



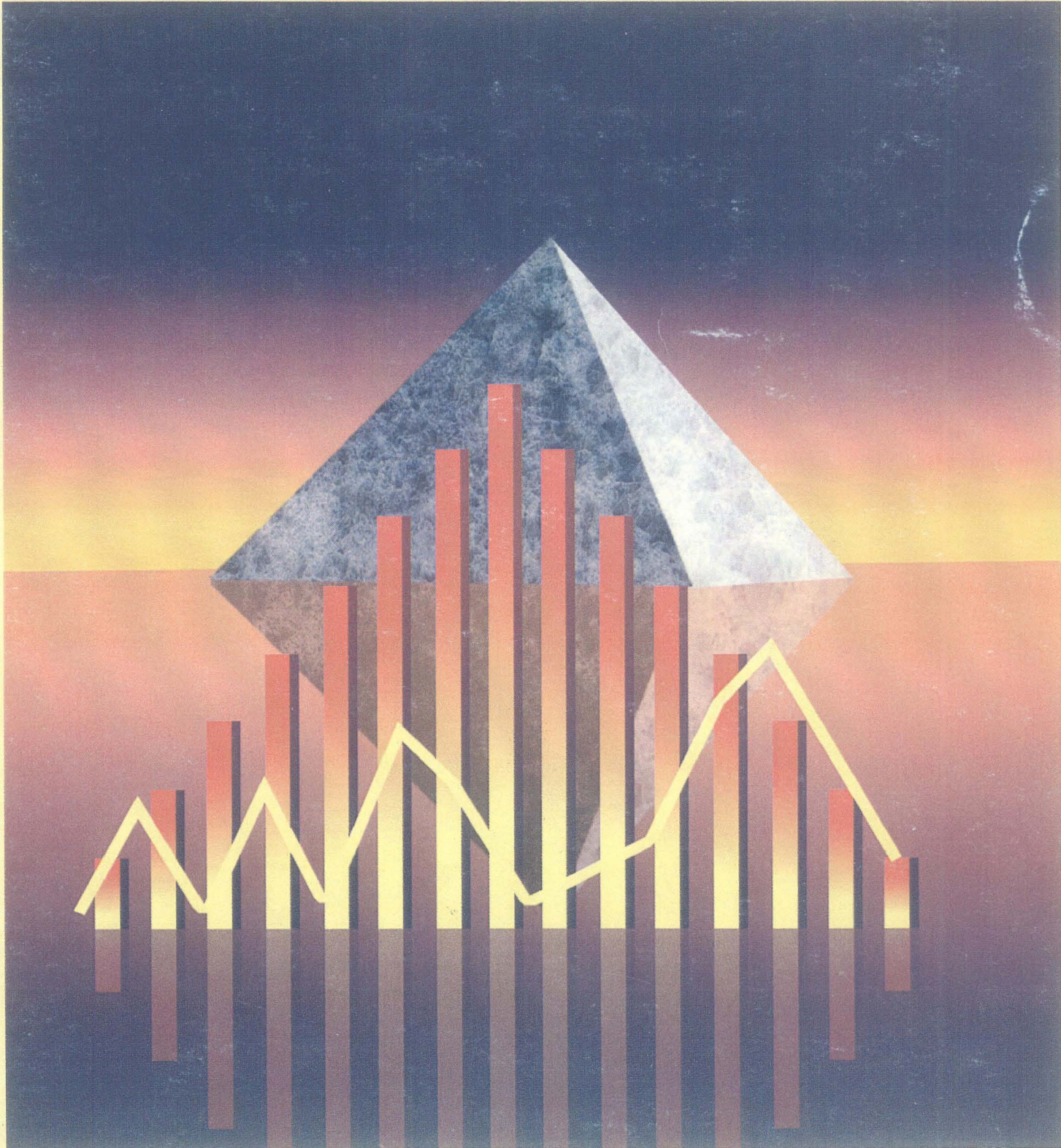
İstatistik Araştırma

Journal of Statistical Research *Dergisi*

Cilt 02 *Volume*

No 01 *Number*

Nisan 2003 *April*



JOURNAL OF STATISTICAL RESEARCH

Owner

On Behalf Of The State Institute of Statistics

Ömer DEMİR

President, The State Institute of Statistics

Executive Editör

M.Akif BAKIR

Editorial Board

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

Soner GÖNEN (Gazi University)
Ceyhan İNAL (Hacettepe University)
Öztaş AYHAN (METU)
Alptekin ESİN (Gazi University)
Zehra MULUK (Başkent University)
Ergün KARAAĞAOĞLU (Hacettepe University)
Ömer ESENSOY (Hacettepe University)
Bedriye SARAÇOĞLU (Gazi University)
İmdat KARA (Başkent University)
Aykut TOROS (Hacettepe University)
Ali YAZICI (Atılım University)

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

Mehmet AKYOL - Gönül ERDEM - Atalay BİÇYAP

Journal of Statistics Research Secretary Address State Institute of Statistics Training Center	Tel: +90 312 417 64 40 / 390 - 391 Faks: +90 312 425 35 85 e-mail: dergi@die.gov.tr URL: http://www.die.gov.tr
Request and Subscription Address State Institute of Statistics Revolving Fund Administration Necatibey Street, No: 114 06100 Yücecepe / ANKARA	Tel: +90 312 417 64 40 / 323 - 319 Fax: +90 312 417 58 86

State Institute of Statistics, Printing Division, April 2003
MTB: 2003 - 0609 - 300 Copies

DEVLET İSTATİSTİK ENSTİTÜSÜ

KÜTÜPHANE VE DOKÜMANTASYON MERKEZİ

GENEL EDİTÖR NOTU

Devlet İstatistik Enstitüsü, istatistiki arařtırmaların niteliğinin yükseltilmesi, istatistik metodolojisi ve uygulamasının geliştirilmesi, literatürde yer alan çalışmaların tartışılması, istatistik uygulamalarıyla ilgili anket çalışmalarının ele alınması, teorik ve uygulamalı alanlardaki arařtırmacılar arasındaki iletişimin ortak çalışmalar ve yayınlarla güçlendirilmesi amacıyla İstatistik Arařtırma Dergisi adı altında yılda üç defa Nisan, Ağustos ve Aralık aylarında hakemli bir dergiyi yayın hayatına kazandırmıştır.

İstatistik arařtırma Dergisi, dergiye gelen arařtırmalardan Nisan ve Aralık aylarında, İstatistik Arařtırma Sempozyumu'nda sunulan bidirilerin hakemlik süreci sonucunda Ağustos ayında özel sayı olarak yayınlanmaktadır. Yayın hayatında 1 yılını doldurmuş olan dergimiz hakkında okuyucularımızdan aldığımız olumlu tepkiler bize gurur vermektedir.

Nisan 2003 itibarı ile deęerlendirme sürecinde olan ve deęerlendirmesi tamamlanan çalışma sayısı 28'dir. Bunlardan 13'nin Hakemlik Süreci devam etmekte, 2'si hakemler tarafından istenen düzeltmelerin yapılması için arařtırma sahibinde, 5'i hakem belirlenmek üzere ilgili Alan Editörü'nde olup, 8 arařtırma Nisan sayımızda yayınlanmak üzere basıma uygun bulunmuştur.

Benden sonra Genel Editörlük görevini üstlenen sayın Prof. Dr. Fetih YILDIRIM'a başarılar diler, derginin yeni sayısını bilim ve istatistik dünyasına takdim etmekten onur duyar, bu ürünün ortaya çıkmasında emeęi geçen Dergi Sekreteryasına şükranlarımı ve özellikle her aşamada yer alarak desteklerini esirgemeyen üniversite öğretim üyelerine en derin saygılarımı sunarım.

Dr. M.Akif BAKIR
Genel Editör
İstatistik Arařtırma Dergisi

BU SAYIDAKİ MAKALELERİN HAKEMLERİ

- Prof.Dr. Zehra MÜLÜK – *Başkent Üniversitesi*
Prof.Dr. Ergun KARAAĞAOĞLU – *Hacettepe Üniversitesi*
Prof.Dr. Semra Oral ERBAŞ – *Gazi Üniversitesi*
Prof. Dr. M. Kemal YOĞURTÇUGİL – *Mimar Sinan Üniversitesi*
Prof. Dr. Nimetullah BURNAK – *Osman Gazi Üniversitesi*
Prof.Dr. Hüseyin TATLIDİL – *Hacettepe Üniversitesi*
Prof.Dr. Soner GÖNEN – *Gazi Üniversitesi*
Doç. Dr. Tülay SARAÇBAŞI – *Hacettepe Üniversitesi*
Doç. Dr. Cafer ÇELİK – *Atatürk Üniversitesi*
Doç. Dr. Reha ALPAR – *Hacettepe Üniversitesi*
Doç. Dr. Gül ERGÜN – *Hacettepe Üniversitesi*

Eksik Blok Düzenlerinin Dual Yapıları

Hülya BAYRAK*

ÖZET

Nesnelerin belirli kriterler altında düzenlenmesi problemi kombinatoriyel analizde önemli bir yer tutar. Böyle bir düzenleme isabet yapısı yada konfigürasyon olarak bilinir. İsbet yapılarının özel bir tipi tamamlanmamış blok düzenleridir.

Bir düzenin duali işlemleri ve blokları sırasıyla orijinal düzenin bloklarına ve işlemlerine karşılık gelen yeni bir düzendir.

