifflin.
- JOHNSON, N.L., KOTZ, S. ve KEMP, A.W. (1992), *Univariate Discrete Distributions*, Boston: Houghton Mifflin.
- KELTON, W.D., SADOWSKI, R.P. ve SADOWSKI, D.A. (2002), *Simulation with ARENA*, Boston: McGraw-Hill.
- KONVALINEN, J. ve WILEMAN, S. (1987), *Programming with Pascal*, Singapore: McGraw-Hill.
- LAW, A.M. ve KELTON, W.D. (2000), *Simulation Modeling and Analysis*, New Jersey: McGraw-Hill.
- LAW, A.M. ve McCOMAS, M.G. (1999), *ExpertFit: Total Support for Simulation Input Modeling*, A.P. Farrington, H.B. Newbhard, D.T. Sturrock ve G.W. Evans (der.), Proceedings of Winter Simulation Conference içinde, 261-266.
- LEEMIS, L. (1999), *Simulation Input Modelling*, A.P. Farrington, H.B. Newbhard, D.T. Sturrock ve G.W. Evans (der.), Proceedings of Winter Simulation Conference içinde, 14-23.
- MARIA, A. (1997), *Introduction to Modeling and Simulation*, S. Andradottir, K.J. Healy, D.H. Withers ve B.L. Nelson (der.), Proceedings of Winter Simulation Conference içinde, 7-13.
- NELSON, B.L. ve YAMNITSKY, M. (1998), *Input Modelling Tools for Complex Problems*, A.P. Farrington, H.B. Newbhard, D.T. Sturrock ve G.W. Evans (der.), Proceedings of Winter Simulation Conference içinde, 105-111.
- NILSEN, N.R. (1991), *Application of AI Techniques to Simulation*, P.A. Fishwick, P.A. ve R.B. Modjeski (der.), Knowledge-Based Simulation Methodology and Application içinde, New York: Springer-Verlag, 1-19.
- ÖREN, T. (1994), *Artificial Intelligence in Simulation*, Annals of Operations Research, 63, 287-319.
- ÖZTÜRK, A. ve DUDEWICZ, E.J. (1992), *A New Statistical Goodness-of-fit Test Based on Graphical Representation*, Biometrika, 34, 403-427.
- PROLOG DEVELOPMENT CENTER (1993), *User's Book for Esta for Window*, Denmark: PDC.

- RICH, E. (1983), *Artificial Intelligence*, New Jersey: MacGraw-Hill.
- ROTHENBERG, J. (1990), *Tutorial: Artifice Intelligence and Simulation*, O. Balci, R.P. Sadowski ve R.E. Nance (der.), Proceedings of Winter Simulation Conference içinde, 22-24.
- RYAN, B.F., JOINER, B.L. ve RYAN, Jr. T.A. (1985), *Minitab Handbook*, Boston: PWS-Kent.
- SCHMEISER, B. (1999) *Advanced Input Modeling for Simulation Experimentation*, A.P. Farrington, H.B. Newbhard, D.T. Sturrock ve G.W. Evans (der.), Proceedings of Winter Simulation Conference içinde, 110-115.
- SELIA, A. (1995), *Introduction to Simulation*, C. Alexopoulos, K. Kang, W.R. Lilegdon ve D. Goldsman (der.), Proceedings of Winter Simulation Conference içinde, 7-14.
- SHANNON, R.E. (1998), *Introduction to Art and Science of Simulation*, D.J. Medeiros, E.F. Watson, J.S. Carson, M.S. Manivannan (der.), Proceedings of Winter Simulation Conference içinde, 7-14.
- SHAPIRO, S.S. ve WILK, M.B. (1965), *An Analysis of Variance Test for Normality*, *Biometrika*, 52, 591-611.
- TANIMOTO, S.L. (1995), *The Elements of AI Using Common Lisp*, Oxford: CS Press.
- VINCENT, S. (1998), *Input Data Analysis*, J. Banks (der.), *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice* içinde, New York: John Wiley and Sons, 55-91.
- YILMAZ, A. ve SABUNCUOĞLU, İ. (2000), *Input Data Analysis Using Neural Networks*, *Simulation*, 74(3), 128-137.

An Automatic Input Analyzer for Simulation

ABSTRACT

It is well known that the success of a simulation study depends on the success of the model of the real system. Based on this fact, in this study, it is aimed to develop a computer system for automatically diagnosing the theoretical distribution which models the data obtained from a stochastic system best. The system developed, named GAYUS, is intended for the naive users, and has the knowledge of several univariate theoretical distributions. In the decision making process, GAYUS uses a hybrid approach to integrate various heuristics, goodness-of-fit tests and graphical displays. Having these features, the system may be an example for the "simulation-based software," or for the "cognizant front-end." A group of users tested the performance of the system. In the testing process, its users utilized random samples generated by the method of Monte Carlo. According to the users, GAYUS has an easy-to-use and a friendly interface. The test results indicated that as the sample size increases, the system

better predicts the correct theoretical distribution from which the sample is drawn. We believe that the performance of GAYUS could be improved by increasing the variety of criteria used, and by utilizing more powerful input analysis techniques in the diagnosis process.

Key Words: *Simulation-based software, Artificial intelligence, Input analysis, Graphical methods, Univariate theoretical distributions*

Tablo 1. Girdi analizi sonuçları.

Teorik DAĞILIŞLAR	AÇIKLAMA	TEST	PUAN
NORMAL (0.594;0.285)	—	0	0.075
BETA (-0.062;-0.042)	min	0	0.015
LogNORMAL (-0.957;1.062)	—	2	0.64850151364
ÜSTEL (0.594)	max	2	0.96386497191
GAMMA (1.238;0.48)	—	2	0.54454456556
UNIFORM (0.00982;2.26975)	—	0	0.03

Tablo 2. Ayrıntılı puan tablosu.

AYRINTILI PUAN TABLOSU						
(0,3) TEST	(0,1) D ailesi	(0,015) FARKh	(0,05) FARKg	(0,05) HİSTg	(0,05) Q-Qg	DAĞILIŞ alfa = % 5 puan
0	0	5	0	0	0	NORMAL 0.075
2	1	10	1	1	1	ÜSTEL 0.96386497191
2	0	7	0	0	0	LOGNORMAL 0.64850151364
0	0	2	0	0	0	UNIFORM 0.03
0	0	1	0	0	0	BETA 0.015
2	0	0	0	0	0	GAMMA 0.54454456556

Tablo 3. Kararın gerekçelendirilmesi.

BİLGİLENDİRME !
1- Veriler SÜREKLİ dağılıyor.
2- İyi uyum testleri için hata oranını % 5 olarak seçtiniz.
3- Girdi verilerinin standart sapmasının ortalamasına oranı 0.89872217848 olduğu için, verilerin ait olabileceği dağılış(lar)ın ÜSTEL olduğu düşümlenebilir.
4- Olasılık grafiklerine baktınız ve verilere en uygun dağılışın ÜSTEL olduğu sonucuna ulaştınız.
5- Kartil grafiklerine baktınız, y=x doğrusuna en çok benzeyen grafiğin ÜSTEL kartil grafiği olduğuna karar verdiniz.
6- Fark grafiklerine baktınız ve y=0 merkez olmak üzere dikey eksene göre en az deęişkenlik gösteren dağılışın ÜSTEL olduğuna karar verdiniz.

Tablo 4. İyi-uyum test sonuçları.

Ki-Kare İYİ UYUM TEST SONUÇLARI			(1:Ho KABUL, 0:Ho RED)
alfa	alfa		
0.05	0.01	DAĞILIŞLAR	
1	1	Üstel (0.594) Ki-Kare Hesap = 2.45863; p = 0.483	
1	1	LogNormal (-0.957;1.062) Ki-Kare Hesap = 3.92424; p = 0.27	
0	0	Uniform (0.00982;2.26975) Ki-Kare Hesap = 50.16656 ; p = 0	
0	0	Normal (0.594;0.285) Ki-Kare Hesap = 31.97382; p = 0	
1	1	Gamma (1.238;0.48) Ki-Kare Hesap = 4.05974; p = 0.255	
0	0	Beta (-0.062;-0.042) Ki-Kare Hesap = 9999; p = 0	
Kolmogorov-Simirnov İYİ UYUM TEST SONUÇLARI			
ALFA	ALFA		
0.05	0.01	DAĞILIŞLAR	
1	1	Üstel (0.594) K-S Hesap = 0.0383	
1	1	LogNormal (-0.957;1.062) K-S Hesap = 0.04969	
0	0	Uniform (0.00982;2.26975) K-S Hesap = 0.98334	
0	0	Normal (0.594;0.285) K-S Hesap = 0.28471	
1	1	Gamma (1.238;0.48) K-S Hesap = 0.04138	
0	0	Beta (-0.062;-0.042) K-S Hesap = 999	

Tablo 5. GAYUS'a ilişkin başarımların sınıma sonuçları.

Sistemin ürettiği sonuçlar		Sonuca ilişkin:	Örneklem ölçümü (n)			
			10	60	100	500
Verilerin türetildiği teorik dağılımın,	En yüksek puanı alması	Olasılık	0.4429	0.7881	0.7881	0.9214
		Ortalama puan	0.7769	0.7891	0.7651	0.7284
	En yüksek 2. puanı alması	Olasılık	0.2881	0.1881	0.1833	0.0690
		Ortalama puan	0.7067	0.6747	0.6735	0.6458
	En yüksek 3. puanı alması	Olasılık	0.1714	0.0238	0.0286	0.0095
		Ortalama puan	0.6415	0.6257	0.6213	0.2875
	En yüksek 4., 5. veya 6. puanı alması	Olasılık	0.0976	0.0000	0.0000	0.0000
		Ortalama puan	0.3750	-	-	-

Kanonik Korelasyon ve Açıklanmış Varyans Oranı İstatistiklerinin Başarı Düzeylerinin Tahminlenmesinde Kullanılması Üzerine bir Araştırma

Öznur İŞÇİ*

ÖZET

İncelenen değişkenleri bağımlı ve bağımsız olmak üzere iki set halinde toplamak mümkün ise, bu setler arasındaki karşılıklı ilişkileri araştırmak amacıyla kullanılan yöntemlerden biriside kanonik korelasyon analizidir. Bu çalışmada Bornova Anadolu Lisesi öğrencilerinin not ortalamaları ile psikolojik skorlar arasındaki ilişki incelenmiştir. Öğrencilerin not ortalamalarını alarak kanonik korelasyon ve açıklanmış varyans oranı istatistiklerini GENSTAT paket programı ile hesaplayarak yorumlamalar yapılmaktadır. Çalışmada kanonik korelasyon analizi, kanonik açıklama analizi ve diğer çok değişkenli yöntemlerle ilişkileri gösterilmiş ve Wollenberg'in ortalama açıklanmış varyans yönteminin 9x9 korelasyon matrisi üzerinde bir uygulaması verilmiştir.

Anahtar Kelimeler : Kanonik korelasyon analizi, açıklanmış varyans oranı istatistikleri, çok değişkenli yöntemler.

1. GİRİŞ

Kanonik Korelasyon Analizi(Canonical Correlation Analysis), $p > 1$ ve $q > 1$ sayıda değişken içeren iki veri seti (x ve y) arasındaki değişkenler arası ilişkileri ortaya koymak amacıyla yararlanılan bir yöntemdir. Basit korelasyon analizi, iki değişken (y_i, x_i $i=1,2,\dots,n$) arasındaki ikili ilişkiyi r_{xy} korelasyon katsayısı aracılığı ile değerlendiren bir yöntemdir. Çoklu korelasyon analizi ise bir bağımlı değişken ile iki ya da daha fazla bağımsız değişken ($y_1, x_1, x_2, \dots, x_p$) arasındaki ilişkileri değerlendiren bir yöntemdir. Kanonik korelasyon ise, iki veya daha fazla değişken içeren değişkenler seti ($x_1, x_2, \dots, x_p; y_1, y_2, \dots, y_q$) arasındaki ilişkiyi doğrusal bileşenler aracılığı ile değerlendiren çok değişkenli bir yöntemdir(Özdamar,1999).

Kanonik açıklama analizinin kanonik korelasyon ve çok değişkenli çoklu regresyon arasında yer aldığı söylenebilir. Çok değişkenli çoklu regresyon bir orijinal uzaydan diğer bir orijinal uzaya tekil bir dönüşüm sağlamaktadır. Halbuki kanonik açıklama analizi bir setin değişkenlerinin tümünü aynı anda dikleştirmekte ve diğer setin orijinal uzayını tahminlemek için kullanılmak üzere dik bir dönüşüm sağlamaktadır.