Bu çalışmanın amacı bazı dual düzenlerin geometrik yapılarını incelemektedir. Bu dual düzenler tamamlanmamış blok düzenler sınıfındadır.

Anahtar Kelimeler : Dual Düzen, Dengeli Tamamlanmamış Blok Düzeni, Kalıntı Düzen, Projektif Geometri.

1. GİRİŞ

Bir taktikal konfigürasyon (tactical configuration) $TC(v, k, t, \lambda)$, v tane elemanın k genişliğe sahip bloklara yerleştirilmesidir. Bu yerleştirmede, farklı elemanlara sahip $t \times 1$ boyutlu sütun vektörleri (t -tuple) tam olarak λ blokta gözüktür. t ve λ sırasıyla, konfigürasyonun kuvveti ve indeksi olarak bilinir.

$t=2$ durumu "Dengeli Tamamlanmamış Blok Düzeni" (Balanced Incomplete Block Design-BIBD) ne karşılık gelir. Bu düzenlerde $v \geq 2$ işlem vardır. b , blok sayısı; k , blok genişliği; r , replikasyon sayısı; λ ($\lambda > 0$), işlemlerin birlikte görünme sayısı olmak üzere parametreler arasında $\lambda(v-1) = r(k-1)$ ilişkisi vardır.

Böyle bir düzende, A , isabet matrisi olmak üzere

$$\begin{aligned} A J_b &= r J_{v \times b} \\ J_v A &= k J_{v \times b} \end{aligned}$$

eşitlikleri sağlanır, burada J_i tüm elemanları 1 olan uygun boyutlu matristir. A^T gözönüne alındığında,

$$\begin{aligned} J_b A^T &= r J_{b \times v} \\ A^T J_v &= k J_{b \times v} \end{aligned}$$

* Gazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, Ankara, Türkiye

yazılabileceği aşıkardır. Yani A^T matrisi, b, v, k, r parametrelili bir blok düzeninin isabet matrisi olacaktır. Bir düzenin duali, işlemleri ve blokları, sırasıyla, orijinal düzenin, blokları ve işlemlerine karşılık gelen yeni bir düzendir.

2. EKSİK BLOK DÜZENLER VE DUALLERİ

Bir düzenin dual düzeni işlemleri ve blokları sırasıyla, orijinal düzenin blokları ve işlemlerine karşılık gelen yeni bir düzendir. BIBD'da $v \leq b$ Fisher eşitsizliği sağlanır. Ancak dual düzenlerde bloklar ve işlemlerin rolleri değiştiğinden Fisher eşitsizliği sağlanmayabilir. Netice olarak, dual düzen dengeli olmayabilir.

BIBD'de $v = b$ ise, düzen simetriktir denir (Simetrik BIBD).

$TC(v, k, k, \lambda)$ konfigürasyonunun duali k -sınıflı birliktelik yapısını tanımlar. Bu birliktelik yapıları geometrik özelliklerin terimleriyle karakterize edilebilir.

Bir BIBD, sonlu projektif geometri $PG(m, p^n)$ ve sonlu öklid geometri $EG(m, p^n)$ de uygun doğruların seçimi ile bir geometrik konfigürasyonla tanımlanabilir.

Çalışmada, $PG(m, p^n)$ 'nin doğruları ile düzenin blokları, $PG(m, p^n)$ 'nin noktaları ile düzenin işlem ve/veya işlem kombinasyonları eşleştirilmiştir.

$PG(m, p^n)$ 'nin doğruları bir uzay oluşturur. Her doğru bir bloğa karşılık geldiğinden, uzaya "bloklar uzayı" denebilir. Doğrular üzerindeki noktaların oluşturduğu uzay ile bloklar uzayı birbirinin dual uzayları olur.

Teorem 2.1: Bir BIBD'nin dualinin yine BIBD olabilmesi için gerek ve yeter koşul $v = b$ olmasıdır (Street and Street, 1987).

Shrikhande, $\lambda = 1$ veya $\lambda = 2$ durumu için asimetric BIBD'nin duallerinin kısmi dengeli tamamlanmamış blok düzeni olduğunu göstermiştir (Shrikhande, 1952).

Aynı sonuçlar, Shrikhande ve Bahagwandas tarafından graf teori ile ilişkilendirilerek yeniden ele alınmıştır (Shrikhande and Bahagwandas, 1965).

Teorem 2.2 : v, b, r, k, λ parametrelili asimetric BIBD'nin dual düzeni, aşağıdaki parametrelere sahip iki birliktelik sınıflı kısmi dengeli tamamlanmamış blok düzeni (Partially Balanced Incomplete Block Design-PBIBD) dir.

$$v^* = b, \quad b^* = v, \quad r^* = k, \quad n_1^* = k(r-1), \quad n_2^* = b-1-n_1$$

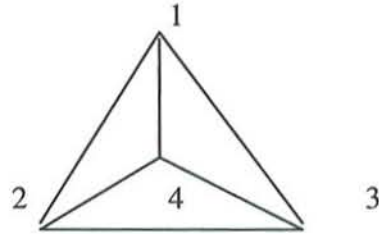
$$\lambda_1^* = 1, \quad \lambda_2^* = 0, \quad p_{11}^{1*} = r-2+(k-1)^2, \quad p_{11}^{2*} = k^2 \quad (\text{Raghavarao, 1971}).$$

Örneğin ; $v = 4, b = 6, k = 2, r = 3, \lambda = 1$ parametrelili bir BIBD'nin birliktelik yapısı aşağıdaki gibidir :

Tablo 1. $v=4, b=6, k=2, r=3, \lambda=1$ Parametrelı BIBD

BLOKLAR						
İŞLEMLER	1	2	3	4	5	6
1	x		x		x	
2	x			x		x
3		X	x			x
4		x		x	x	

Bu düzenin geometrik yapısı $GF(2)$ üzerine kurulan sonlu öklid geometrisi $EG(2, 2)$ dir.



Şekil 1. Tablo 1' deki Düzenin $EG(2,2)$ de Bir Gösterimi

Tablo 1 de verilen düzenin dual düzenine ait birliktelik yapısı aşağıdaki gibidir

Tablo 2. Dual Düzenin Birliktelik Yapısı

BLOKLAR				
İŞLEMLER	1	2	3	4
1	X	x		
2			x	x
3	X		x	
4		x		x
5	X			x
6		x	x	

$$\lambda_1^* = 1$$

(1, 3)(1, 4)(1, 5)(1, 6)
 (2, 3)(2, 4)(2, 5)(2, 6)
 (3, 1)(3, 2)(3, 5)(3, 6)
 (4, 1)(4, 2)(4, 5)(4, 6)
 (5, 1)(5, 2)(5, 3)(5, 4)
 (6, 1)(6, 2)(6, 3)(6, 4)

$$\lambda_2^* = 0$$

(1, 2)
 (2, 1)
 (3, 4)
 (4, 3)
 (5, 6)
 (6, 5)

Tablo 4 de verilen düzen $v^* = 6, b^* = 4, r^* = 2, k^* = 3, \lambda_1^* = 1, \lambda_2^* = 0, n_1^* = 4, n_2^*$

$= 1, p_{ij}^{1*} = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}, p_{ij}^{2*} = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$ parametrelerine PBIBD 'dir.

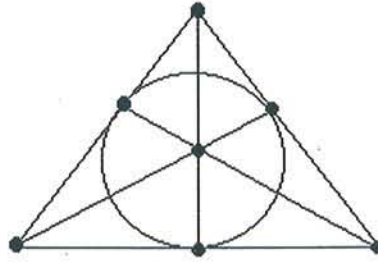
Tanım 2.1 : Bir G grafında birleştirilmiş olan iki köşenin her ikisi tam p_{11}^1 tane diğer köşelerle birleştirilmiş , fakat p_{11}^2 tanesi ile birleştirilmemiş ise, G grafi , $n_1, n_2, p_{11}^1, p_{11}^2$ parametrelili “Güçlü Düzgün Graf”(Strongly Regular Graph-SRG) adını alır (Bose, 1963).

Tanım 2.2 : SRG, aşağıdaki parametrelere sahip ise “Bağlantılı Blok Graf” (Linked Block graph-LBG) adını alır (Bose, 1963).

$$\begin{aligned} n_1 &= r(k-1) & , & & n_2 &= (k-r)(r-1)(k-1) / r \\ p_{11}^1 &= (r-1)^2 + k-2 & , & & p_{11}^2 &= r^2 \end{aligned}$$

Sonuç olarak , dual düzenin birliktelik yapısı LBG’a izomorftur.

$b=v=7, r=k=3, \lambda=1$ parametrelili Simetrik BIBD’i, $TC(7,3,2,1)$ Konfigurasyonudur. Bu konfigurasyon, geometrik olarak $GF(2)$ üzerine kurulan $PG(2, 2)$ ile tanımlanır. Bir Fano düzlemidir. Aynı zamanda Hadamard $2-(7, 3, 1)$ düzeni tanımlar (ve deficiency $d = v-r(k-1) - 1 = 0$ olduğundan Steiner sistemdir).



Şekil 2. $TC(7,3,2,1)$ Konfigurasyonunun $PG(2,2)$ de Bir Gösterimi

$TC(7,3,2,1)$ konfigurasyonunun tanımladığı Simetrik BIBD’nin, duali parametreleri aynı olan Simetrik BIBD dir. Dual düzenlerin izomorf olması gerekmez.

Eğer $\lambda=2$ ise simetrik $TC(v,k,2,2)$ düzenler “Biplane” tanımlarlar. $TC(7,3,2,1)$ konfigurasyonunun tamamlanışı olan $TC(7,3,2,2)$ konfigurasyonu bir Biplane dir (Bayrak ve Gönen, 2002).

3. LATİS DENEME DÜZENİ VE DUAL DÜZENLERİ

Kare Latis Düzeni’nde p^n tane işlem vardır. Bunlar , p genişlikli bloklara yerleştirilir. Düzen , p düzeyli n tane etkenin oluşturduğu faktöriyel deney düzeni gibi düşünülür.

Diğer taraftan çok sayıda işlem , p yerine k genişlikli ($k < p$) bloklara dağıtılarak bir çözüm yolu denenebilir. Bu durumda BIBD’ından faydalanılır.

Latis düzenlerin en önemli kısmını oluşturan homojen replikasyonları elde etmenin bir yolu Duallik kavramını kullanmak olabilir (Bayrak ve Gönen , 1998).

PG (2, 3) geometrisi gözönüne alındığında, PG(2, 3) deki doğrular ve üzerindeki noktalar aşağıdaki gibidir:

Tablo 3. PG(2, 3) deki Doğrular ve Üzerindeki Noktalar

Doğru	A	B	AB	AB ²	C	AC	AC ²	BC	BC ²	ABC	ABC ²	AB ² C	AB ² C ²
	100	010	110	1-10	001	101	10-1	011	01-1	111	11-1	1-11	1-1-1
[0,0,1]	X	X	X	X									
[0,1,0]	X				X	X	X						
[0,-1,1]	X							X		X			X
[0,1,1]	X								X		X	X	
[1,0,0]		X			X			X	X				
[1,0,-1]		X				X				X		X	
[1,0,1]		X					X				X		X
[1,-1,0]			X		X					X	X		
[-1,1,1]			X				X	X				X	
[1,-1,1]				X	X							X	X
[1,1,0]			X			X			X				X
[1,1,-1]				X		X		X			X		
[1,1,1]				X			X		X	X			

Bu geometride, paralel doğru gruplarından herbirinin kapsadığı işlem ya da işlem kombinasyonları homojen replikasyonları verir. Böylece 4 homojen replikasyon elde edilir. Bu homojen replikasyonlar veya 4 tane doğru demeti, düğüm noktalarının dual uzayını meydana getirir. Dolayısıyla, bir homojen replikasyon, geometrik anlamda düğüm noktası çıkarılmış doğrular demetidir. Homojen replikasyonlar öyle bloklar oluşturur ki bu bloklar EG(2,3) geometrisinin tüm özelliklerini sağlar. Böyle bir geometrik yaklaşımla, blok yapma keyfi ve yanlış olmaz. Düğüm noktaları çıkarılmış doğru demetlerinin paralellik özelliğinden, hangi bloğa hangi kombinasyonun gireceği bellidir. Homojen replikasyonlar aşağıdaki gibidir :

Homojen Replikasyon 1

(1 0 0)	(1 1 0)	(1-1 0)	A	AB	AB ²
(1 0-1)	(1 1-1)	(1-1-1)	AC ²	ABC ²	AB ² C ²
(1 0 1)	(1 1 1)	(1-1 1)	AC	ABC	AB ² C

Homojen Replikasyon 2

(1 0 0)	(1 0 1)	(1 0-1)	A	AC	AC ²
(1-1 1)	(1-1-1)	(1-1 0)	AB ² C	AB ² C ²	AB ²
(1 1 0)	(1 1 1)	(1 1-1)	AB	ABC	AB ² C

Homojen Replikasyon 3

(1 0 0)	(1 1 1)	(1-1-1)	A	ABC	AB ² C
(1-1 0)	(1 0 1)	(1 1-1)	AB ²	AC	ABC ²
(1 1 0)	(1 0-1)	(1-1 1)	AB	AC ²	AB ² C

Homojen Replikasyon 4

(1 1 1)	(1-1 0)	(1 0-1)	ABC	AB ²	AC ²
(1 0 0)	(1-1 1)	(1 1-1)	A	AB ² C	ABC ²
(1 1 0)	(1 0 1)	(1-1-1)	AB	AC	AB ² C ²

Tanım 3.1 : (s, r, μ) -net : \mathcal{T} , bir isabet yapısı olsun. B ve G , \mathcal{T} yapısının iki bloğu olmak üzere bu yapı üzerindeki paralellik , $B // G$, $B = G$, veya $|B, G| = 0$ olarak tanımlanır. \mathcal{T} yapısının (s, r, μ) -net oluşturabilmesi için $v = s^2\mu$ sayıda nokta , $b = sr$ sayıda blok olmalıdır, her paralel sınıfta s blok ve her blokta $k = \mu$ nokta bulunmalıdır. Eğer $B // H$ ise $|B, H| = \mu$ ve ayrıca ;

(1) Paralel sınıflar $s \geq 2$ bloğa sahip ve $r \geq 3$ paralel sınıf var ise ,

(2) $p, q \in \mathcal{T}$ herhangi iki nokta ($p \neq q$) olmak üzere $|p, q| = 0$ veya $|p, q| = \lambda$ ise

\mathcal{T} yapısı “Afin Çözülebilir Kısmi Düzlem” (Affine resolvable partial plane-ARPP) veya (s, r, μ) -ARPP olarak adlandırılır (Drake, 1979).

Bu yapılar, Bruck tarafından tanımlanan $(s, r, 1)$ -netlerinin genelleştirilmiş halidir. (s, r, μ) -netlerin duallerinin Transversal düzenler olduğu Hanani tarafından gösterilmiştir (Hanani , 1974).

Verilen homojen replikasyonlar quasi simetrik, $(s, r, 1)$ -net yapısına uyar ($\lambda = \mu$ hali). Yani quasi simetrik ARPP'dir. Deficiency; $d = 0$ olduğundan ($d = r - k + 1$) Afin düzleme karşılık gelir (Bayrak ve Gönen, 2002). Bu replikasyonlar aynı zamanda Afin yeniden çözülebilirdir.

Homojen replikasyonların duali dikkate alındığında, düğüm noktalarının oluşturduğu uzayın dual uzayı $T(4,1,3)$ transversal düzeninin transversal gruplarını meydana getirir. Bu gruplar aşağıdaki gibidir.

$$G_1 : [0 1 0] : (0 0 1) (1 0 1) (1 0-1) \rightarrow C, AC, AC^2$$

$$G_2 : [0 0 1] : (0 1 0) (1 1 0) (1-1 0) \rightarrow B, AB, AB^2$$

$$G_3 : [0 1 1] : (0 1-1) (1 1-1) (1-1 1) \rightarrow BC^2, ABC^2, AB^2C$$

$$G_4 : [0 1-1] : (1 1 1) (0 1 1) (-1 1 1) \rightarrow ABC, BC, AB^2C^2$$

4 tane transversal grup mevcuttur. Bu transversal düzen , parametreleri $v=12$, $b = 9$, $k = 4$, $r = 3$, $\lambda_1 = 0$, $\lambda_2 = 1$ olan yarı-düzgün gruplarına ayrılabilen düzene (semi-regular group divisible-SRGD) izomorftur.

Teorem 3.1 : Transversal düzenler, Afin düzenlerin dualidir (Mavron, 2000).

Dual yapıların, Bose ve Connor tarafından tanımlanan SRGD'nin özel haline karşılık geldiği görülür (Drake, 1979).

Dual düzenin geometrik yapısı pseudo SLBG'a (pseudo Strongly Linked Block Graph) izomorftur.

Teorem 3.2 : $(\beta (\gamma(\beta-1) + 1) , \beta\gamma + 1 , \beta k)$ parametrelili bir Afin çözülebilir düzeninin dualinin parametreleri aşağıdaki gibidir :

$$v = \beta (\gamma\beta^2 + \beta + 1) , \quad b = \beta^2(\gamma(\beta-1) + 1) , \quad r = \beta(\gamma(\beta-1) + 1) ,$$

$$k = \beta^2\gamma + \beta + 1 , \quad m = \beta^2\gamma + \beta + 1 , \quad n , \quad \lambda_1 = 0 , \quad \lambda_2 = \gamma(\beta-1) + 1$$

Bazı düzgün gruplarına ayrılabilen (regular group divisible-RGD) düzenlerin dualleri yine GD düzenlerdir. RGD düzen orijinal düzene izomorfik olabilir.

4. KALINTI DÜZENLER , BİPLANE VE DUAL DÜZENLERİ

Bir kalıntı düzen (residual design) $PG(2,q)$ dan herhangi bir doğrunun çıkarılmasıyla kurulabilir.Yeni düzen bir $EG(2 , q)$ olur. Benzer bir sonuç Biplane($\lambda=2$) lerin kalıntı düzenlerinin dualleri gözönüne alınarak bulunabilir.

Bir biplane $v = b = (n^2+3n+4) / 2 , r = k = n+2 , \lambda = 2$ parametrelerine sahip BIBD ise, kalıntı düzen

$v = n(n+1) / 2 , b = (n+1)(n+2) / 2 , r = n+2 , k = n , \lambda = 2$ parametrelerine sahiptir.