* Dr., Celal Bayar Üniversitesi, Bilgisayar Araştırma ve Uygulama Merkezi

Kanonik korelasyon analizinde yapısal korelasyonların (bir setin değişkenleri ile o setin faktörleri arasındaki korelasyonlar) zayıf olması durumunda, değişkenlerin faktörler tarafından açıklanmış varyans oranları da oldukça düşük gerçekleşecektir. Buna karşılık bu iki önemsiz faktör arasında yüksek bir kanonik korelasyon bulunabilir. Gerçekte bu güçlükler kanonik korelasyon analizinin yapısal korelasyonlar üzerinde durmamasından kaynaklanmaktadır ve faktörlerin yorumlanmasında güçlükler neden olmaktadır(Wollenberg,1977). Halbuki en yüksek korelasyonlu olarak elde edilen bu faktörlere bir setteki değişkenlerin ne derece katkıda bulunduğu bilinmesi gerçekten önemlidir. Bu da yapısal korelasyonların bilinmesini gerektirir.

2. KANONİK KORELASYON ANALİZİ

Kanonik korelasyon analizi çok değişkenli bir istatistik tekniktir. Kanonik korelasyon analizi gittikçe artan bir uygulama alanına sahip olmakla beraber, bağımlı (kriter) ve bağımsız (tahmin) değişken setleri arasında ilişki araştırıldığı zaman kullanılan tekniklerden biridir(Şen H., Kalyoncu, C.,2001).

Kanonik korelasyon analizi, iki değişken seti arasındaki ilişkilerin incelenmesi ve bunların yorumlanması amacıyla kullanılır. Çok boyutlu sistemdeki değişkenlerin bağımlı ve bağımsız olmak üzere iki sete ayrılabilmesinin mümkün olduğu durumlarda uygulanan yöntemlerden birisi olan kanonik korelasyon analizi, ilk olarak Hotelling(1936) tarafından geliştirilmiştir. Analizin esas amacı, her iki değişken setine ait olmak üzere, aralarında en yüksek korelasyon bulunan doğrusal bileşenlerin elde edilmesidir. Bu en yüksek korelasyona kanonik korelasyon, elde edilen doğrusal bileşenlere ise, kanonik değişkenler adı verilmektedir. Bununla birlikte önemli kanonik korelasyon sayısı birden çok olabilmektedir.

Kanonik korelasyon analizi, rassal değişkenler kümesinin doğrusal fonksiyonları arasındaki en büyük korelasyonları hesaplamaya çalışır ve çok boyutlu ana kümeden çekilmiş iki veya daha çok değişken kümesi arasındaki ilişki ile ilgilenir. Rassal değişkenler kümesinin doğrusal fonksiyonları arasındaki maksimum korelasyonları hesaplamaya çalışır. Kanonik korelasyon ile, rassal değişkenler kümesinin maksimum korelasyonlu ve birim varyanslı doğrusal bileşimi elde edilmek istenir. Kanonik analiz çoklu regresyon analizinin özel bir halidir. Çoklu regresyon analizi bir bağımlı, birden fazla bağımsız değişken arasındaki ilişkiyi araştırmasına karşın kanonik analizde p ve q tane değişken bulunmaktadır.

2.1 Kanonik Değişkenlerin Elde Edilmesi

$q < p$ olmak üzere, $x(p+q) \times 1$ şans vektörünün $x_1(q \times 1)$ ve $x_2(p \times 1)$ şeklinde iki alt sete ayrıldığı varsayalım. Bu alt setler,

$$x_1 = (x_1, x_2, \dots, x_q)^T$$

ve

$$x_2 = (x_{q+1}, x_{q+2}, \dots, x_p)^T$$

şeklinde gösterilebilir. Burada x_1 seti açıklayıcı değişkenler seti ve x_2 seti ise bağımlı değişkenler seti adı verilmektedir. Bunun sonucu olarak, x şans vektörünün ortalaması ve varyans-kovaryans matrisi de,

$$\mu_x = \begin{pmatrix} \mu_{x_1} \\ \mu_{x_2} \end{pmatrix}, \quad \text{Cov}(x) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad (1)$$

şeklinde parçalanabilecektir. Böylece,

$$\Sigma_{11} = \text{Cov}(x_1), \quad \Sigma_{12} = \text{Cov}(x_1, x_2) = \Sigma_{21}^T, \quad \Sigma_{22} = \text{Cov}(x_2)$$

olduğu açıktır. x_1 setine ait doğrusal bileşen $a = w^T x_1$ ve x_2 setine ait doğrusal bileşen ise $b = v^T x_2$ olsun. Görüldüğü gibi, a ve b yapay değişkenleri orijinal değişkenlerin doğrusal kombinasyonları olarak ifade edilmektedir. a ve b yapay değişkenleri için,

$$\text{Var}(a) = w^T \Sigma_{11} w, \quad \text{Var}(b) = v^T \Sigma_{22} v, \quad \text{Cov}(a, b) = \text{Cov}(b, a)^T = w^T \Sigma_{12} v$$

ifadeleri geçerlidir. Buna dayanarak a ve b doğrusal bileşenleri arasındaki korelasyon,

$$\rho_{a,b} = \frac{w^T \Sigma_{12} v}{\left[(w^T \Sigma_{11} w) (v^T \Sigma_{22} v) \right]^{1/2}} \quad (2)$$

olacaktır. $\text{Var}(a) = w^T \Sigma_{11} w$ ve $\text{Var}(b) = v^T \Sigma_{22} v$ şeklinde birim varyans kısıtlamaları altında, $\rho_{a,b}$ korelasyonu en yüksek yapan w ve v vektörlerinin elde edilmesi kanonik analizin çözüm aşamasını oluşturmaktadır. Kısıtlar için iki lagrange çarpanı kullanılarak amaç fonksiyonu şu şekilde verilebilir:

Amaç fonksiyonunun çözümü araştırıcısı,

$$Z = w^T \Sigma_{12} v - 1/2 \lambda_1 (w^T \Sigma_{11} w - 1) - 1/2 \lambda_2 (v^T \Sigma_{22} v - 1) \quad (3)$$

$$(\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} - \lambda_1^2 I) w = 0 \quad (4)$$

ve

$$(\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} - \lambda_2^2 I) v = 0 \quad (5)$$

özdeğer-özvektör problemlerine götürmektedir. (4) ve (5) ifadelerinden elde edilecek özdeğerler (λ_1^2, λ_2^2) birbirlerine ve aynı zamanda da kanonik korelasyonun karesine eşittirler. Burada, $\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ ve $\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$ matrisleri pozitif tanımlı matrisler olsunlar ve rankı küçük olanın rankı k olsun. Böylece, yukarıda verilen (4) ve (5) ifadelerinden elde edilecek özdeğerler için,

$$\lambda_{11}^2 > \lambda_{12}^2 > \dots > \lambda_{1k}^2 > 0$$

yazılabilir ve $(w_1, v_1), (w_2, v_2), \dots, (w_k, v_k)$ ise bu özdeğerlere karşılık gelen özvektörlerdir.

2.2 Kanonik Değişkenlerin Kovaryans Yapısı

Σ_{11} ve Σ_{22} matrislerinin tekil olmadıkları varsayılın ve $K(q \times q)$ matrisi,

$$K = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} \quad (6)$$

şeklinde tanımlansın. Buna dayalı olarak da N_1 ve N_2 matrisleri,

$$N_1 = KK^T \quad N_2 = K^TK \quad (7)$$

olsun. Böylece,

$$M_1 = \Sigma_{11}^{-1/2} N_1 \Sigma_{11}^{1/2} \quad M_2 = \Sigma_{22}^{-1/2} N_2 \Sigma_{22}^{1/2} \quad (8)$$

matrisleri için,

$$M_1 = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \quad (9)$$

$$M_2 = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \quad (10)$$

yazılabilecektir. (9) ve (10) ifadelerinden görüldüğü gibi M_1 ve M_2 matrisleri kanonik korelasyon analizinde özdeğer-özvektör ayrışımı uygulanan matrisler olmaktadır.

3. AÇIKLANMIŞ VARYANS ORANLARI

İki değişken seti arasındaki ilişkilerin incelenmesi amaçlandığında, kanonik korelasyon istatistiğine alternatif bir istatistik Stewart ve Love(1968) tarafından ileri sürülen açıklanmış varyans oranı(redundancy index) istatistiğidir. Bilindiği gibi kanonik korelasyon analizinde, aralarında en yüksek korelasyona sahip doğrusal bileşenler elde edilmekte, diğer bir ifadeyle a_i ve b_i gibi iki yapay değişken arasındaki korelasyon en yüksek yapılmakta fakat elde edilen bu yapay değişkenlerin kendi setlerindeki değişimin ne kadarını açıklayabildikleri üzerinde hiç durulmamaktadır.

Bu durum en yüksek korelasyonlu olarak elde edilen kanonik değişkenlerin yorumlanmasında araştırmacıya önemli güçlükler getirmektedir [Van Den Wollenberg, A.L.(1977) ve De Sarbo, W.S.(1981)]. Çünkü x_1 ve x_2 değişken setlerinden elde edilen en yüksek korelasyonlu a_1 ve b_1 doğrusal bileşenleri, kendi setlerindeki değişimin çok küçük bir kısmını açıklıyor olabilmektedir. İki set arasındaki karşılıklı ilişkilerin yorumlanmasının ve değerlendirilmesinin kendi setlerini temsil gücü bu derece zayıf olan doğrusal bileşenlere dayandırılmasının uygun olmayacağı açıktır.

Bu sorunun çözümü için, Wollenberg(1977), kendi setlerindeki varyasyonun mümkün olan en yüksek kısmını açıklayabilme özelliğine sahip (elde edildikleri setleri temsil gücü bakımından en güçlü faktörleri veren) doğrusal bileşenlerin elde edilmesi ile ilgili bir yöntem geliştirilmiştir.

3.1 Yapısal Korelasyonlar ve Açıklanmış Varyans İndeksi

x_1 setinin değişkenleri ile bu setin i . kanonik faktörü arasındaki yapısal korelasyonları (faktör yüklerini) içeren $h_i(p \times 1)$ sütun vektörü,

$$h_i = \text{Cov}(x_1, c_i) = E\{[x_1 - E(x_1)][t_i^T x_1 - E(t_i^T x_1)]^T\} \\ = E\{[x_1 - \mu_{x1}][x_1 - \mu_{x1}]^T t_i\} = \sum_{11} t_i = R_{11} t_i \quad (11)$$

olarak elde edilir. Benzer olarak, y seti değişkenleri ile bu setin i . kanonik faktörü arasındaki yapısal korelasyonları içeren $g_i(q \times 1)$ sütun vektörü,

$$g_i = \text{Cov}(x_2, d_i) = E\{[x_2 - E(x_2)][s_i^T x_2 - E(s_i^T x_2)]^T\} \\ = E\{[x_2 - \mu_{x2}][x_2 - \mu_{x2}]^T s_i\} = \sum_{22} s_i = R_{22} s_i \quad (12)$$

şeklinde elde edilir. Stewart ve Love(1968), bu yapısal korelasyonlardan hareket ederek "açıklama indeksi" (redundancy index) olarak adlandırılan bir açıklanmış varyans indeksi tanımlamışlardır. Bu indeks, bir setin değişkenlerinin diğer setin verilen bir faktörü tarafından açıklanan varyans olarak tanımlanır. i . kanonik değişken çifti dikkate alınarak, x_1 ve x_2 setlerine ait açıklama indeksleri,

$$R_{x1i} = \lambda^2_{1i} (h_i^T h_i) / q \quad (13)$$

$$R_{x2i} = \lambda^2_{2i} (g_i^T g_i) / p \quad (14)$$

şeklinde tanımlanırlar (Wollenberg, 1977). Ayrıca Stewart ve Love (1968), bir sete ait k tane açıklama indeksini toplayarak o setin bütün kanonik faktörleri tarafından açıklanabilen toplam varyans oranını bir indeks olarak ifade etmişlerdir. Buna göre, x_1 ve x_2 setlerinin toplam açıklanmış varyans oranları sırasıyla,

$$R_{d1} = \sum R_{x1i} \quad (15)$$

$$i = 1, 2, \dots, k$$

$$R_{d2} = \sum R_{x2i} \quad (16)$$

olarak tanımlanır.

3.2 Açıklama İndeksinin En Yüksek Yapılması Sorunu

Açıklama indeksini maksimum yapan faktörler bir anlamda tahminleme gücü optimum olan faktörler olacağından, daha çok tahminleme amacına dayanan araştırmalarda açıklama indeksini en yüksek yapan faktörlerin kullanılması kanonik faktörlerin kullanılmasından daha uygundur.

Örneğin bir pazarlama yöneticisi tüketici marka tercihleri(y), bireysel psikolojik skorlarla(x) tahminleyebilmek için, y seti değişkenlerinin, $w \times x$ faktörü tarafından açıklanan varyansını maksimum yapacak bir w vektörü araştıracaktır (De Sarbo, 1981).

Van Den Wollenberg(1977), kanonik korelasyon istatistiği yerine, açıklama indeksi istatistiğini en yüksek yapan bir yöntem geliştirmiştir. Yöntemin esas amacı R_{x1i} ve R_{x2i}

istatistiklerini en yüksek yapan birim varyanslı $c=t^T x_1$ ve $d=s^T x_2$ doğrusal bileşenlerin belirlenmesidir. Burada, c ve d değişkenleri “en yüksek varyans açıklayıcı değişkenler” olarak tanımlanır. Bu işi yapan analize ise, “kanonik açıklama analizi” denilmektedir.

$R_{x_{1i}}$ ve $R_{x_{2i}}$ indekslerini en yüksek yapmak için, bu ifadelerin payındaki açıklanmış varyansları en yüksek yapmak yeterli olacaktır. Daha önce verilen birim varyans kısıtlamaları altında, kısıtlı optimizasyon probleminin amaç fonksiyonları,

$$\begin{aligned} Z_1 &= t^T R_{12} R_{21} t - \mu_1 (t^T R_{11} t - 1) \\ \text{ve} \\ Z_2 &= s^T R_{21} R_{12} s - \mu_2 (s^T R_{22} s - 1) \end{aligned} \quad (17)$$

şeklinde verilir. (17) nolu ifadenin t ve s’ye göre kısmi türevleri alınıp sıfıra eşitlenirse,

$$\begin{aligned} (R_{11}^{-1} R_{12} R_{21} - \mu_1 I) t &= 0 \\ \text{ve} \\ (R_{22}^{-1} R_{21} R_{12} - \mu_2 I) s &= 0 \end{aligned} \quad (18)$$

ifadeleri elde edilir. (18) ifadesi ile verilen ilk homojen eşitliğin $t=0$ dan farklı çözümünün olabilmesi için μ_1 ’in $R_{11}^{-1} R_{12} R_{21}$ matrisinin bir özdeğeri, t’nin de bu özdeğere karşılık gelen özvektör olması gerekir. $R_{11}^{-1} R_{12} R_{21}$ matrisinin tam ranklı olduğu ve özdeğerlerinin birbirlerinden farklı oldukları varsayıldığında,

$$\mu_{11} > \mu_{22} > \dots > \mu_{1p} > 0$$

yazılabilecektir. Daha önce verilen birim varyans kısıtlaması gereği,

$$\mu_{1i} = t_i^T R_{12} R_{21} t_i \quad (19)$$

elde edilir. Bu sonuca göre, x_2 seti değişkenlerinin, x_1 setinin i.inci en yüksek varyans açıklayıcı değişkeni tarafından açıklanan varyansı μ_{1i} özdeğerine eşit olmaktadır. μ_{1i} özdeğeri x_2 setinin i.inci açıklama indeksi $R_{x_{2i}}$ ’nin q katı olarak yorumlanabilir. Bu durumda x_2 setinin toplam açıklama indeksini gösteren (16) ifadesi,

$$R_{d_2} = 1/q \sum \mu_{1i} \quad i = 1, 2, \dots, p \quad (20)$$

şeklinde yazılabilecektir. Diğer taraftan, $R_{x_{1i}}$ indeksini en yüksek yapmak için, (18) ifadesi ile verilen ikinci homojen eşitlik, μ_{2i} ve s_i için çözüldüğünde ve ilgili basitleştirmeler yapıldığında,

$$\mu_{2i} = s_i^T R_{21} R_{12} s_i \quad (21)$$

elde edilir. Bu son eşitliğe göre, μ_{2i} özdeğeri, x_1 setinin i.inci açıklama indeksi $R_{x_{1i}}$ ’nin p katı olarak yorumlanabilir. Bu sonuca dayanarak, x_1 setinin toplam açıklama indeksini gösteren (15) ifadesi aşağıda verildiği gibidir,

$$R_{d_1} = 1/p \sum \mu_{2i} \quad i = 1, 2, \dots, q \quad (20)$$

4. UYGULAMA

Bu bölümde, kanonik korelasyon analizi ve kanonik açıklama analizi daha önce üzerinde durulan istatistiksel özellikler bakımından karşılaştırılabilmek amacıyla, Tablo 1'deki korelasyon matrisi esas alınarak her iki analizin uygulaması yapılmıştır. Bornova Anadolu Lisesi öğrencilerinin fen dersleri ve sosyal derslerindeki başarı düzeylerinin, bağımsız değişkenler setini oluşturan psikolojik, sosyal ve çevre ile ilgili skorlara dayalı olarak tahminlenmesi amacıyla Lise 2'de okuyan 6 sınıftaki 216 öğrenciye anket uygulanmıştır. Bazı öğrencilerin derste bulunmaması ya da anket formundaki eksik bilgiler dolayısıyla 150 öğrencinin formu değerlendirilmiştir. Lise öğrencilerinin başarı düzeylerini etkileyebileceği düşünülen değişkenler ve 3 döneme ilişkin her dersten aldıkları notlar dikkate alınmıştır. (İşçi, 1995). Bağımsız değişkenler seti (x) aşağıdakileri kapsar:

- x₁= Babanın mesleği
- x₂= Annenin mesleği
- x₃= Arkadaşlık ilişkileri
- x₄= Okulu kendi isteği ile seçip seçmediği?
- x₅= Anne ve baba ayrı mı?
- x₆= Özel ders alıyor mu?
- x₇= Kimin yanında kalarak okuyor?
- x₈= Okulun bulunduğu şehrin kültürel yapısını benimsiyor mu?
- x₉= Ailede geçimsizlik ya da huzursuzluk ortamı var mı?

Bağımlı değişkenler seti (y) aşağıdakileri kapsar:

- y₁= Matematik
- y₂= Geometri, Analitik Geometri
- y₃= Fen Bilimleri(Fizik, Kimya, Biyoloji)
- y₄= Coğrafya
- y₅= Tarih
- y₆= Yabancı Dil(İngilizce, Almanca)
- y₇= Türk Dili Edebiyatı
- y₈= Din
- y₉= Beden

Bu anket ile öğrencilerin başarılarını etkileyebileceği düşünülen faktörler ve psikografik cevaplar alınmıştır. Öğrenciler 1. soruda anne ve babasının mesleğini işaretlemişlerdir. 2. soruda da arkadaşlarıyla olan ilişkilerinde; çok kötü(-2), kötü(-1), arkadaşlık benim için önemli değil(0), normal(1), iyi(2), çok iyi(3) olarak derecelendirilmiştir. 3,4,5 ve 8. sorular da evet(1), hayır(2) olarak belirlenmiştir. 6. soruda kimin yanında kalarak okuyorsunuz sorusuna ailemle(1), öğrenci yurdunda(2), arkadaşlarımla aynı evde kalıyorum(3); 7.soruda okulunuzun bulunduğu şehrin kültürel yapısını ne derece benimsiyorsunuz sorusuna hiç benimsemiyorum(0), çok az benimsiyorum(1), orta derece benimsiyorum(2), tam benimsiyorum(3) olarak dikkate alınmıştır. Öğrencilerin 3 dönemde aldıkları ortak ya da seçmeli dersler birleştirilerek her derse ilişkin tek bir not hesaplanmıştır. Örneğin matematik 1. dönem 4, 2. dönem 3, 3. dönem 5 olsun, matematik notu 4 olarak değerlendirilmiştir. Diğer ders notları da benzer şekilde hesaplanmıştır. Tablo 1'de kanonik korelasyon analizi ve kanonik açıklama analizi için korelasyon matrisi görülmektedir. Tablo 1'de verilen korelasyon matrisi, ilk 9 değişken bağımsız değişkenler setini, diğer 9 değişken bağımlı değişkenler setini meydana getirecek şekilde parçalanmıştır.

Tablo 1: Orijinal Değişkenlerin Korelasyon Matrisi (R)

	R ₁₁									R ₁₂									
1.0000																			
0.2500	1.0000																		
0.1200	0.1900	1.0000																	
-0.1800	-0.0200	-0.3200	1.0000																
0.0900	0.0900	0.2200	-0.2700	1.0000															
-0.0800	-0.0200	-0.0700	0.1100	-0.1600	1.0000														
-0.1400	0.0200	-0.0900	0.2500	-0.0100	0.0100	1.0000													
0.2500	0.2300	0.2400	-0.3800	0.2600	-0.0200	-0.0200	1.0000												
0.1200	0.1600	0.2500	-0.1700	0.0300	0.0000	-0.0100	0.2000	1.0000											
-0.0700	0.0600	0.0500	-0.0500	0.0600	-0.0100	-0.0500	-0.1000	0.1400	1.000										
0.0900	0.0700	-0.0600	-0.0100	0.0700	0.0800	0.0600	-0.1000	0.1200	0.6400	1.000									
0.1000	0.1200	-0.0100	0.0500	0.1700	-0.0100	0.0000	-0.1300	0.0600	0.7600	0.7800	1.0000								
0.0200	-0.0700	-0.0500	0.0200	0.0100	0.0500	0.0200	0.0100	0.1400	0.5500	0.5200	0.5900	1.0000							
0.0300	0.0100	0.0600	0.0500	0.0100	0.0700	0.0000	-0.0500	0.0400	0.6200	0.5400	0.6700	0.6100	1.0000						
0.1900	0.0700	0.1900	-0.0400	0.1100	0.0100	0.0400	-0.0700	0.1000	0.5800	0.5700	0.7300	0.5100	0.6200	1.0000					
0.0400	0.0700	0.1800	-0.0400	0.1700	-0.0100	-0.0800	-0.0700	0.0900	0.6400	0.5200	0.6600	0.4200	0.6500	0.6200	1.0000				
-0.0200	0.0200	0.0500	-0.0100	0.1400	0.1100	-0.0100	-0.0600	0.0200	0.5100	0.5200	0.5500	0.5000	0.6300	0.5500	0.5900	1.0000			
0.0800	0.0400	0.0000	0.0200	-0.0100	0.0900	-0.1400	-0.0700	0.0900	0.2100	0.1400	0.1700	0.1800	0.1700	0.0500	0.2100	0.1700	1.0000		

R₂₁

R₂₂

56

Kanonik Korelasyon ve Açıklanmış Varyans Oranı İstatistiklerinin Başarı Düzeylerinin Tahminlenmesinde Kullanılması Üzerine bir Araştırma

Kanonik korelasyon analizinin x ve y setleri ile ilgi sonuçlar sırasıyla, Tablo 2 ve 3'te verilmiştir.

Tablo 2: x Setine Ait Kanonik Korelasyon Analizi Sonuçları

Kök no	λ_i	λ_i^2	H_i	$R_{xi} = \lambda_i^2 H_i$	R_{xi} / R_{d1}
1	0.4690	0.2199	0.0946	0.0208	0.2845
2	0.4116	0.1694	0.0993	0.0168	0.2298
3	0.3620	0.1310	0.1169	0.0153	0.2063
4	0.3012	0.0907	0.0820	0.0074	0.1012
5	0.2096	0.0439	0.0822	0.0034	0.0465
6	0.1990	0.0396	0.1079	0.0043	0.0588
7	0.1880	0.0353	0.1187	0.0042	0.0574
8	0.0790	0.0062	0.1458	0.0009	0.0123
9	0.0138	0.0002	0.1526	0.0000	0.0000

Toplam **1.0000** **$R_{d1} = 0.0731$**

Tablo 3: y Setine Ait Kanonik Korelasyon Analizi Sonuçları

Kök no	λ_i	λ_i^2	G_i	$R_{yi} = \lambda_i^2 G_i$	R_{yi} / R_{d2}
1	0.4690	0.2199	0.1193	0.0262	0.3819
2	0.4116	0.1694	0.0399	0.0068	0.0991
3	0.3620	0.1310	0.0554	0.0073	0.1064
4	0.3012	0.0907	0.1781	0.0162	0.2361
5	0.2096	0.0439	0.0931	0.0041	0.0598
6	0.1990	0.0396	0.0857	0.0034	0.0496
7	0.1880	0.0353	0.0740	0.0026	0.0379
8	0.0790	0.0062	0.3170	0.0020	0.0291
9	0.0138	0.0002	0.1526	0.0000	0.0000

Toplam **1.0000** **$R_{d2} = 0.0686$**

Tablolarda yer alan $H_i = h_i^T h_i / p$ değeri a_i faktörlerinin x setinde açıkladıkları varyans oranını, $G_i = g_i^T g_i / q$ değeri b_i faktörlerinin y setinde açıkladıkları varyans oranını ifade etmektedir. λ_i ise i.inci kanonik korelasyondur. Tablo 2 ve Tablo 3 incelendiğinde x ve y setlerinin aralarında en yüksek korelasyona sahip en önemli faktörlerin, kendi setlerindeki değişimin sırasıyla ancak 0.0946 ve 0.1193'lük kısımlarını açıklayabildikleri görülmektedir. Bu, birinci kanonik faktör çiftinin bile açıklanan varyans anlamında kendi setlerini temsil etme gücünün çok zayıf olduklarını göstermek bakımından oldukça anlamlıdır. x setinin birinci faktörü, y setindeki değişimin 0.0262'sini açıklayabilmektedir ve bu değer de y setinde açıklanması mümkün olan varyans oranının 0.3819'una karşılık gelmektedir. y setinin birinci

faktörü, x setindeki değişimin ancak 0.0208'ini açıklayabilmektedir ve bu değer de x setinde açıklanması mümkün olan varyans oranının yalnızca 0.2845'idir. y setinin en önemsiz faktörlerinden biri olarak dikkate alınabilecek olan $v_8(0.3170)$ doğrusal bileşeni, bu sete ait tüm doğrusal bileşenler içerisinde en yüksek varyans açıklama yüzdesine sahip olan bileşendir. Benzer şekilde, x setinin en önemsiz faktörlerinden birisi olarak dikkate alınabilecek $w_9(0.1526)$ doğrusal bileşeni, bu sete ait tüm doğrusal bileşenler içerisinde en yüksek varyans açıklama yüzdesine sahip olan bileşendir.