Teorem 4.1: $v = n(n+1) / 2 , k = n , \lambda = 2$ parametrelili bir düzen alınsın, Dual düzen aşağıdaki parametrelere sahip bir PBIBD dir.

$$v^* = (n+1)(n+2) / 2 , \quad b^* = n(n+1) / 2 , \quad r^* = n , \quad k^* = n+2$$

$$n_1 = 2n , \quad n_2 = n(n-1) / 2 , \quad \lambda_1 = 1 , \quad \lambda_2 = 2 , \quad p_{11}^1 = n , \quad p_{11}^2 = 4$$

$TC(7,4,2,2)$ konfigurasyonu $v = b = 7 , r = k = 4 , \lambda = 2$ parametrelili BIBD tanımlar.Geometrik olarak Biplane dir.Bu parametrelere sahip Biplane ile ilişkili grafların seti , Hussain graflarının tam setini verir.Bu sette $(k-1)(k-2) / 2 = 3$ tane graf vardır.Residual düzen ise $v = 3 , b = 6 , k = 2 , \lambda = 2$ parametrelili bir yapıdır (Bayrak ve Gönen, 2002).

Kalıntı düzenin duali düşünüldüğünde ,

$$v^* = 6 , \quad b^* = 3 , \quad r^* = 2 , \quad k^* = 4 , \quad \lambda_1 = 1 , \quad \lambda_2 = 2 ,$$

$$n_1 = 4 , \quad n_2 = 1 , \quad p_{11}^{1*} = 2 , \quad p_{11}^{2*} = 4$$

parametrelili Üçgensel PBIBD (Triangular PBIBD) elde edilir.

Sonuç olarak, üçgensel düzenler, $\lambda=2$ için SBIBD'lerin kalıntı düzenlerinin dualleri gibi düşünülür.

KAYNAKLAR

- BAYRAK, H. ve GÖNEN, S. (1998) , *Sanal Deney Düzeninde Dualite*, İstatistik Konferansı, Gazi Üniv., ss 245-250.
- BAYRAK, H. ve GÖNEN, S. (2002), *The Geometrical Structures of Some Dual Designs*, İstatistik Günleri 2002 Sempozyumu, Hacettepe Üniversitesi.
- BOSE, R. C. (1963) , *Strongly Regular Graphs, Partial Geometries and Partially Balanced Designs*, Pacific. J.of Maths., 13, pp 389-419.
- DRAKE, A. D. (1979), *Partial λ -Geometries and Generalized Hadamard Matrices Over Groups*, Can. J. Math., Vol. 31, No:3, pp 617-627.
- HANANI, H. (1974) , *On Transversal Designs, Proceedings of the Advanced Study Institute On Combinatorics*, Breukelen, Math. Centre Tract 55, Amsterdam, pp 42-52.
- MAVRON, V. C. (2000), *Frequency Squares and Affine Designs*, The Electronic Journal of Combinatorics 7, R56, pp 1-6.
- RAGHAVARAO, D. (1971) , *Constructions and Combinatorial Problems in Design of Experiments* , Dover Publications, Inc. , New York.
- STREET, A. P. and STREET, D. J. (1987) , *Combinatorics of Experimental Design*, Oxford Univ. Press,
- SHRIKHANDE, S.S. (1952), *On The Dual of Some Balanced Incomplete Block Designs* , Biometrics , 8 , pp 66-72.
- SHRIKHANDE, S.S. and BAHAGWANDAS, (1965), *Dual of Incomplete Block Designs*, J. Indian Stat. Assn., 3, pp 30-37.

The Dual Structure Of Incomplete Block Designs

ABSTRACT

The problem of arranging object so that certain criteria are fulfilled is of great generality in combinatorial analysis. Such an arrangement is known as an incidence system or tactical configuration. A special type of incidence system is balanced incomplete block design(BIBD).

The dual of a design is defined as a new design whose treatments and blocks are correspondance with blocks and treatments of the original design, and incidence is preserved.

Purpose of this study is to work on geometrical structure of some dual designs. These dual designs are members of class of the incomplete block design.

Key Words : *Dual Design, Balanced Incomplete Block Design, Projective Geometry, Residual Design.*

Çok Değişkenli Normal Dağılıma Sahip Örneklerdeki Aykırı Gözlemlerin Belirlenmesi İçin Bayesgil Bir Yaklaşım

Ufuk EKİZ*

ÖZET

Bu çalışmada çok değişkenli normal dağılıma sahip örneklerdeki aykırı gözlemlerin belirlenmesi için önerilen Bayesgil bir metod tanıtılmaktadır. Örnekleme teorisinde de kullanılan karesel formların dağılımından yararlanarak aykırı gözlemleri belirlemek fikrinden hareketle , gerçekleşmiş hataların karesel formlarının sonsal (posterior) dağılımı , bu gözlemlerin belirlenmesi için kullanılmaktadır. Uygulamada , gerçek bir veri üzerinde her bir gözleme ilişkin sonsal olasılıklar elde edilmiş ve aykırı (outlier) gözlem(ler) belirlenmiştir.

Anahtar Kelimeler: Baskın Olmayan Önsel , Aykırı Gözlem , Bayes Faktör , Önsel İhtimal , Sonsal İhtimal .

1. GİRİŞ

Aykırı gözlemler , verinin genel özelliklerini göstermeyen gözlemlerdir. Bazı temel modellere göre aşırılık (extremeness) gösterirler. Barnett ve Lewis ,1994,sh.7 aşırılığın , gözlem değerlerinin "sıralanmasının" bazı formlarından kaynaklanabileceğini vurgulamışlardır. Ayrıca aşırı gözlemleri ifade etmek için uygun bir alt sıralama özelliğinin benimsenmesine ihtiyaç olduğunu tartışmışlardır. Barnett ,1976 çok değişkenli problemlerde alt-sıralama kurallarının özelliklerini göz önünde bulundurmakta ve onları dört sınıfa ayırmaktadır. Bu dört alt-sıralama sınıfları ; düşürülmüş , marjinal , kısmi ve koşullu olarak isimlendirilmiştir . Düşürülmüş alt-sıralama çok değişkenli problemlerde aykırı gözlemleri belirlemede kullanılan neredeyse tek örnektir. Bu alt-sıralama , p boyutlu Y_1, Y_2, \dots, Y_n rasgele değişkenlerinin gözlenmiş değerlerinden oluşan örneği , tek değişkenli $R(Y)$ istatistiğinin değerleri bakımından sıralayarak uygulanmaktadır. Şayet

$$R(Y_i) = \max\{R(Y_j); j = 1, 2, \dots, n\}$$

ise , i. gözlemden aykırı olarak şüphelenilebilir.

Ayrıca , göz önünde bulundurulan temel model altında $R(Y_i)$, istenmeyecek ölçüde dağılımın konumundan(location) uzakta ise , Y_i , aykırı gözlem olarak açıklanabilir.

* Gazi Üniv.Fen-Ed.Fak.İstatistik Böl. Teknikokullar/ANKARA

Literatürde genellikle , η örneğin ya da yığının konumunu ve V 'de örneğin veya yığının yayılımını(scale) ifade etmek üzere , çok değişkenli aykırı gözlemlerin belirlenmesine ilişkin olarak örnekleme teorisinde ;

$$R(Y_j, \eta, V) = (Y_j - \eta)^T V^{-1} (Y_j - \eta) \quad (1)$$

formundaki istatistikler kullanılmaktadır. Eğer ilgilenilen dağılımın ortalama vektörü μ ve kovaryans matrisi Σ , daha önceden biliniyorsa , (1)

$$R(Y_j, \mu, \Sigma) = (Y_j - \mu)^T \Sigma^{-1} (Y_j - \mu) \quad (2)$$

şeklinde yazılabilir. Bu karesel formda , μ ve Σ çoğunlukla bilinmedikleri için (2) de onların yerine örnekten elde edilen tahmin edicileri kullanılır. Bu tahminler aykırı gözlemlerden etkilenebilmektedirler. Bunun için bazı yazarlar μ ve Σ için sağlam (robust) tahminlerinin kullanılmasını önermektedirler (Campell ,1980 ; Rousseeuw ve Vant Zameren ,1990).

Çok değişkenli aykırı gözlemlerin belirlenmesine yönelik olarak Guttman ,1973 Bayesgil bir yaklaşım öne sürmektedir. Bu yaklaşım , n birimlik rasgele bir örnekteki pek çok gözlemin parametreleri μ ve Σ , geriye kalan gözlemlerin ise parametreleri $\mu + \alpha$ ve Σ olan normal dağılımdan geldiğini varsaymaktadır. Guttman , aykırı gözlemleri belirlemek için α 'nın sonsal dağılımını kullanmayı öngörmektedir. Buna göre , örnekteki tüm gözlemlere ilişkin kovaryans matrisinin determinantı ile ters orantılı olarak j.gözleme bir c_j ağırlığı atanır. Eğer bir aykırı gözlem var ise buna karşılık gelen ağırlığın incelenmesi çok belirleyici olacaktır.

Çok değişkenli aykırı gözlemlerin belirlenmesine ilişkin diğer çalışmalar Gnanadesikan ve Kettenring ,1972 , Hawkins ,1980 , Rousseeuw ve Leroy ,1987 , Justel ve Pena ,2001 nin makalelerinde yer almaktadır.

Bu çalışmada çok değişkenli normal dağılıma sahip bir örnekteki aykırı gözlemlerin belirlenmesine yönelik önerilen Bayesgil bir yaklaşım tanıtılmaktadır. Bu , tek değişkenli doğrusal modellerde Chaloner ve Brant ,1988 tarafından önerilen aykırı gözlemlerin belirlenmesine yönelik bir yaklaşımın genellemesidir. Chaloner ve Brant , aykırı gözlemleri belirlemek için modele ilişkin , rasgele hataların sonsal dağılımını kullanmışlardır. Burada sunulan Bayesgil yaklaşımda , (2) de tanımlanan $R(Y_j, \mu, \Sigma)$ karesel formuna ilişkin sonsal dağılım , aykırı gözlemlerin belirlenmesinde kullanılmaktadır.

Bölüm 2 'de çok değişkenli normal dağılıma sahip p boyutlu Y_1, Y_2, \dots, Y_n rasgele değişkenlerinin gözlenmiş değerlerinden oluşan bir örnekteki aykırı gözlemlerin tanımlanması ele alınmakta , Bölüm 3'te de yapılan tanım için gerekli olan sonsal olasılıkların elde edilmesi anlatılmaktadır. Uygulamada , gerçek bir verideki herhangi bir gözlemin aykırı olup olmadığını test etmekte kullanılan bayes faktör değerlerinin

hesaplanması için gerekli sonsal olasılıklar , Monte Carlo simülasyon tekniği ile elde edilmekte ve olası aykırı gözlemler belirlenmektedir.

2. ÇOK DEĞİŞKENLİ NORMAL DAĞILIMA SAHİP YIĞINDAN ÇEKİLMİŞ RASGELE ÖRNEKLERDE AYKIRI GÖZLEMLERİN BELİRLENMESİ

Bu bölümde çok değişkenli normal dağılıma sahip rasgele bir örneğin aykırı gözlemlerinin belirlenmesine ilişkin Bayesgil bir yaklaşım sunulmaktadır. p değişkenli, $Y \sim N(\mu, \Sigma)$ olan dağılımın n çaplı rasgele bir örneği Y_1, Y_2, \dots, Y_n olsun.

$$\varepsilon_i = \Sigma^{-1/2} (Y_i - \mu) \quad , \quad i=1,2,\dots,n \quad (3)$$

$$\delta_i = \varepsilon_i^T \varepsilon_i = R(Y_i, \mu, \Sigma)$$

şeklinde tanımlansın. Burada $\varepsilon_i \sim N(0, I)$ dır. Eğer δ_i nin sonsal dağılımı üzerinden $\delta_i > k$ olasılığı , uygun k değerini belirlemek için kullanılacak olasılıktan büyük ise , i . gözlemin aykırı olduğu söylenebilir. Herhangi bir gözlemin aykırı olarak nitelenmesine izin vermeyecek kadar büyük bir önsel(prior) olasılığa karşılık gelecek şekilde bir k değeri belirlenebilir. Eğer olasılık 0.95 alınırsa ;

$$0.95 = \text{pr}(\delta_1 \leq k) \cdot \text{pr}(\delta_2 \leq k) \dots \text{pr}(\delta_n \leq k)$$

yazılabilir. Diğer bir ifade ile

$$0.95 = \text{pr}(\delta_i \leq k, \text{ tüm } i \text{ ler için}) = \{F_p(k)\}^n \quad (4)$$

dır. Burada , $F_p(\cdot)$, p serbestlik dereceli merkezsiz ki-kare dağılımına sahip rasgele değişkenin dağılım fonksiyonudur. Yukarıda verilen eşitliğin çözümünden ;

$$k = F_p^{-1}(0.95^{1/n})$$

elde edilir. δ_i 'nin sonsal dağılımının elde edilmesi ve $p_i = \text{pr}(\delta_i > k / Y)$ sonsal olasılıklarının hesaplanabilmesi için μ ve Σ 'nın bilgi vermeyen(noninformative) önselleri kullanılacaktır. Bulunan p_i olasılık değerlerinden , muhtemel aykırı gözlemler belirlenebilir. Bu p_i olasılık değerleri sıralanabilirler ve en büyük p_i olasılık değerine sahip gözlemler aykırı gözlemler olarak ifade edilebilir. Bu çalışmada , i . gözlemin aykırı olup olmadığına ilişkin ,

$$H_{0i} : \delta_i > k \quad , \quad (Y_i \text{ aykırı bir gözlem ise })$$

$$H_{1i} : \delta_i \leq k \quad , \quad i = 1,2,\dots,n$$

hipotezi test etmek için Bayes Faktörleri kullanılmaktadır. H_{1i} 'ye karşı H_{0i} hipotezinin testi için B_i Bayes faktörü ;

$$B_i = \frac{p_i F_p(k)}{(1-p_i) \{1 - F_p(k)\}}$$

H_{0i} 'ye ilişkin sonsal olasılıkların (posterior odds) , önsel olasılıklara (prior odds) oranıdır. Kass ve Raftery ,1995 , B_i değerinin 10 'dan büyük olmasının H_{0i} için güçlü , 100 'den büyük olmasının ise çok güçlü bir kanıt olduğunu öne sürmektedir. Bayes faktör değerlerinin elde edilmesinde gerekli p_i sonsal olasılıkları , μ ve Σ 'nın bilgi vermeyen önsellerinin kullanılması ile Σ ve Y 'nin biliniyor olması koşulu altında bulunacak δ_i 'nin sonsal dağılımından elde edilecektir.

3. μ VE Σ İÇİN BASKIN OLMAYAN ÖNSEL FONKSİYONU VE p_i SONSAL DEĞERLER

δ_i 'nin sonsal (posterior) dağılımını elde etmek ve $p_i = \text{pr}(\delta_i > k / Y)$ olasılığını hesaplamak için μ ve Σ 'nın bilgi vermeyen bileşik önsel fonksiyonu , μ ve Σ 'nın yaklaşık olarak bağımsız oldukları varsayımı altında;

$$p(\mu, \Sigma) = p(\mu) \cdot p(\Sigma)$$

dır. Box ve Tiao ,1973, sh.426 çok değişkenli normal dağılımda μ ve Σ 'nın bilgi vermeyen önsel fonksiyonlarını sırası ile ;

$$p(\mu) \propto 1$$

$$p(\Sigma) \propto |\Sigma|^{-\frac{p+1}{2}}$$

ve birleşik önsel fonksiyonu da ,

$$p(\mu, \Sigma) = p(\mu) \cdot p(\Sigma) \propto |\Sigma|^{-\frac{p+1}{2}}$$

şeklinde vermektedir.

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} \quad , \quad S = \sum_{i=1}^n (Y_i - \bar{Y})(Y_i - \bar{Y})^T$$

olarak tanımlanırsa , gözlenen değerlere ilişkin olabilirlik fonksiyonu ;

$$f(Y/\mu, \Sigma) \propto |\Sigma|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\{iz(\Sigma^{-1}S) + n(\bar{Y} - \mu)^T \Sigma^{-1}(\bar{Y} - \mu)\}\right]$$

formunda ifade edilir. Bileşik önsel fonksiyonu ve Y 'nin biliniyor olması koşulu altında μ ve Σ 'ya ilişkin olabilirlik fonksiyonlarını kullanarak (μ, Σ) nin bileşik sonsal fonksiyonu;

$$p(\mu, \Sigma/Y) \propto p(\mu, \Sigma).f(\mu, \Sigma/Y)$$

$$\propto |\Sigma|^{-\frac{n+p+1}{2}} \exp\left\{-\frac{1}{2}iz(\Sigma^{-1}S)\right\} \exp\left\{-\frac{n}{2}(\mu - \bar{Y})^T \Sigma^{-1}(\mu - \bar{Y})\right\}$$

olarak elde edilir. Bileşik sonsal fonksiyon

$$p(\mu, \Sigma/Y) = p(\mu/\Sigma, Y).p(\Sigma/Y)$$

şeklinde ifade edildiğinden ,

$$\mu/\Sigma, Y \sim N(\bar{Y}, n^{-1}\Sigma)$$

$$\Sigma/Y \sim W^{-1}(S, p, n-p)$$

bulunur. Burada $W^{-1}(S, p, n-p)$ ters (inverted) Wishart dağılımıdır (Box ve Tiao ,1973 , sh.460-464).