Tablo 4: Kanonik Korelasyon Analizinde x Setine Ait Kanonik Faktörler

	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈	a ₉
x ₁	0.4781	0.6731	-0.5572	-0.0851	-0.1996	-0.1393	0.0223	0.2753	-0.1314
x ₂	0.3749	-0.3304	0.1156	0.2061	0.3366	-0.1092	0.5797	0.0288	0.6216
x ₃	0.1037	0.6630	0.7307	-0.3486	-0.0107	0.1591	0.0125	-0.3085	0.0393
x ₄	0.2085	-0.1135	-0.1439	-0.6016	-0.6931	-0.0002	-0.0231	-0.6716	0.1984
x ₅	0.6904	-0.3889	0.0375	0.0726	-0.0557	0.2244	-0.6890	-0.0124	0.0005
x ₆	0.0320	0.1652	-0.2332	0.2254	0.0310	0.9379	0.0367	-0.0249	0.1543
x ₇	-0.0583	0.3442	-0.3384	0.0316	0.7821	-0.1572	-0.2941	-0.3644	0.0546
x ₈	-0.5942	-0.0916	-0.1460	-0.5950	-0.2387	-0.0108	-0.3128	0.2286	0.6245
x ₉	-0.1501	0.0672	-0.0674	0.7676	-0.4665	-0.2446	-0.2656	-0.3292	0.1922

Tablo 5: Kanonik Korelasyon Analizinde y Setine Ait Kanonik Faktörler

	b ₁	b ₂	b ₃	b ₄	b ₅	b ₆	b ₇	b ₈	b ₉
y ₁	-0.6774	-0.3088	0.8970	0.8054	0.3056	-0.3082	0.4278	-0.3014	-0.5130
y ₂	-0.2756	0.4419	-0.7054	0.8805	0.5164	0.2459	0.0872	0.2010	0.8801
y ₃	1.4593	-1.2296	-0.5097	-0.6724	-0.2602	-0.2428	0.0362	-0.2548	-0.3899
y ₄	-0.5321	-0.0027	-0.3683	0.1100	-0.6986	-0.2547	-0.9343	-0.0817	-0.1711
y ₅	-0.3802	0.2892	-0.2155	-0.6307	-0.0626	0.3600	0.8660	-0.9609	0.4331
y ₆	0.2234	1.3818	0.1741	0.1365	0.2582	-0.1971	-0.0249	0.0743	-0.6294
y ₇	0.2348	0.0725	0.7158	-0.0021	-0.4970	-0.0708	-0.6440	0.2945	1.0632
y ₈	0.1843	-0.4606	0.1268	-0.0133	0.4147	1.0343	-0.3940	0.1213	-0.5638
y ₉	0.1925	0.2093	-0.1156	0.2879	-0.6399	0.2948	0.5112	0.3349	-0.2228

Kanonik korelasyon analizi uygulaması sonucu elde edilen bağımsız ve bağımlı değişkenler setine ait kanonik değişkenler(kanonik bileşenler) sırasıyla Tablo 4 ve Tablo 5'de verilmiştir. Aralarında en yüksek korelasyona sahip birinci kanonik değişken çifti(w_1, v_1) dikkate alınsın. w_1 kanonik faktörün yorumuna, kendi setine ait değişkenler arasında en yüksek katkıyı sağlayan değişkenlerin sırasıyla x_5 (anne ve babanın ayrı olup olmaması), x_8 (okulun bulunduğu çevrenin kültürel yapısına uyum derecesi), x_1 (babanın mesleği) ve x_2 (annenin mesleği) değişkenleri olduğu görülmektedir. y setine ait en önemli faktör olan v_1 faktörü dikkate alındığında ise, bu doğrusal bileşenin yorumuna en yüksek katkıda bulunan değişkenlerin sırasıyla y_3 (fen bilimleri), y_1 (matematik), y_4 (coğrafya), y_5 (tarih) dersleri olduğu

görülmektedir. x seti, öğrencinin başarısını etkileyen psikolojik, sosyal ve ekonomik değişkenler seti ve y seti ise öğrencinin fen ve sosyal bilimlerle ilgili derslerindeki başarı durumunu yansıtan değişkenler olduğuna göre, öğrencinin içinde bulunduğu psikolojik, sosyal ve ekonomik yapıyı en fazla belirleyen bu değişkenlerin anne ve babanın ayrı olup olmaması, okulun bulunduğu çevrenin kültürel yapısına uyum derecesi, babanın mesleği ve annenin mesleği olduğu ifade edilebilir.

Tablo 4 ve Tablo 5'deki bilgiler dikkate alındığında, hangi değişkenlerin her iki setin en önemli faktörleri olan birinci faktör çiftine veya daha sonra gelen faktör çiftlerinin yorumuna en çok katkıda buldukları görülebilir. Kanonik korelasyon analizinde bağımlı ve bağımsız değişkenler setine ait kanonik faktörlerin tümü için sonuçlar mevcut ise de ilk faktörün yorumu yapılmıştır.

Tablo 6: x Setine Ait Kanonik Açıklama Analizi Sonuçları

Kök no	h_i	μ_{2i}	$R_{xi} = \mu_{2i}/p$	R_{xi} / R_{d1}
1	0.0873	0.2004	0.0223	0.3042
2	0.1020	0.1655	0.0184	0.2510
3	0.0905	0.1231	0.0137	0.1869
4	0.1030	0.0626	0.0069	0.0941
5	0.1098	0.0396	0.0044	0.0600
6	0.1052	0.0352	0.0039	0.0532
7	0.0826	0.0272	0.0030	0.0409
8	0.1662	0.0067	0.0007	0.0095
9	0.1534	0.0002	0.0000	0.0000

Toplam 1.0000 $R_{d1} = 0.0733$

Tablo 7: y Setine Ait Kanonik Açıklama Analizi Sonuçları

Kök no	g_i	μ_{1i}	$R_{yi} = \mu_{1i}/p$	R_{yi} / R_{d2}
1	0.1218	0.3749	0.0417	0.6079
2	0.0245	0.0753	0.0084	0.1224
3	0.0453	0.0575	0.0064	0.0933
4	0.2062	0.0403	0.0045	0.0655
5	0.0782	0.0277	0.0031	0.0452
6	0.0912	0.0238	0.0026	0.0379
7	0.0545	0.0131	0.0014	0.2040
8	0.3405	0.0041	0.0005	0.0073
9	0.0378	0.0001	0.0000	0.0000

Toplam 1.0000 $R_{d2} = 0.0686$

Tablo 6 ve Tablo 7'deki açıklanan varyans oranları incelenirse, kanonik açıklama analizinin faktörleri elde edildikleri setleri temsil gücü bakımından optimum bir şekilde sıralandığı görülmektedir. Tablo 6 dikkate alındığında, y setine ait tüm en yüksek varyans açıklayıcı faktörlerin, x setinin toplam değişimin genellikle %7.3'ünü açıklayabildiği görülmektedir. Bu değer, öğrencilerin fen ve sosyal derslerindeki başarı düzeylerine dayalı faktörlerin, bu öğrencilerin aile yapısı ve sosyal yapıları ile ilgili değişkenler setinde açıklayabilmesi mümkün olan en yüksek değişim yüzdesidir. Tablo 7 dikkate alındığında ise, x setine ait tüm en yüksek varyans açıklayıcı faktörlerin, y setindeki toplam değişimin % 6.9'unu açıklayabildiği görülmektedir.

Kanonik korelasyon analizi sonuçlarıyla karşılaştırma yapılırsa, kanonik açıklama analizinde x setinin ilk iki faktörü y setinde açıklanabilecek toplam varyasyonun 0.7303'ünü açıkladığı halde bu değer kanonik korelasyon analizinde 0.4810 düzeyinde kalmaktadır. y setinin ilk iki en yüksek varyans açıklayıcı faktörü x setinde açıklanabilecek toplam varyasyonun 0.5552'sini açıkladığı halde, bu değer kanonik korelasyon analizinde 0.5143'da kalmaktadır. Bu durum, kanonik korelasyon analizinin aksine, kanonik açıklama analizinde her iki sete ait tahminleme gücünün setlerin ilk iki faktöründe toplandığını göstermektedir.

Tablo 8: Kanonik Açıklama Analizinde x Setine Ait Kanonik Faktörler

	c ₁	c ₂	c ₃	c ₄	c ₅	c ₆	c ₇	c ₈	c ₉
x ₁	-0.3357	0.0615	0.8771	-0.1457	0.3083	-0.2378	-0.0791	0.2677	0.1254
x ₂	-0.2320	-0.0167	-0.2960	0.5051	0.2650	-0.3679	0.3867	-0.1610	-0.6134
x ₃	-0.1563	0.9461	-0.0034	-0.4243	-0.1870	0.0676	0.0855	-0.2832	-0.0374
x ₄	-0.0003	0.0985	0.0832	-0.2910	0.4886	0.1829	-0.7364	-0.6758	-0.1943
x ₅	-0.6160	-0.0027	-0.1703	0.3674	0.0766	0.5579	-0.4859	0.2574	0.0013
x ₆	-0.2777	-0.2150	0.1084	-0.3195	0.2123	0.6497	0.5564	0.0209	-0.1596
x ₇	0.0208	0.0212	0.6130	0.4543	-0.6023	0.1924	0.2043	-0.2627	-0.0513
x ₈	0.7886	0.0192	0.918	-0.2221	0.0467	0.2590	-0.3839	0.2081	-0.6305
x ₉	-0.4176	-0.5758	-0.0985	-0.4143	-0.4932	-0.2187	-0.2960	-0.0989	-0.1958

Tablo 9: Kanonik Açıklama Analizinde y Setine Ait Kanonik Faktörler

	d ₁	d ₂	d ₃	d ₄	d ₅	d ₆	d ₇	d ₈	d ₉
y ₁	-0.3769	0.6868	0.8908	-0.5799	-0.0301	0.3310	-0.7588	0.2151	0.5067
y ₂	-0.2892	0.3123	-0.8176	-0.6815	-0.1704	-0.4652	-0.5286	-0.3338	-0.8623
y ₃	0.5326	-1.9912	0.2356	0.1280	-0.1916	0.5198	0.2939	0.2855	0.3727
y ₄	-0.4448	0.2735	-0.1779	-0.2009	-0.6403	0.0141	1.0252	0.2407	0.1699
y ₅	-0.4576	0.0492	-0.4264	0.5847	0.8790	0.2420	-0.3756	0.8609	-0.4667
y ₆	0.8270	0.8818	-0.6804	0.2211	-0.1429	-0.0176	-0.3118	-0.0395	0.6510
y ₇	0.5882	0.3447	0.5680	-0.0127	-0.2059	-0.0469	0.7035	-0.1486	-1.0523
y ₈	-0.0701	-0.3468	0.4085	-0.0066	0.2811	-1.1327	0.0232	-0.1340	0.5650
y ₉	0.1809	0.0668	-0.2956	-0.4409	0.7028	0.2595	0.2712	-0.3024	0.2267

Kanonik açıklama analizi sonucu elde edilen bağımlı ve bağımsız değişkenler setine ait kanonik değişkenler sırasıyla Tablo 8 ve Tablo 9'da verilmiştir. Aralarında en yüksek korelasyona sahip birinci kanonik değişken çifti (t_1, s_1) dikkate alınsın. t_1 kanonik faktörünün yorumuna kendi setine ait değişkenler arasında en yüksek katkıyı sağlayan değişkenlerin sırasıyla, x_8 (okulun bulunduğu çevrenin kültürel yapısına uyum derecesi), x_5 (anne ve babanın ayrı olup olmaması), x_9 (ailede geçimsizlik ya da huzursuzluk ortamının olup olmadığı) ve x_1 (babanın mesleği) değişkenleri olduğu görülmektedir. y setine ait en önemli faktör olan s_1 faktörü dikkate alındığında ise, bu doğrusal bileşenin yorumuna en yüksek katkıda bulunan değişkenlerin sırasıyla y_6 (yabancı dil), y_7 (edebiyat), y_3 (fen bilimleri) ve y_5 (tarih) dersleri olduğu görülmektedir.

Bu araştırmada yapılan her iki uygulama için kullanılan bilgisayar programları Ege Üniversitesi Bilgisayar Mühendisliği bölümündeki GENSTAT istatistiksel paket programından yararlanılarak hazırlanmıştır. Uygulamada kullanılan kanonik korelasyon analizi ve kanonik açıklama analizi programlarının hazırlanmasında Okur, M.C ve Ertaş K.(1990)'daki programlardan faydalanılmıştır.

5. SONUÇ

Analizler arasındaki farkların açık bir biçimde gösterilmesini amaçlayan uygulamada, Bornova Anadolu Lisesi öğrencilerinin psikolojik, sosyal ve ekonomik faktörlere dayalı olarak başarı düzeylerinin belirlenmesi amacıyla elde edilen korelasyon matrisi esas alınarak her iki analizin uygulaması yapılmıştır. Bu çalışmada, bağımlı ve bağımsız değişken setleri için olmak üzere toplam 18 değişken bakımından 150 öğrenci üzerinde ölçümler yapılmıştır.

Kanonik korelasyon analizi uygulamalarında gözlem sayısının çok olması analiz sonuçlarının güvenilirliğini arttırmaktadır. Kanonik korelasyonların önemliliği Wilk's Lambda, Hotelling-Lawley, Pillai Trace, Roy's maksimum kök gibi test istatistikleri ile hesaplanabilir.

Çalışmamızda çok değişkenli istatistiksel analiz yöntemlerinden birisi olan kanonik korelasyon analizi ve kanonik açıklama analizi yöntemlerinin GENSTAT paket programında çözümü ele alınmıştır. Anket verileri düzenlendikten sonra GENSTAT paket programında veriler girilmiştir. Kanonik korelasyon analizi ve kanonik açıklama analizi yöntemlerinin tüm aşamaları gerçekleştirildikten sonra sonuçlar alınmıştır.

Bazı uygulamalarda kanonik korelasyonların iki değişken seti arasındaki karşılıklı ilişkiler hakkında önemli fikirler vereceği açıktır. Özellikle aralarında yüksek ilişki bulunan vektörlere ihtiyaç duyulan araştırmalarda bu analizi uygulamak yerinde olacaktır. Bununla beraber uygulamalarda birkaç kanonik vektör çifti en güvenilir faktör çiftleri olarak değerlendirilmekte ve yorumlamalar esasta bu çiftlere dayandırılmaktadır. Bazı durumlarda bu yorumlamanın tamamen geçersiz olabileceği ortaya konulmaya çalışılmış ve sözü edilen yorumlama güçlüklerini ortadan kaldırmak amacıyla Wollenberg(1977) tarafından gerçekleştirilen ve elde edildikleri setleri temsil etme bakımından en güçlü faktörleri veren kanonik açıklama analizi incelenmiştir. Analizler arasındaki farkların açık bir şekilde gösterilmesini amaçlayan uygulamada korelasyon matrisi esas alınarak her iki analiz uygulanmış ve sonuçların bir değerlendirmesi verilmiştir.

KAYNAKLAR

- DESARBO, W.S. (1981), *Canonical/ Redundancy Factoring Analysis*. Psychometrica,46,307-329.
- HOTELLING, H. (1936), *Relations Between Two Sets of Variates*. Biometrika, 28,321-377.
- İŞÇİ, Ö.(1995), *Kanonik Korelasyon ve Açıklanmış Varyans Oranı İstatistiklerini Optimum Yapan Faktörlerin Bornova Anadolu Lisesi Öğrencilerinin Başarı Düzeylerinin Tahminlenmesinde Kullanılması Üzerine Bir Araştırma*, DEÜ Sosyal Bilimler Enstitüsü, Yüksek lisans Tezi.
- OKUR, M. CUDİ, ERTAŞ, K.,(1990), *Kanonik Analizlerde Açıklanmış Varyans Oranlarının Maksimum Yapılması Üzerine Bir Araştırma*, Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt:5, Sayı:1-2,s:23-46.
- ÖZDAMAR, K.(1999), *Paket Programlar İle İstatistiksel Veri Analizi (Çok Değişkenli Analizler)-2*, Kaan Kitabevi, 385-427.
- STEWART, D.K., LOVE, V.A(1968), *A General Canonical Correlation Index*, Psychological Bulletin, 70,160-163.
- VAN DEN WOLLENBERG, A.L.(1977), *Redundancy Analysis:As an Alternative for Canonical Correlation Analysis*, Psychometrika, 42, 207-219.
- ŞEN, H., KALYONCU, C.(2001), *Eskişehir Osmangazi Üniversitesi Tıp Fakültesindeki Beslenme Bozukluğu (Malnütrisyon Sıklığı) İle İlgili Araştırmanın Kanonik Korelasyon Analizi İle Çözümlemesi*, Çukurova Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü V. Ulusal Ekonometri ve İstatistik Sempozyumu 19-22 Eylül 2001.

A Research on Using Canonical Correlation and Redundancy Analysis Statistics to Estimate Accomplishment Levels

ABSTRACT

If it is possible to collect examined variables into two sets as dependent and independent, canonical correlation analysis is one of the methods to search the relationship between these sets. In this study the relationship between Bornova Anatolian High School Students' average grades and psychological scores is examined. Comments are given by taking the students' average grades and calculating the canonical correlation and the redundancy statistics by using GENSTAT program. In the study the relationship among correlation analysis, redundancy analysis and the other multivariate statistical methods is shown and the application of Wollenberg's redundancy method on 9x9 correlation matrix is given.

Key Words : *Canonical correlation analysis, redundancy analysis, multivariate statistical methods.*

Lineer Modellerde Yarı Uzay Derinliğine Dayalı Dengeli Bootstrap Güven Bölgeleri

İhsan KARABULUT*

Fikri ÖZTÜRK*

ÖZET

Bir lineer modelde parametrelere ilişkin güven bölgelerinin oluşturulmasında son yıllarda uygulaması yaygınlaşmaya başlayan derinlik kavramı kullanılarak dengeli bootstrap güven bölgelerinin nasıl oluşturulabileceği bir uygulama ile gösterilmiştir.

Anahtar Kelimeler: Lineer model, bootstrap, dengeli güven bölgesi, yarı uzay derinliği.

1. GİRİŞ

Bilindiği gibi

$$Y = X\beta + \varepsilon$$

lineer modelinde X $n \times p$ boyutlu matris olup $\text{rank}(X) = p$, $n \geq p$ olması durumunda ve $E(\varepsilon) = 0$, $\text{Cov}(\varepsilon) = \sigma^2 I$ varsayımları altında $\beta \in R^p$ parametre vektörü ile $\sigma^2 \in (0, \infty)$ parametresi için alışılmış tahmin ediciler

$$\hat{\beta}_n = (X'X)^{-1} X'Y$$

ve

$$\hat{\sigma}^2 = \frac{Y'(I - X(X'X)^{-1}X')Y}{n - p}$$

dır.

$\hat{\beta}_n$ tahmin edicisi Gauss-Markov teoremi ile belirtilen özelliklere sahiptir. $\hat{\sigma}^2$ tahmin edicisi de Y 'nin karesel formu biçiminde olan yansız tahmin ediciler arasında enküçük varyanslıdır (Graybill(1976)). Ancak $\hat{\beta}_n$ ve $\hat{\sigma}^2$ 'nin dağılımları bilinmediğinde β ve σ^2 parametreleri ile ilgili hipotez testi yapmak ve güven bölgeleri oluşturmak mümkün olmamaktadır. Bu çalışmada β parametre vektörü için $\hat{\beta}_n$ tahmin edicisine bağlı bootstrap güven bölgesi ile derinliklere dayalı güven bölgeleri teorik ayrıntılarına inilmeden kısaca tanıtılıp açıklanmaktadır.

* Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü, 06100 Tandoğan Ankara. e-posta: kbulut@science.ankara.edu.tr

Dağılımdan bağımsız olan bu güven bölgelerinin işlerliğini görmek amacıyla $\alpha \in (0,1)$ olmak üzere $\varepsilon \sim N(0, \sigma^2 I)$ durumunda $1-\alpha$ güven düzeyli

$$P((\hat{\beta}_n - \beta)' X'X (\hat{\beta}_n - \beta) \leq p \hat{\sigma}^2 F_{1-\alpha; p, n-p}) = 1-\alpha \quad (1)$$

güven bölgesi ile karşılaştırılma yapılmaktadır.

Çalışmanın ikinci kısmında bootstrap güven bölgesi, üçüncü kısmında yarı uzay derinliğine dayalı dengeli güven bölgesi kavramları tanıtılmaktadır. Dördüncü kısımda β parametre vektörü için $\hat{\beta}_n$ en küçük kareler tahmin edicisine dayalı $1-\alpha$ düzeyinde bootstrap güven bölgesinin oluşturulmaktadır. Son kısımda iki açıklayıcı değişkenli ve sabit terimi olmayan yapay bir model üzerinde üretilmiş verilerle bir uygulama yapılmaktadır.

2. BOOTSTRAP GÜVEN BÖLGESİ

Parametre vektörleri için güven bölgesi kavramına geçmeden önce bazı hatırlatmalar yapalım. $Z \sim N(0,1)$ olmak üzere

$$K_{1-\alpha} = \{ z : z_{\alpha/2} \leq z \leq z_{1-\alpha/2} \}$$

aralığı için

$$P(Z \in K_{1-\alpha}) = 1-\alpha$$

dır.

$X_1, X_2, X_3, \dots, X_n$ bir örneklem ve θ bu örneklemin geldiği kitleye ait bilinmeyen bir parametreyi gösterebilir. $\hat{\theta}_n$, θ için ve $\hat{\sigma}_{\hat{\theta}_n}^2$, $Var(\hat{\theta}_n) = \sigma_{\hat{\theta}_n}^2$ için tahmin ediciler olsunlar. Z standart normal dağılımlı rasgele değişkeni göstermek üzere

$$n^{1/2} (\hat{\sigma}_{\hat{\theta}_n}^2)^{-1/2} (\hat{\theta}_n - \theta) \xrightarrow{d} Z$$

olduğunu varsayalım. Bu durumda $n^{1/2} (\hat{\sigma}_{\hat{\theta}_n}^2)^{-1/2} (\hat{\theta}_n - \theta)$ pivotuna dayalı yaklaşık $1-\alpha$ güven düzeyli bir güven aralığı

$$\{ \hat{\theta}_n - n^{-1/2} (\hat{\sigma}_{\hat{\theta}_n}^2)^{-1/2} z : z \in K_{1-\alpha} \}$$

dır.

$X_1, X_2, X_3, \dots, X_n$ örnekleminde üretilen B tane bootstrap örnekleminde

$$Z_n^b = n^{1/2} (\hat{\sigma}_{\hat{\theta}_n^b}^2)^{-1/2} (\hat{\theta}_n^b - \hat{\theta}_n), b=1,2,3,\dots,B$$

gözlenen bootstrap değerleri küçükten büyüğe sıralansın. Baştan ve sondan $\alpha/2$ oranlarında gözlemin dışında kalanların kümesi $W_{n,1-\alpha}$ ile gösterilsin.

$$BGA = \{ \hat{\theta}_n - (\hat{\sigma}^2)^{-1/2} w : w \in W_{n,1-\alpha} \} \quad (2)$$

kümesini kapsayan en küçük kapalı aralık, θ için $1-\alpha$ güven düzeyinde dengeli bir bootstrap güven aralığı (BGA) olarak tanımlanabilir. Bir diğer bootstrap güven aralığı Hall(1992)'de anlatıldığı gibi

$$\{\hat{\theta}_n^b : Z_n^b \in K_{1-\alpha}, b = 1, 2, 3, \dots, B\}$$

kümesini kapsayan en küçük kapalı aralık olarak tanımlanır. Burada bootstrap değerleri için bir sıralama yapılmadığı ve güven aralığının yukarıda tanımlanan $K_{1-\alpha}$ aralığında yer alabilenlerin oluşturduğunu belirtelim.

Benzer düşüncelerle p -boyutlu bir θ parametre vektörü için bootstrap güven bölgesinin oluşturulmasında $Z \sim N(0, I_{p \times p})$ olmak üzere

$$K_{1-\alpha} = \{z \in R^p : z'z \leq r_\alpha\}$$

küresi,

$$P(Z \in K_{1-\alpha}) = 1 - \alpha$$

olacak şekilde tanımlansın. n birimlik bir örnekleme dayalı bir $\hat{\theta}_n$ tahmin edicisi için

$$\sqrt{n} \hat{\Sigma}_{\hat{\theta}_n}^{-\frac{1}{2}} (\hat{\theta}_n - \theta) \xrightarrow{d} Z$$

olduğunda $\sqrt{n} \hat{\Sigma}_{\hat{\theta}_n}^{-\frac{1}{2}} (\hat{\theta}_n - \theta)$ pivotuna dayalı yaklaşık $1-\alpha$ güven düzeyli bir güven aralığı

$$\{\hat{\theta}_n - n^{-\frac{1}{2}} \hat{\Sigma}_{\hat{\theta}_n}^{-\frac{1}{2}} z z \in K_{1-\alpha}\}$$

dır. $\hat{\theta}_n$ ve $\hat{\Sigma}_{\hat{\theta}_n}$ tahmin edicilerinin bootstrap karşılıkları $\hat{\theta}_n^b$ ve $\hat{\Sigma}_{\hat{\theta}_n^b}$, $b = 1, 2, 3, \dots, B$ olmak üzere

$$Z_n^b = \sqrt{n} \hat{\Sigma}_{\hat{\theta}_n^b}^{-\frac{1}{2}} (\hat{\theta}_n^b - \hat{\theta}_n)$$

değerlerine dayalı

$$\{\hat{\theta}_n^b : Z_n^b \in K_{1-\alpha}, b = 1, 2, 3, \dots, B\}$$

güven bölgesi θ parametre vektörü için bir bootstrap güven bölgesidir. Bu güven bölgesi çok değişkenli standart normal dağılımın küresel bir dağılım olması düşüncesi altında, bu dağılımın $1-\alpha$ 'lık kısmını oluşturan $K_{1-\alpha}$ küresinin içine dönüştürülen $\hat{\theta}_n^b$ bootstrap tahminleri ile oluşturulmuştur.