$\mu/\Sigma, Y$ ve Σ/Y 'nin sonsal fonksiyonlarını kullanarak $\varepsilon_i/\Sigma, Y$ ve $\delta_i/\Sigma, Y$ 'nin koşullu sonsal fonksiyonları bulunabilir. ε_i 'ler , Σ verildiğinde μ 'nün doğrusal fonksiyonlarıdır. Y ve Σ 'nın biliniyor olması koşulu altında μ 'ye ilişkin sonsal dağılım ;

$$p(\mu/\Sigma, Y) = c \left[\exp\left\{-\frac{n}{2}(\mu - \bar{Y})^T \Sigma^{-1}(\mu - \bar{Y})\right\} \right]$$

olarak ifade edilir.

$$\varepsilon_i = \Sigma^{-\frac{1}{2}}(Y_i - \mu)$$

$$\mu = Y_i - \Sigma^{\frac{1}{2}}\varepsilon_i$$

dönüşümü uygulandığında , dönüşüme ilişkin Jacobian terimi bir sabit olacağından c ifadesi içinde yer alır. Y ve Σ 'nın biliniyor olması koşulu altında ϵ_i 'nin sonsal fonksiyonu ;

$$p(\epsilon_i / Y, \Sigma) = c. \left[\exp \left\{ -\frac{n}{2} (Y_i - \Sigma^{\frac{1}{2}} \epsilon_i - \bar{Y})^T \Sigma^{-1} (Y_i - \Sigma^{\frac{1}{2}} \epsilon_i - \bar{Y}) \right\} \right]$$

$$= c. \left[\exp \left\{ -\frac{n}{2} [\epsilon_i - \Sigma^{-\frac{1}{2}} (Y_i - \bar{Y})]^T [\epsilon_i - \Sigma^{-\frac{1}{2}} (Y_i - \bar{Y})] \right\} \right]$$

elde edilir. $\gamma_i = \Sigma^{-\frac{1}{2}} (Y_i - \bar{Y})$ biçiminde tanımlanırsa, $\epsilon_i / \Sigma, Y \sim N(\gamma_i, n^{-1}I)$ olur. Burada c , fonksiyona ilişkin sabit terimdir. Böylece Σ ve Y verildiğinde $W_i = n\delta_i$, serbestlik derecesi p , merkezsiz olmama parametresi $\lambda_i = n(Y_i - \bar{Y})^T \Sigma^{-1} (Y_i - \bar{Y})$ olan merkezsiz olmayan bir ki-kare dağılımıdır.

Yukarıda verilen dağılımdan ,

$$p_i = E_{\Sigma/Y} [\text{pr}(\delta_i > k / Y, \Sigma)]$$

$$= E_{\Sigma/Y} [\text{pr}(W_i > nk / Y, \Sigma)] \quad ; i=1,2,\dots,n$$

yazılabilir. Hesaplama kolaylığı bakımından p_i 'leri $\Sigma^{-1} = V$ metriğinde yazmak daha uygundur.

$$p_i = E_{V/Y} [\text{pr}(W_i > nk / Y, V)] \quad (5)$$

yazılabilir. Burada $\lambda_i = n(Y_i - \bar{Y})^T V (Y_i - \bar{Y})$ ve

$$V / Y \sim W(S^{-1}, p, n - p)$$

dir. Her bir i için p_i 'nin tahmin değerlerinin elde edilmesinde Monte Carlo teknikleri uygulanabilir.

4. UYGULAMA

Aşağıda , Tablo 1 'de verilen çok değişkenli normal dağılıma sahip rasgele değişkenlere ilişkin gözlenen değerler Khattree R. ve Naik D.N.,1999 Appendix B den alınmıştır. Burada $n = 28$, $p = 4$ tür.

Tablo 1. (Cork Data)

	Y_1	Y_2	Y_3	Y_4		Y_1	Y_2	Y_3	Y_4
1	72	66	76	77	15	91	79	100	75
2	60	53	66	63	16	56	68	47	50
3	56	57	64	58	17	79	65	70	61
4	41	29	36	38	18	81	80	68	58
5	32	32	35	36	19	78	55	67	60
6	30	35	34	26	20	46	38	37	38
7	39	39	31	27	21	39	35	34	37
8	42	43	31	25	22	32	30	30	32
9	37	40	31	25	23	60	50	67	54
10	33	29	27	36	24	35	37	48	39
11	32	30	34	28	25	39	36	39	31
12	63	45	74	63	26	50	34	37	40
13	54	46	60	52	27	43	37	39	50
14	47	51	52	43	28	48	54	57	43

Hiçbir gözlemin aykırı olmaması olasılığı 0.95 olarak seçildiğinde $pr(\delta_i < k) = F_k(.) = 0.999$ ve $k=17.1211$ bulunur. MAT LAB (6.0) paket programında , Monte Carlo simülasyon tekniği kullanılarak yazılan programdan, (5) 'teki p_i olasılıklarına ilişkin tahminler elde edilmiştir. Her bir gözlem için , $W(S^{-1},4,24)$ dağılımından rasgele bir matris üretilmiş ve varyansın biliniyor olması koşullu , merkezsiz olmayan ki-kare olasılığı hesaplanmıştır ($pr(W_i > nk / Y, V)$). Bu döngü 20.000 defa tekrarlanmış ve p_i bu koşullu olasılıkların ortalaması olarak alınmıştır. Her bir gözlem için elde edilen Bayes Faktör ve p_i olasılık değerleri aşağıda verilen Tablo 2 'de yer almaktadır.

Tablo 2. Bayes Faktör (B_i) ve p_i Olasılık Değerleri

\hat{I}	B_i	p_i	\hat{I}	B_i	p_i
1	0.5457	0.0010	15	3.5766	0.0066
2	2.1698E-015	3.9801E-018	16	14.0862	0.0265
3	1.2283E-014	2.2532E-017	17	1.9158E-008	3.5142E-011
4	5.9138E-014	1.0848E-016	18	1.8242	0.0034
5	5.2341E-014	9.6012E-017	19	0.1000	1.8338E-004
6	7.0795E-014	1.2986E-016	20	1.0472E-013	1.9209E-016
7	3.6632E-014	6.7196E-017	21	7.9652E-014	1.4611E-016
8	2.6495E-009	4.8601E-012	22	1.1157E-013	2.0465E-016
9	1.7219E-014	3.1586E-017	23	5.9686E-014	1.0948E-016
10	2.9206E-013	5.3575E-016	24	4.1208E-009	7.5591E-012
11	4.2923E-014	7.8737E-017	25	9.8705E-014	1.8106E-016
12	0.0050	9.2551E-006	26	4.6401E-006	8.5116E-009
13	5.8215E-014	1.0679E-016	27	0.0015	2.8117E-006
14	3.5794E-014	6.5659E-017	28	1.5924E-009	2.9211E-012

Sadece on altıncı gözleme ilişkin Bayes Faktör değeri $B_{16} = 14.0862$, 10 'dan büyük çıkmıştır. Buna göre on altıncı gözlemin aykırı olarak düşünülmesi için çok güçlü bir kanıt mevcuttur. Khattree R. ve Naik D.N. , bu verideki aykırı gözlemlerin belirlenmesi için örnekleme teorisindeki farklı yöntemleri kullanmışlar ve aynı sonucu elde etmişlerdir.

5. SONUÇ

Bu çalışmada , önerilen Bayesgil yaklaşımda , çok değişkenli normal dağılıma sahip örneklerdeki aykırı gözlemleri belirlemek için hata terimlerinin karesel formlarının koşullu sonsal dağılımı kullanılmıştır. Eğer parametre değerleri bilinmiyorsa (2)'deki istatistiğin iyi bir alternatifi yoktur. Parametrelerdeki belirsizlik , (2) nin sonsal dağılımının kullanılması ile açıklanmış olur. Bu yaklaşımın en az iki avantajı bulunmaktadır. Birincisi, ifade etmesi kolaydır ve ikincisinde , tahmin edicilerin aykırı gözlemlerden etkilenmesi bu analizde daha az etkili olur . Parametreler için bir önsel dağılım mevcut değilse Bilgi vermeyen önseli seçmek mantıklıdır. Bilgi veren (Informative) önselleri kullanmakta mümkün olabilir. Çok değişkenli normal dağılımda , ortalama parametresi için normal önseli ve kovaryans parametresi için ters Wishart önseli kullanıldığında elde edilen bileşik önsel fonksiyon uygun parametre değerleri için , bilgi vermeyen bileşik önsel fonksiyon ile aynı sonucu vermektedir (Peter M.Lee ,1989.sh.73). Eğer başka bilgi veren önseller kullanılırsa , hatalara ilişkin sonsal dağılım bilinen bir formda ortaya çıkmayabilir . Bu durumda Marcov Chain Monte Carlo (MCMC) teknikleri p_i 'lerin tahmini için gerekli olur.

KAYNAKLAR

- BARNETT,V.(1976),*The ordering multivariate data (withdiscussion)*, J.Roy.Statist.Soc.Ser.A,139,318-54.
- BARNETT,V.,LEWIS,T(1994),*Outlier in Statistical Data*,3rd.ed.Chichester:John Wiley & Sons.
- BOX,G.E.P.,TIAO,G.C.(1973),*Bayesian Inference in Statistical Analysis*, Reading, MA:Addison-Wesley .
- CAMPBELL,N (1980) , *Robust Procedures in Multivariate analysis I :Robust Covariance Estimation.*, Applied Statistics, 29, 231-37.
- CHALONER,K.,BRANT,R.(1988),*A Bayesian Approach to Outlier Detection and Residual Analysis.* , Biometrika,75,651-9.
- GNANADESIKAN,R.,KETTENRING,J.(1972),*Robust Estimates,Residuals and Outlier Detection With Multiresponse Data.*, Biometrics,28,81-124.
- GUTTMAN,I.(1973),*Care and Handling of Univariate or Multivariate Outliers in Detecting Spuriousity-a Bayesian Approach.*, Technometrics.15.723-38.
- HAWKINS,D.(1980) , *Identification of Outliers* , London , Chapman and Hall.
- JUSTEL,A.,PENA,D.(2001),*Bayesian unmasking in Linear Models*, Computational Statistics & Data Analysis.36,69-84.
- PETER M.LEE.(1989) , *Bayesian Statistics* , London,Chapman and Hall.
- KASS,R.,RAFTERY,A.(1995) , *Bayes Factors* , JASA , 90 , 773-95.
- KHATTREE R.,NAIK N.D.(1999),*Applied Multivariate Statistics with SAS Software* , SAS Institute.
- ROUSSEEUW,P.,LEROY,A.(1987),*Robust Regression and Outlier Detection*, NewYork:Wiley.
- ROUSSEEUW,P.,VAN ZAMEREN,B.(1990),*Unmasking Multivariate Outliers and Leverage Points* , JASA , 85 , 633-9.