$\hat{\theta}_n$ parametre vektörü bir boyutlu olduğunda yukarıda anlatıldığı gibi $\hat{\theta}_n^b$ bootstrap tahminleri veya bunların dönüşümleri olan Z_n^b değerleri küçükten büyüğe sıralanıp sıra istatistiklerine dayalı olan (2)'deki dengeli bootstrap güven aralığı oluşturulur. Bunu parametrenin vektör olması durumunda yapabilir miyiz? Parametre vektörünün $\hat{\theta}_n^b$ bootstrap tahminleri çok boyutlu Euclide uzayında noktalar olarak ele alınıp, bu noktaların ortalaması (ağırlıklı ortalaması) olan noktaya göre Euclide

uzaklıkları hesaplanır ve noktalar bu uzaklıklara göre sıralanırsa, en uzak noktalardan α kadarının atılmasından sonra geriye kalanlar yardımıyla güven bölgesi oluşturulabilir. $\hat{\theta}_n$ 'nin dağılımı küresel olmadığında böyle bir düşünce pek isabetli olmayabilecektir.

Çok değişkenli dağılımlarda sıra istatistikleri gibi bir kavrama ihtiyaç duyulduğu açıktır. Bu ihtiyaç bir ölçüde derinlik kavramıyla karşılanabilir. Sonraki kısımda bu kavram kısaca tanıtilip parametre vektörleri için derinliklere dayalı bootstrap güven bölgelerinin oluşturulmasında kullanılacaktır.

3.DERİNLİK ÖLÇÜLERİ

$x \in R^d$ verilen bir nokta ve F , d boyutlu X_i rasgele vektörünün R^d de tanımlı dağılım fonksiyonu olmak üzere, derinlik kavramı x noktasının F 'nin "merkezine" yakınlığını ölçmenin bir yoludur. Bunun örneklem karşılığı, $x \in R^d$ noktasının X_1, X_2, \dots, X_n rasgele örneğine ait gözlem kümesinin (bulutunun) merkezine yakınlığının ölçüsü olarak ifade edilebilir. Tek boyutlu rasgele değişkenlere ilişkin sıra istatistiği tanımında herbir rasgele değişken için küçükten büyüğe ya da büyükten küçüğe sıralama sözkonusu olabilmekte iken rasgele vektörler için böyle bir sıralama yapılamamaktadır. Rasgele vektörlerde bir sıralama derinlik kavramı ile en derinde(merkezde) yer alan vektörden en dışta yer alan vektörlere doğru yapılabilmektedir. Derinlik ölçmek için değişik ölçüler tanımlanmıştır. Bunlar için diğerlerinin yanında Liu, Parelius ve Singh(1999), Liu(1990), Rousseeuw ve Ruts(1999)'un çalışmaları örnek gösterilebilir.

Bu yazımızda derinlik ölçülerinden sadece yarı uzay derinliği(half-space depth, HD) üzerinde durulacaktır. Affin değişmezlik, merkezde en büyük derinlik, en derin noktadan uzaklaştıkça monoton olarak azalan derinlik ve sonsuzda sıfırlanan derinlik özelliklerine sahip olan bu derinlik ölçüsünün diğer özellikleri hakkında bilgi Danoho ve Gasko(1982) ile Zuo ve Serfling(2000)'de bulunabilir.

$x \in R^d$ noktasının F dağılımına göre yarı uzay derinliği

$$HD(F; x) = \inf_H \{P(H) : H, R^d \text{ de } x' \text{ i içeren kapalı bir yarı hiperdüzlem}\}$$

olarak tanımlanır. Yarı uzay derinliğinin örnek karşılığı

$$\hat{HD}(F; x) = \inf_H \left\{ \frac{n(X_i; X_i \in H)}{n} ; H, R^d \text{ de } x' \text{ i içeren kapalı bir yarı hiperdüzlem} \right\}$$

biçimindedir. Burada $n(A)$, A kümesinin eleman sayısını göstermektedir.

Derinlik ölçülerinde derinlik değerleri $[0, 1]$ aralığındadır. Derinliği en büyük olan noktaya derinlik merkezi ya da kısaca merkez denir. En büyük derinliğe sahip birdençok nokta bulunduğu bunların ortalaması merkez olarak alınmaktadır. Gözlemlerin örnek derinlik değerleri gözlemleri sıralamak için doğal bir araçtır. Bu

sıralamada derinlik değeri en büyük olan gözlem 1. sıraya konulmakta, sıra numarasının büyümesi o gözlemin merkezden dışarıya doğru uzaklaşması anlamındadır. Aynı derinlik değerine sahip birden çok gözlem bulunduğu tek değişkenli rasgele değişkenlerde olduğu gibi bunlara aynı sıra numarası verilmeyip ayrı sıra numarası verilmektedir; kaç tane gözlem varsa o kadar sıra numarası vardır. X_1, X_2, \dots, X_n örnekleminin derinliklerine göre sıralanmış $X_{[1]}, X_{[2]}, \dots, X_{[n]}$ 'lere derinlik sıra istatistikleri denir.

Rasgele örneklemin alındığı kitlenin eliptik bir dağılıma sahip olması durumunda dağılıma ait, varsa ortak yoğunluk fonksiyonunun konturları ile derinlik konturları örtüşecektir, Liu, Parelius ve Singh(1999).

Şimdi, ikinci kısımda sözü edilen $Z_n^b, b=1,2,3,\dots,B$ bootstrap gerçekleştirmelerini gözönüne alalım. Bunlar p boyutlu Euclide uzayında birer nokta olarak düşünülün ve örnek yarı uzay derinliğine göre sıralansın. Büyük derinlikli olanından küçüğüne doğru sıralanan Z_n^b 'ler $Z_{n:1}^b, Z_{n:2}^b, Z_{n:3}^b, \dots, Z_{n:B}^b$ olmak üzere bunlardan en derinde olan $1-\alpha$ 'lık kısmını oluşturanlar

$$BW_{n,1-\alpha} = \{Z_{n:1}^b, Z_{n:2}^b, Z_{n:3}^b, \dots, Z_{n:[(1-\alpha)B]}^b\} \quad (3)$$

olsun. Burada,

$$[(1-\alpha)B] = \begin{cases} (1-\alpha)B & , (1-\alpha)B \text{ tamsay} \\ [(1-\alpha)B]+1 & , (1-\alpha)B \text{ tamsay} \text{ de} \end{cases}$$

olup, $[\cdot]$ sayının tam kısmını göstermektedir. $BW_{n,1-\alpha}$ kümesine dayalı olarak oluşturulan

$$BGB(\theta) = \{\hat{\theta}_n - n^{-1/2} \sum_{w \in BW_{n,1-\alpha}} \frac{1}{2} w\} \quad (4)$$

kümesini kapsayan en küçük konveks bölge θ parametre vektörü için yaklaşık $1-\alpha$ güven düzeyinde derinliklere dayalı dengeli bir bootstrap güven bölgesi (BGB) olarak tanımlanır. Bu bootstrap güven bölgesi n örneklem hacmine bağlı olarak tam $1-\alpha$ güven bölgesine hemen hemen her yerde yakınsaması Yeh ve Singh(1997) tarafından ispatlanmıştır.

4. LİNEER MODELLERDE β PARAMETRE VEKTÖRÜ İÇİN DERİNLİKLERE DAYALI BOOTSTRAP GÜVEN BÖLGESİ

$Y = X\beta + \varepsilon$ lineer modelinde $\varepsilon \sim N(0, \sigma^2 I_{n \times n})$ olması durumunda β parametre vektörünün ençok olabilirlik tahmin edicisi $\hat{\beta}_n = (X'X)^{-1} X'Y$ için

$$\hat{\beta}_n \sim N(\beta, \sigma^2 (X'X)^{-1})$$

dir. Ancak hata vektörü ε 'nün dağılımı bilinmediğinde β için $1-\alpha$ güven düzeyinde

Yeh ve Singh(1997)'in elde ettikleri doğrulukta (hassasiyette) BGB oluşturmak istenirse ε_i rasgele değişkenlerinin dördüncü dereceden $E\varepsilon_i^4$ beklenen değerinin var olması gerekecektir. Bu saptama Qumsiyeh(1994) Teorem 3.3 ve Hatırlatma 3.4'e dayanarak yapılmıştır.

Bu varsayım yapıldığında $\hat{\beta}_n$, β 'nin enküçük kareler tahmin edicisi olmak üzere

$$\sqrt{n}\hat{\Sigma}_{\hat{\beta}_n}^{-\frac{1}{2}}(\hat{\beta}_n - \beta) \xrightarrow{d} Z$$

olması için gerekli olan üçüncü dereceden $E\varepsilon_i^3$ beklenen değerinin var olması koşulu da yerine gelmektedir(Lehmann(1999) Teorem 2.7.3, Teorem 2.5.2 ve Teorem 5.4.6).

Bu ve önceki kısımlarda anlatılanları kullanarak β parametre vektörü için $1-\alpha$ güven düzeyinde BGB oluşturmak istensin. Bir bootstrap tahmini elde etmek için bootstrap örnekleminin nasıl elde edildiğini özetleyelim(Efron(1982)):

1. n çaplı rasgele örneklem kullanılarak $Y = X\beta + \varepsilon$ lineer modelindeki β parametre vektörünün $\hat{\beta}_n$ enküçük kareler tahmin edicisini kullanarak artıklar tahmin edilir ve artık terimleri $e_i = y_i - X\hat{\beta}_n$ için örneklem dağılım fonksiyonu F_n oluşturulur.
2. F_n dağılımından n birimlik bir örneklem $e_1^b, e_2^b, e_3^b, \dots, e_n^b$ olmak üzere b . bootstrap verisi

$$y_i^b = X\hat{\beta}_n + e_i^b, i=1,2,3,\dots,n$$

elde edilir ve enküçük kareler tahmin edicisi ile $\hat{\beta}_n^b$ bootstrap tahmini elde edilir.

3. İkinci adım birbirlerinden bağımsız olarak B defa tekrar edilerek $\hat{\beta}_n^1, \hat{\beta}_n^2, \hat{\beta}_n^3, \dots, \hat{\beta}_n^B$ bootstrap örneklemleri elde edilir.

Burada ilk adım için Qumsiyeh(1994), örneklem dağılım fonksiyonu F_n 'nin e_i kalıntıları yerine $\hat{\mu} = n^{-1} \sum_{i=1}^n e_i$ olmak üzere $\hat{\varepsilon}_i = e_i - \hat{\mu}$ için oluşturulmasının daha uygun olacağını ileri sürmektedir. Bu durumda F_n dağılım fonksiyonu altında beklenen değer $E_n \hat{\varepsilon}_i = 0$ olacaktır.

Şimdi yarı uzay derinliğine dayalı yaklaşık $1-\alpha$ güven düzeyinde dengeli bootstrap güven bölgesi ifade edilecektir. $\hat{\beta}_n$ enküçük kareler tahmin edicisinin varyans kovaryans matrisinin tahmin edicisi $\hat{\Sigma}_{\hat{\beta}_n}$ ve bunun bootstrap karşılığı $\hat{\Sigma}_{\hat{\beta}_n^b}$ ile gösterilsin.

$$\sqrt{n}\hat{\Sigma}_{\hat{\beta}_n}^{-\frac{1}{2}}(\hat{\beta}_n - \beta) \xrightarrow{d} Z$$

olmak üzere,

$$Z_n^b = \sqrt{n}\hat{\Sigma}_{\hat{\beta}_n^b}^{-\frac{1}{2}}(\hat{\beta}_n^b - \hat{\beta}_n), b=1,2,3,\dots,B \quad (5)$$

bootstrap gerçekleştirmeleri yarı uzay derinliğine göre büyük derinlikli olanından küçük derinlikli olanına doğru $Z_{n:1}^b, Z_{n:2}^b, Z_{n:3}^b, \dots, Z_{n:B}^b$ olarak sıralansın. Bunların en derinde olan $1-\alpha$ 'lık kısmını oluşturanlar (3)'deki gibi $BW_{n,1-\alpha}$ kümesi ile gösterilsin. $BW_{n,1-\alpha}$ kümesine dayalı olarak (4)'deki gibi

$$BGB(\beta) = \left\{ \hat{\beta}_n - n^{-\frac{1}{2}} \hat{\Sigma}_{\hat{\beta}_n}^{-\frac{1}{2}} w : w \in BW_{n,1-\alpha} \right\} \quad (6)$$

kümesini kapsayan en küçük konveks bölge β parametre vektörü için $1-\alpha$ güven düzeyli derinliklere dayalı dengeli bootstrap güven bölgesini oluşturmuş olur.

5. UYGULAMA

Bu kısımda, β parametre vektörü için 4. kısımda ifade edilen derinliklere dayalı dengeli bootstrap güven bölgeleri rasgele üretilmiş veriler kullanılarak örneklenecektir. Veriler 4. kısımda sözü edilen varsayımların sağlanması ve karşılaştırma yapmak için elimizde mevcut bir güven bölgesinin bulunması amacıyla normal dağılımdan üretilmiştir. Örneklem yarı uzay derinlik hesaplamaları Rousseeuw ve Ruts(1996)'un iki boyutlu verilerde simpleks ve yarı uzay derinliklerini hesaplayan Fortran 77 ile yazılmış bilgisayar programı ile yapılmıştır. Uygulamaya konu olan

$$Y_{30 \times 1} = X_{30 \times 2} \beta + \varepsilon_{30 \times 1}$$

sabit terimsiz lineer modelinde, parametre vektörü $\beta' = [2, 3]$, $\varepsilon_{30 \times 1} \sim N(0_{30 \times 1}, 4I_{30 \times 30})$ ve $X_{30 \times 2}$ matrisi

$$X'_{30 \times 2} = \begin{bmatrix} -2 & -3 & -3 & -2 & -2 & -1 & -1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 3 & 2 & 1 & 1 & 0 & -1 \\ 1 & -1 & -2 & -1 & -2 & 0 & 1 & 2 & 2 & 1 & -1 & -2 & 0 & -1 & -2 & 3 & 2 & 1 & 0 & -1 & 0 & 1 & 2 & 3 & 1 & 1 & -1 & 1 & 0 & -1 \end{bmatrix}$$

olarak tasarlanmıştır. Modele uygun rasgele veriler üretildiğinde Y rasgele vektörüne ait 30 gözlem değeri satır sırasıyla aşağıda olduğu gibi kaydedilmiştir:

0.2256	-5.0870	-7.4673	-7.7479	-5.5239	-2.3192	-0.4066	5.1270	5.8994	5.3272
-3.6824	-11.1002	-6.0582	-1.9188	-8.0180	12.8161	11.1646	3.0418	4.0158	-0.6830
2.8261	10.1482	8.9667	15.4556	12.1678	2.8221	4.8991	7.7122	2.1001	-6.5343

Bu verilerle parametre vektörünün en küçük kareler tahmini, $\hat{\beta}'_n = [1.7537, 3.2309]$ ve $\hat{\sigma}^2 = 8.2996$ olup

$$\hat{Y}_i = 1.7537 X_{1i} + 3.2309 X_{2i}, \quad i = 1, 2, 3, \dots, 30$$

elde edilmiş ve artıklar $e_i = Y_i - \hat{Y}_i$ aşağıdaki gibi bulunmuştur:

0.5021	3.4050	4.2556	-1.0096	4.4453	-0.5655	-1.8838	0.4189	-0.5624	2.0963
1.3022	-2.8847	-6.0582	1.3121	-1.5562	1.3697	2.9491	-1.9428	2.2621	0.7942
-0.6813	3.4099	-1.0025	2.2555	3.6758	-3.9162	6.3763	2.7276	2.1001	-1.5497

Kısım 4'te sunulan bootstrap yöntemi kullanılarak (5)'deki gibi elde edilen 200 gerçekleşmesi olan $z_{30}^1, z_{30}^2, z_{30}^3, \dots, z_{30}^{200}$ değerleri yarı uzay derinliğine göre $z_{30.1}^b, z_{30.2}^b, z_{30.3}^b, \dots, z_{30.200}^b$ olarak sıralanıp

$$BW_{30,0.90} = \{z_{30.1}^b, z_{30.2}^b, z_{30.3}^b, \dots, z_{30.180}^b\}$$

kümesi elde edilir.

$$\frac{1}{\sqrt{n}} \hat{\Sigma}^{\frac{1}{2}} \frac{1}{\hat{\beta}_n} = \frac{1}{\sqrt{n}} (\hat{\sigma}^2 (X'X)^{-1})^{\frac{1}{2}} = \begin{bmatrix} 0.06996 & -0.02046 \\ -0.02046 & 0.07315 \end{bmatrix}$$

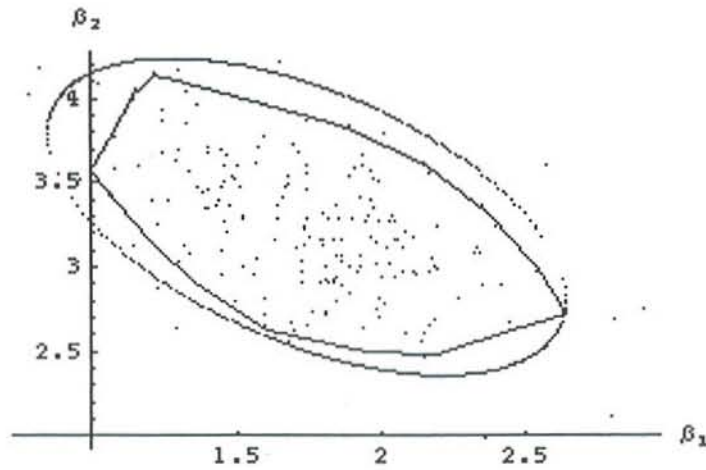
olmak üzere yaklaşık 0.90 güven düzeyli ve yarı uzay derinliğine dayalı bootstrap güven bölgesinin içinde bulunacak elemanlarının kümesi (6)' da olduğu gibi

$$BGB(\beta) = \left\{ \hat{\beta}_n - n^{-\frac{1}{2}} \hat{\Sigma}^{\frac{1}{2}} \frac{1}{\hat{\beta}_n} w : w \in BW_{30,0.90} \right\}$$

olarak bulunur. Elde edilen bu kümenin elemanlarını içeren en küçük konveks bölge bir çokgen olarak Şekil 1.'de sunulmuştur.

$BGB(\beta)$ kümesini içeren çokgen tam 0.90 güven düzeyli elips biçimli güven bölgesinin de içinde yer almıştır. Çokgenin dışında kalan 20 bootstrap tahmini 200 bootstrap tahmininin derinliklere göre derinliği en az olanlarıdır.

Elde edilen yaklaşık güven bölgesinin tam 0.90 güven düzeyli bölgenin içinde bir güven bölgesi olduğu dikkat çekmektedir. Bu durum ayrıca incelenmeli ve genellemede bulunup bulunulamayacağı araştırılmalıdır.



Şekil 1. β için dağılıma bağlı olarak elde edilen 0.90 güven düzeyli güven bölgesi ve yarı uzay derinliğine dayalı $BGB(\beta)$ kümesini içeren en küçük konveks çokgen ile çizilen yaklaşık 0.90 güven düzeyli güven bölgesi.

KAYNAKLAR

- DONOHU, L. DAVID and GASKO, M. (1992). Estimates Based on Halfspace Depth and Projected Outlyingness. *Ann. Statist.*, Vol. 20, No. 4, 1803-1827.
- EFRON, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans. SIAM, Philadelphia, PA, U.S.A
- GRAYBILL, F. A. (1976). Theory and Application of the Linear Model. Duxbury Press, North Scituate, Mass., U.S.A..
- HALL, P. (1992). The Bootstrap and Edgeworth Expansion. Springer-Verlag, New York, U.S.A..
- LEHMANN, E. L. (1999). Elements of Large-Sample Theory. Springer-Verlag, New York, U.S.A..
- LIU, R. Y. (1990). On a Notion of Data Depth Based on Random Simplices. *Ann. Statist.*, Vol. 18, No. 1, 405-414.
- LIU, REGINA Y., PARELIUS, JESSE M. and SINGH, K. (1999). Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference. *Ann. Statist.*, Vol. 27, No. 3, 783-858.
- QUMSIYEH, MAHER B. (1994). Bootstrapping and Empirical Edgeworth Expansions in Multiple Linear Regression Models. *Communications in Statistics: Theory and Methods*, 23(11), 3227-3239.
- ROUSSEEUW, PETER J. and RUTS, I. (1996). Bivariate Location Depth. *Appl. Statist.* Vol. 45, No.4, 516-526.
- ROUSSEEUW, PETER J. and RUTS, I. (1999). The Depth Function of a Population Distribution. *Metrika*, 9, 213-244.
- YEH, ARTHUR B. and SINGH, K. (1997). Balanced Confidence Regions Based on Tukey's Depth and the Bootstrap. *J. R. Statist. Soc. B*, Vol. 59, No. 3, 639-652.
- ZUO, Y. and SERFLING, R. (2000). Structural Properties and Convergence Results for Contours of Sample Statistical Depth Functions. *Ann. Statist.*, Vol. 28, No. 2, 483-499.

Balanced Bootstrap Confidence Regions in Linear Models Based on Half Space Depth

ABSTRACT

A balanced bootstrap confidence region has been studied and exemplified for a linear model parameters by data depth a notion which have become increasingly used in multivariate data analysis in recent years.

Key Words : Linear model, bootstrap, balanced confidence region, half space depth.

MAKALE İNDEKSİ

MAKALE İSMİ*	CİLT	NO	SAYFA	YIL
Ankara İlinde Gelir Farklılıklarını Belirleyen Etmenler	01	03	65	2002
Ardışık Örnekleme Planında Cusum Kontrol Kartlarının Kullanımı	01	01	87	2002
Basılı Türkçe'nin Önemli Bazı İstatistiksel Özellikleri	01	01	113	2002
Bayes Ağlarda Koşullu Bağımsızlıkların İncelenmesi Üzerine Bir Çalışma	02	01	89	2003
Benzetimde Girdi Analizi Yapan Otomatik bir Sistem	02	03	31	2003
Bernoulli Sayıları ve Lojistik Dağılım Fonksiyonu Üzerine	01	01	1	2002
Bir Yatırım Problemine Analitik Hiyerarşi Prosesi Yönteminin Uygulanması	01	01	39	2002
Çok Değişkenli Kalite Kontrol Yaklaşımlarının Bir Değerlendirmesi	01	01	67	2002
Çok Değişkenli Normal Dağılıma Sahip Örneklerdeki Aykırı Gözlemlerin Belirlenmesi İçin Bayesgil Bir Yaklaşım	02	01	11	2003
Damar Darlık Derecesi ve Risk Faktörlerinin Homojenite Analizi ile İncelenmesi	02	01	21	2003
Deney Tasarımında Kesirli Çoketkenli Tasarımlar ve Uygulamaları	01	01	11	2002
Eksik Blok Düzenlerinin Dual Yapıları	02	01	1	2003

MAKALE İSMİ*	CİLT	NO	SAYFA	YIL
Farklı Ortalama Vektörü ve Farklı Kovaryans Matrisi Koşullarında Dört Değişkenli Lojistik Regresyon Modeli ve Diskriminant Analizine Ait Doğru Sınıflandırma Olasılıklarının Simülasyon Tekniği Yardımıyla Karşılaştırılması	01	03	13	2002
Genelleştirilmiş T (Gt) Dağılımına Dayalı Regresyon Analizi	02	03	1	2003
Hotelling T 2 Testinde Kovaryans Matrislerinin Heterojenliğinin I. Tip Hata Üzerine Etkisi: Bir Simülasyon Çalışması	01	03	1	2002
İki ve Üç Yönlü Tabloların Gözelerinde 5'den Küçük Beklenen Frekans Olması Durumunda, I. Tip Hata Olasılığının Durumu	02	01	47	2003
Kanonik Korelasyon ve Açıklanmış Varyans Oranı İstatistiklerinin Başarı Düzeylerinin Tahminlenmesinde Kullanılması üzerine bir Araştırma	02	03	49	2003
Kardiyolojik Verilerin Doğrusal Olmayan Kanonik Korelasyon Analizi ile İncelenmesi	01	03	55	2002
Koşullu Gauss Dağılımı ve Etkileşimleri	02	03	11	2003
Küçük Alan Tahminlerinde Sentetik Kestiriciler ve Özellikleri	01	03	81	2002
Lineer Modellerde Yarı Uzay Derinliğine Dayalı Dengeli Bootstrap Güven Bölgeleri	02	03	63	2003
Matematiksel Finans	01	01	53	2002
Oransal ve Çarpımsal Tahmin Ediciler	01	01	101	2002
Ortogonal Düzenler	01	03	25	2002
Sürekli Değişken İçeren Grafiksiz Modeller	02	01	79	2003

MAKALE İSMİ*	CİLT	NO	SAYFA	YIL
Tiroit Bezi Verilerinin Bayes ve En Yakın K-Komşu Gibi Eğitici Yöntemlerle Sınıflanması	01	01	131	2002
Üç Yönlü Tablolarda İstatistiğinin Kullanılması	01	01	23	2002
Üçgensel Olumsuzluk Tablolarında Yarı-Bağımsızlık Modeli için Güç Analizi	02	03	21	2003
Varyansı Bilinmeyen Normal Dağılımlı Kitlenin Ortalamasının Ardışık Tahmini	01	03	35	2002
Yapay Sinir Ağı Performansına Etki Eden Faktörlerin Analizinde Taguchi Yöntemi:Hisse Senedi Fiyat Tahmini Uygulaması	02	01	29	2003
Yaşam Çözümlemesinde Yarışan Riskler ve Bir Uygulama	02	01	63	2003