A Bayesian Method to Identification of Outlier Observations in Multivariate Normal Distribution

ABSTRACT

In this work , a method of Bayesian which is suggested for determine of outlier observation in the samples which has multivariate normal distribution is introduced. Outlier observation is introduced using distribution of quadratic forms in also sampling theory. Posterior odds of quadratic forms of errors are used to determine of observations. In aplication , posterior odds in the real data corresponding to every observation are found and outlier observations are determined.

Key Words: *Noninformative Prior , Outlier Observation , Bayes Factor ,Prior Odds , Posterior Odds.*

Damar Darlık Derecesi ve Risk Faktörlerinin Homojenite Analizi ile İncelenmesi

Necdet SÜT*

Mevlüt TÜRE*

ÖZET

Bu çalışmada, Homojenite Analizi (HA) ile anjiyografik olarak saptanan aterosklerotik damar darlık derecesi üzerine cinsiyet, yaş, sigara, hipertansiyon, ailede koroner kalp hastalığı olup olmadığı ve diyabet değişkenlerinin etkilerinin incelenmesi amaçlandı.

Anjiyografik olarak saptanan aterosklerotik damar darlık derecelerini etkileyebilecek faktörlerin HA ile incelenmesi sonucunda 54 ve daha düşük yaşlarda hipertansiyon, diyabet ve ailede koroner kalp rahatsızlığı gözlenmeyen hastaların damar darlık derecesinin %50 ve daha düşük olduğu, bununla birlikte 55 ve daha yukarı yaşlarda hipertansiyon, diyabet ve ailede koroner kalp rahatsızlığı gözlenen hastaların damar darlık derecesinin %70 ve daha yüksek olduğu bulundu. Ayrıca erkeklerin sigara içme, kadınların içmeme eğiliminde oldukları da bulundu.

Sonuç olarak kategorik yapıda değişkenlerden oluşan veri setini incelemede HA'nin kullanışlı bir yöntem olduğu söylenebilir.

Anahtar Sözcükler: Homojenite Analizi, Optimal Ölçekleme, Damar Darlık Derecesi.

1. GİRİŞ

Günümüzde sağlık alanında verilerin doğru ve tutarlı analizi büyük önem taşımaktadır. Kategorik yapıya sahip verilerin analizinde ki-kare analizinin uygulanması çapraz tabloların satır ve sütun sayılarının fazla olduğu durumlarda satır ve sütunların birbirleri ile karşılaştırmasını zorlaştırmaktadır. Bu nedenle değişken sayılarının fazla olduğu çapraz tablolarda çok değişkenli analizlerden yararlanmak gerekmektedir.

RxC boyutlu çapraz tabloların satır ve sütunlarının hem kendi içlerinde hem de birbirleri arasındaki türdeşliğini belirlemek için Basit Uyum Analizi'nden, RxCxKx. biçimindeki çok boyutlu çapraz tabloların analizi için ise HA'den yararlanılmaktadır (Özdamar, 1999; Türe, 1993, Meulman, 1982; Michailidis G. & De leeuw J, 2000).

* Trakya Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı – Edirne

2. GEREÇ VE YÖNTEM

2.1. Homojenite Analizi

HA, $R \times C \times K \times \dots$ biçiminde iç içe değişik biçimlerde çaprazlanmış tablolarda yer alan değişkenlerin alt kategorileri arasındaki bağımlılığı ortaya koyan bir yöntemdir. Satırlarda birimlerin, sütunlarda ise değişken kategorilerinin yer aldığı ve değerlerinin 0 ve 1'lerle gösterildiği matrisin iç çarpımı sonucu elde edilen tabloların basit uyum analizidir (Greenacre ve Blasius, 1994; De Leeuw, 1984; Van der Heijden ve De Leeuw, 1989; Özdamar, 1999; Van der Heijden ve ark., 1999; SAS Inc., 1994; SPSS Categories, 1999; Statsoft Inc., 2000).

Satırlarında birimlerin ve sütunlarında değişkenlerin yer aldığı $(n \times m)$ boyutlu çok değişkenli bir veri matrisi varsayalım. Bu varsayılan değişkenlerden k_j kategorili j değişkeninin gösterge matrisi G_j $(n \times k_j)$ olarak tanımlanır. Gösterge matrisinde, birimler ile kategori skorları yer alır. Gösterge matrisinin satırlarında birimlerin değerleri, sütunlarında kategoriler gösterilmektedir. Gösterge matrisinin elemanları 0 ve 1 değerlerinden oluşmaktadır.

Her bir değişkenin kategori açıklayıcılıklarının, türdeşliğin maksimize edilmesiyle belirlendiği HA'de alternatif en küçük kareler yöntemi yardımıyla kayıp fonksiyon minimize edilerek çözümler sağlanır. (Gifi, 1990; Meulman, 1982; Van der Burg E. 1988; Michailidis G. and De leeuw J., 2000; Michailidis G. and De leeuw J, 1999; Bond J. and Michailidis G., 1997).

HA'de $(n \times p)$ boyutlu X obje skorları matrisiyle, dönüştürülmüş değişkenler $(G_j Y_j)$ arasındaki farkın kareleri toplamı minimize edilerek çözümler sağlanır (Gifi, 1990; Van der Burg, 1988). HA'de

$$\left. \begin{array}{l} X'X = nI \\ \text{ve} \\ u'X = 0 \end{array} \right\} \text{olmak üzere}$$

kayıp fonksiyon,

$$\min \sigma(X, Y) = \sum_{j=1}^m (X - G_j Y_j)^2 \quad (1)$$

minimize edilerek çözüm sağlanır. Formülde yer alan u , n birimden oluşan birim vektörü; X , obje skorları matrisini (ortalaması 0, varyansı 1); G_j , j değişkeninin gösterge matrisini; $Y_j = (G_j' G_j)^{-1} G_j' X$ kategori açıklayıcılıklarını ifade etmektedir.

$(n \times k_j)$ boyutlu G_j matrisi, j değişkeninin gösterge matrisini ifade etmektedir. Eğer i birimi j değişkeninin l . kategorisine aitse G_j matrisinin (i, l) elemanı 1'e eşittir. (Van der Burg, 1988).

Tablo 1’de, m=3 kategorik değişkenin ölçümlerinden oluşan örnek bir veri matrisi, Tablo 2’de, Tablo 1’de yer alan 3 değişkenin gösterge matrisi yer almaktadır. Tablo 3’te ise değişkenlerin tek boyutlu gösterimi yer almaktadır (Van der Burg, 1988).

Tablo 1. Veri Matrisi

	1,	...	m
1	a	+	α
2	a	-	β
.	b	-	β
.	b	+	τ
.	a	-	τ
n	b	x	α

Tablo 2. Gösterge Matrisler

	a	b	+	-	x	α	β	τ
1	1	0	1	0	0	1	0	0
2	1	0	0	1	0	0	1	0
.	0	1	0	1	0	0	1	0
.	0	1	1	0	0	0	0	1
.	1	0	0	1	0	0	0	1
n	0	1	0	0	1	0	0	0

\uparrow G1 \uparrow G2 \uparrow G3

Tablo 3. Değişkenin Tek Boyutlu Gösterimi

$$\begin{array}{c} \alpha \\ \beta \\ \beta \\ \tau \\ \tau \\ \alpha \end{array} \begin{array}{c} \leftarrow \rightleftarrows \rightarrow \\ \leftarrow \rightleftarrows \rightarrow \\ \leftarrow \rightleftarrows \rightarrow \\ \leftarrow \rightleftarrows \rightarrow \\ \leftarrow \rightleftarrows \rightarrow \\ \leftarrow \rightleftarrows \rightarrow \end{array} \begin{array}{c} Y_{13} \\ Y_{23} \\ Y_{23} \\ Y_{33} \\ Y_{33} \\ Y_{13} \end{array} = \begin{array}{c} \left| \begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{array} \right| \times \begin{array}{c} Y_{13} \\ Y_{23} \\ Y_{33} \end{array} \end{array} = G_3 Y_3$$

HA'nin sonuçları birimleri ve kategori açıklayıcılıklarını kapsamakta ve kategoriler homojen alt gruplara ayrılmaktadırlar. Değişkenlerin aynı kategorideki birimleri aynı alt gruplarda sınıflandırıldığında homojendirler.