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

YAZAR İNDEKSİ

YAZAR ADI	MAKALE İSMİ*	CİLT	NO	SAYFA	YIL
AKTAŞ Tülay	Serpil,SARAÇBAŞI Üçgensel Olumsuzluk Tablolarında Yarı-Bağımsızlık Modeli için Güç Analizi	02	03	21	2003
AKYOL Fikret	Mehmet, GÜRBÜZ İki ve Üç Yönlü Tabloların Gözelerinde 5'den Küçük Beklenen Frekans Olması Durumunda, I. Tip Hata Olasılığının Durumu	02	01	47	2003
AKYOL Fikret	Mehmet, GÜRBÜZ Üç Yönlü Tablolarda İstatistiğinin Kullanılması	01	01	23	2002
ALBAYRAK Songül	Tiroit Bezi Verilerinin Bayes ve En Yakın K-Komşu Gibi Eğitici Yöntemlerle Sınıflanması	01	01	131	2002
BAKIR Sezar	M. Akif, KARACA Çok Değişkenli Kalite Kontrol Yaklaşımlarının Bir Değerlendirmesi	01	01	67	2002
BAYRAK Hülya	Hülya Eksik Blok Düzenlerinin Dual Yapıları	02	01	1	2003
BAYRAK Fikri	Hülya, GÖKPINAR Koşullu Gauss Dağılımı ve Etkileşimleri	02	03	11	2003
BAYRAK Aslıhan	Hülya, ALHAN Ortogonal Düzenler	01	03	25	2002
BAYRAK Fikri ,	Hülya, GÖKPINAR Sürekli Değişken İçeren ÖZKAYA Berrin Grafiksel Modeller	02	01	79	2003
DALKILIÇ DALKILIÇ	Mehmet E., Basılı Türkçe'nin Önemli Bazı Gökhan İstatistiksel Özellikleri	01	01	113	2002
DÜNDAR Samim	Samim Bernoulli Sayıları ve Lojistik Dağılım Fonksiyonu Üzerine	01	01	1	2002
EKİZ Ufuk	Ufuk Çok Değişkenli Normal Dağılıma Sahip Örneklerdeki Aykırı Gözlemlerin Belirlenmesi İçin Bayesgil Bir Yaklaşım	02	01	11	2003
GENÇ Olçay	Ali İhsan, ARSLAN Genelleştirilmiş T (Gt) Olçay Dağılımına Dayalı Regresyon Analizi	02	03	1	2003

YAZAR ADI	MAKALE İSMİ*	CİLT	NO	SAYFA	YIL
HAMURKAROĞLU BACANLI Sevil	Canan, Ardışık Örnekleme Planında Cusum Kontrol Kartlarının Kullanımı	01	01	87	2002
İŞÇİ Öznur	Kanonik Korelasyon ve Açıklanmış Varyans Oranı İstatistiklerinin Başarı Düzeylerinin Tahminlenmesinde Kullanılması üzerine bir Araştırma	02	03	49	2003
KADILAR Cem, MULUK Zehra	Deney Tasarımında Kesirli Çoketkenli Tasarımlar ve Uygulamaları	01	01	11	2002
KADILAR Cem, Hülya ÇINGİ	Oransal ve Çarpımsal Tahmin Ediciler	01	01	101	2002
KANIK E. ÇAMDEVİREN GÜRBÜZ Fikret	Arzu, Farklı Ortalama Vektörü ve Handan , Farklı Kovaryans Matrisi , Koşullarında Dört Değişkenli Lojistik Regresyon Modeli ve Diskriminant Analizine Ait Doğru Sınıflandırma Olasılıklarının Simülasyon Tekniği Yardımıyla Karşılaştırılması	01	03	13	2002
Kanık E. Arzu, ÇAMDEVİREN Handan , GÜRBÜZ Fikret	Hotelling T 2 Testinde Kovaryans Matrislerinin Heterojenliğinin I. Tip Hata Üzerine Etkisi: Bir Simülasyon Çalışması	01	03	1	2002
KARABULUT İhsan, ÖZTÜRK Fikri	Lineer Modellerde Yarı Uzay Derinliğine Dayalı Dengeli Bootstrap Güven Bölgeleri	02	03	63	2003
KOÇBERBER Güler, Hülya ÇINGİ	Varyansı Bilinmeyen Normal Dağılımlı Kitlenin Ortalamasının Ardışık Tahmini	01	03	35	2002
OLMUŞ Hülya, ERBAŞ Semra Oral	Bayes Ağlarda Koşullu Bağımsızlıkların İncelenmesi Üzerine Bir Çalışma	02	01	89	2003
ÖNALAN Ömer	Matematiksel Finans	01	01	53	2002

YAZAR ADI	MAKALE İSMİ*	CİLT	NO	SAYFA	YIL
ÖZALP Alperen, ANAGÜN A. Sermet	A. Yapay Sinir Ağı Performansına Etki Eden Faktörlerin Analizinde Taguchi Yöntemi:Hisse Senedi Fiyat Tahmini Uygulaması	02	01	29	2003
SELİM Sibel	Ankara İlinde Gelir Farklılıklarını Belirleyen Etmenler	01	03	65	2002
SERTKAYA Durdu, M. Tekin	SÖZER Yaşam Çözümlemesinde Yarışan Riskler ve Bir Uygulama	02	01	63	2003
SEVİNÇ Volkan	Küçük Alan Tahminlerinde Sentetik Kestiriciler ve Özellikleri	01	03	81	2002
SÜT Necdet, TÜRE Mevlüt	Damar Darlık Derecesi ve Risk Faktörlerinin Homojenite Analizi ile İncelenmesi	02	01	21	2003
TANIL Halil,BATMAZ İnci	Benzetimde Girdi Analizi Yapan Otomatik bir Sistem	02	03	31	2003
TÜRE Mevlüt, SÜT Necdet , KÜRÜM Turhan , GÜLTAÇ	Kardiyolojik Verilerin Doğrusal Olmayan Kanonik Korelasyon Analizi ile İncelenmesi	01	03	55	2002
UYAR Yavuz, ERCÜMENT N., KURT Mustafa	DİZDAR Bir Yatırım Problemine Analitik Hiyerarşi Prosesi Yönteminin Uygulanması	01	01	39	2002

* 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 başlığı metne uygun olmalıdır. Araştırma başlığı ortalı, her sözcüğün ilk harfi büyük 14 punto harf büyüklüğünde, 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 ayraçlarla yapılmalıdır. Ayraç 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 belirtilmeli; 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].Erişim Tarihi: 22.12.1996

- Araştırmanın ingilizce dilde özeti araştırmanın sonunda verilmelidir. Araştırmanın ingilizce 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 ingilizce 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 harfi ayırt edilebilmelidir. Çoklu indislerden sakınılmalıdır.
- Denklemler word, standart (default) ölçülerde 1 tab (1,27 cm) içerden ve numara vermek gerekiyorsa 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 ayraçlar aynı formülde yer aldığı anda, sıra düzeni örneğin $\{[(0)]\}$ biçiminde olmalıdır.

10. Araştırmanın Türkçe yazım kurallarına uygun olması yazarın sorumluluğu altındadır.

DUYURU PANOSU

2004 TÜRKİYE İKTİSAT KONGRESİ

“2004 Türkiye İktisat Kongresi” 5-9 Mayıs 2004 tarihlerinde İzmir’de toplanıyor. Ayrıntılı bilgi için bakınız: www.tik2004.org.tr

KONFERANS DUYURUSU

“Cumhuriyetin 80.Yılı Kutlama Programı” kapsamında Devlet İstatistik Enstitüsü tarafından hazırlanan programda 10 Mayıs 2004 tarihinde “80’inci Yılda Türk İstatistik Sistemi” konulu konferans yer almaktadır.

İSTATİSTİK GÜNLERİ 2004 SEMPOZYUMU

Dokuz Eylül Üniversitesi, Fen-Edebiyat Fakültesi İstatistik Bölümü tarafından düzenlenen “İstatistik Günleri 2004 Sempozyumu” 20-21 Mayıs 2004 tarihlerinde İzmir’de gerçekleştirilecektir. Ayrıntılı bilgi için bakınız: <http://igs2004.deu.edu.tr>

KANADA İSTATİSTİK DERNEĞİ YILLIK TOPLANTISI

30 Mayıs-02 Haziran 2004, Montreal, Quebec, Kanada.

Ayrıntılı bilgi için: Christian LEGER,
Departement de Mathematiques et de Statistique, Universite de Montreal
C.P. 6128, succursale Centre-ville, Montreal-Quebec/Canada H3C 3J7
e-mail: leger@dms.umontreal.ca

IASE 2004 YUVARLAKMASA TOPLANTISI

28 Haziran – 03 Temmuz 2004, Lund Üniversitesi, Lund / İsveç
Ayrıntılı bilgi için bakınız: http://www.hobbes.lite.msu.edu/-IASE_2004_Roundtable
<http://www.cbs.nl/isi/iase.htm>

ULUSLARARASI BİOMETRİ KONFERANSI 2004

11-16 Temmuz 2004, Queensland / Avustralya.
Ayrıntılı bilgi için bakınız: <http://www.ozaccomm.com.au/ibc2004>

İSTATİSTİK KUTLAMASI

Sir David Cox’un 80. doğumyılı onuruna Neuchatel Üniversitesi tarafından 14-18 Temmuz 2004 tarihleri arasında uluslararası konferans düzenlenmiştir.

Ayrıntılı bilgi için: Prof.Yadolah Dodge,
Statistics Group, Universite de Neuchatel

Espace de l'Europa 4, P.O. Box 805, 2002 Neuchatel / Switzerland
e-mail: Yadolah.Dodge@unine.ch

**BERNOULLI DERNEĞİ ve MATEMATİKSEL İSTATİSTİK ENSTİTÜSÜ
6. ORTAK TOPLANTISI**

26-31 Temmuz 2004, Barselona / İspanya.
Ayrıntılı bilgi için bakınız: <http://www.imub.ub.es/events/wc2004/>

AMERİKAN İSTATİSTİK DERNEĞİ ORTAK İSTATİSTİK TOPLANTISI

08-12 Ağustos 2004,
Ayrıntılı bilgi için bakınız : www.amstat.org/meetings

ULUSLARARASI KLİNİK BİOİSTATİSTİK DENEĞİ 25. YILLIK TOPLANTISI

15-19 Ağustos 2004, Leiden / Hollanda.
Ayrıntılı bilgi için bakınız : <http://iscb2004.clinicalresearch.nl>
e-mail : iscb2004@leidscongresbureau.nl

COMPSTAT 2004

23-27 Ağustos 2004, Prag / Çek Cumhuriyeti.
Ayrıntılı bilgi için bakınız : <http://compstat2004.cuni.cz>
e-mail : compstat2004@cuni.cz

ISI 55. DÖNEM TOPLANTISI

05-12 Nisan 2005, Sydney / Avustralya.
Ayrıntılı bilgi için : ISI Permanent Office, Prinses Beatrixlaan 428, P.O. Box 950
2270 AZ Voorburg / The Netherlands
e-mail : isi@cbs.nl
<http://www.cbs.nl/isi>

KANADA İSTATİSTİK DERNEĞİ YILLIK TOPLANTISI

12-15 Haziran 2005, Kanada.
Ayrıntılı bilgi için: Mik BICKIS,
Mathematical Sciences Group, University of Saskatchewan Room142
McLean Hall
106 Wiggins Road, Saskatoon, SK / Canada S7N 5E6
e-mail: bickis@math.usask.ca

Genelleştirilmiş T (Gt) Dağılımına Dayalı Regresyon Analizi

A Regression Analysis Based on the Generalized T (GT) Distribution with Known Shape Parameters

Ali İhsan GENÇ
Olca ARSLAN

1

Koşullu Gauss Dağılımı ve Etkileşimleri Dağılımı ve Etkileşimleri

Conditional Gaussian Distribution and Interactions

Hülya BAYRAK
Fikri GÖKPINAR

11

Üçgensel Olumsuzluk Tablolarında Yarı-Bağımsızlık Modeli için Güç Analizi

Power Analysis of Quasi-Independence Model for Triangular Contingency Tables

Serpil AKTAŞ
Tülay SARAÇBAŞI

21

Benzetimde Girdi Analizi Yapan Otomatik bir Sistem

An Automatic Input Analyzer for Simulation

Halil TANIL
İnci BATMAZ

31

Kanonik Korelasyon ve Açıklanmış Varyans Oranı İstatistiklerinin Başarı Düzeylerinin Tahminlenmesinde Kullanılması Üzerine bir Araştırma

A Research on Using Canonical Correlation and Redundancy Analysis Statistics to Estimate Accomplishment Levels

Öznur İŞÇİ

49

Lineer Modellerde Yarı Uzay Derinliğine Dayalı Dengeli Bootstrap Güven Bölgeleri

Balanced Bootstrap Confidence Regions in Linear Models Based on Half Space Depth

İhsan KARABULUT
Fikri ÖZTÜRK

63