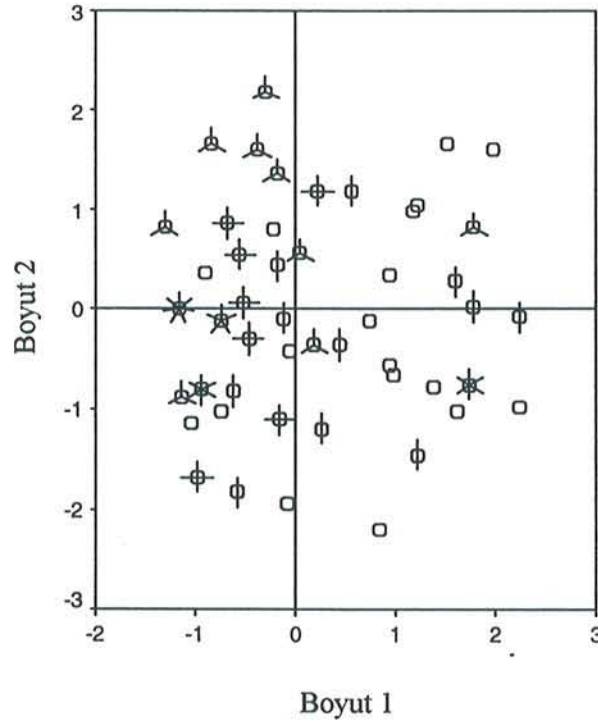
3. BULGULAR

Trakya Üniversitesi Tıp Fakültesi Hastanesine 2000-2002 yılları arasında başvuran 290 hastanın damar darlık derecesi, hipertansiyon hastası olup olmadığı, cinsiyeti, sigara içip içmediği, diyabeti olup olmadığı, ailede koroner kalp hastalığı olup olmadığı ve yaş değişkenleri kategorik veri analizi yöntemi olan HA ile iki boyutlu incelenerek toplam açıklayıcılık çok yüksek olmamakla birlikte %65 olarak bulundu. Olayın %51,1'lik kısmı Boyut-1 tarafından, %48,9'lik kısmı Boyut-2 tarafından açıklandı. Değişkenler ve alt kategorileri Tablo 4'te verildi.

Tablo 4. Değişkenler ve Kategorileri.

<i>Değişkenler</i>		<i>Kategoriler</i>
DARLIK	Damar darlık derecesi	≥ 70 , $50 \leq$
YAŞ	Yaş grupları	<45, 45-54, 55-64, 65-74
CİNSİYET	Cinsiyet	kadın, erkek
DİYABET	Diyabet olup olmadığı	hayır, evet
HT	Hipertansiyon	hayır, evet
SİGARA	Sigara içip içmediği	hayır, evet
AİLE	Ailede koroner kalp rahatsızlığı olup olmadığı	hayır, evet

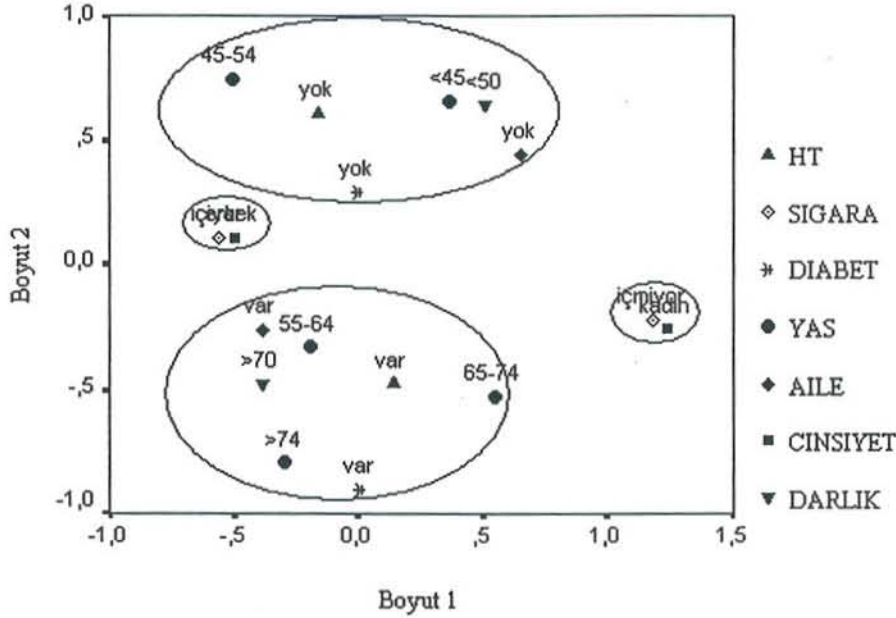
Şekil 1'de yer alan obje skorları grafiği incelendiğinde analiz sonuçlarını olumsuz etkileyebilecek sapan değerlerin olmadığı görüldü. Bunun sonucu olarak analiz edilecek değişkenlerin kategorilerinde herhangi bir birleştirme yada sapan değerleri çıkarma işlemi yapılmadı.



Şekil 1. Objeler Skorum Grafiği

Boyut-1 ile Boyut-2'nin açıklayıcılık değerleri arasında büyük farklılık olmadığından (%2,2) ve Boyut-2'nin sonuçlarının olayı açıklamada daha tutarlı olması sebebiyle Şekil-2'de verilen kategori açıklayıcılığı grafiği Boyut-2 ele alındı. Boyut-2 incelendiğinde 54 ve daha düşük yaşlarda hipertansiyon, diyabet ve ailesinde koroner kalp rahatsızlığı gözlenmeyen hastaların damar darlık derecesinin %50 ve daha düşük olduğu, bununla birlikte 55 ve daha yukarı yaşlarda hipertansiyon, diyabet ve ailesinde koroner kalp rahatsızlığı gözlenen hastalarının damar darlık derecesinin %70 ve üzerinde olduğu görülmektedir. Cinsiyet ile damar darlık derecesi arasında bir ilişki

saptanmamıştır. Ayrıca erkeklerin sigara içme eğiliminde, kadınların içmeme eğiliminde oldukları da görülmektedir.



Şekil 2. Kategori Açıklayıcılığı Grafiği

4. TARTIŞMA VE SONUÇ

χ^2 uzaklıklarını ele alarak değişkenler arasındaki türdeşliği maksimize ederek çözümler sağlayan HA kategorik veri analizinde kullanılan bir yöntemdir. İki'den daha fazla kategorik yapıya sahip değişken söz konusu olduğu durumlarda çapraz tabloların analizi HA ile doğru ve tutarlı olarak yapılabilir. Verilerin kategorilerine ait birimlerin çoğunluğunun hangi kategorilerde birlikte yer aldığı kategori açıklayıcılığı grafiğinde gösterildiğinden değişkenler HA yardımıyla kümelere ayrılırlar.

HA'nin görsel (grafiksel) sonuçları yardımıyla damar darlık derecesini yaş, hipertansiyon, diyabet ve ailede koroner kalp rahatsızlığı olup olmamasının etkilediği, ilerleyen yaşlarda (55 ve daha yüksek) diyabet ve hipertansiyon gözlenme sıklığı arttığı, bununla birlikte ailede koroner kalp rahatsızlığı da var ise damar darlık derecesi %70 ve üzerinde olduğu Şekil-2'de görülmektedir. Çalışmamızın sonuçları Onat A. (2000, 2001) yaptığı çalışma sonuçları ile paralellik göstermektedir.

Sonuç olarak HA, RxCxKx... biçiminde iç içe değişik biçimlerde çaprazlanmış isimsel ve sıralı ölçüm düzeyine sahip verilerden oluşan tabloların analizinde oldukça yararlı olan bir yöntemdir.

KAYNAKLAR

- BOND, J. and MİCHAİLİDİS, G. (1997), *Interactive CA in a Object Oriented Environment*, Erişim: [www.stat.ucla.edu/preprints]. Erişim Tarihi: 03.01.2002
- DE LEEUW, L. (1984), *Canonical Analysis of Categorical Data*, Leiden: Dswo Press.
- GİFİ, A. (1990), *Nonlinear Multivariate Analysis*, Chichester: John Wiley & Sons.
- GREENACRE, M. and BLASİUS, J. (1994), *Correspondence Analysis in the Social Sciences*, New York: Academic Press.
- MEULMAN, J. (1982), *Homogeneity Analysis with Incomplete Data*, Leiden: Dswo Press.
- MİCHAİLİDİS, G. and DE LEEUW, J.(2000), *Constrained Homogeneity Analysis with Applications to Hierarchical Data*, Erişim: [www.stat.ucla.edu/preprints]. Erişim Tarihi: 03.01.2002
- MİCHAİLİDİS, G. and DE LEEUW, J. (1999), *The GIFİ System of Descriptive Multivariate Analysis*, Erişim: [www.stat.ucla.edu/preprints]. Erişim Tarihi: 03.01.2002
- ONAT, A., ÇETİNKAYA, A., SANSOY, V., YILDIRIM, B. ve KELEŞ İ. (2000), *Koroner Kalp Hastalığı Riskini Yükselten Diyabet Sıklığı Erişkinlerimizde Hızla Artıyor, Türk Kardiyoloji Derneği Arşivi*, 28 (1).
- ONAT, A. (2001), *Risk factors and cardiovascular disease in Turkey, Atherosclerosis*, 156: 1-10
- ÖZDAMAR, K. (1999), *Paket Programlarla İstatistiksel Veri Analizi-2*, Eskişehir: Kaan Kitapevi.
- SAS INC. (1994), *SAS/STAT User's Guide, Volume 2, GLM-VARCOMP, Version 6.4th edit*, SAS Institute Inc., Cary, NC, USA, 1994.
- SPSS INC. (1999), *SPSS Categories 10.0*, Chicago: SPSS inc.
- STATSOFT INC. (2000): *STATASOFT Electronic Text Book*. Erişim: [www.statsoft.com]. Erişim Tarihi: 16.01.2002
- TÜRE, M. (1993), *Frekans Tablolarında Satır, Sütun ve Birim Sayısının Uyum Analizi Sonuçları Üzerine Etkisinin Araştırılması (Doktora Tezi)*, Eskişehir: A.Ü. Sağlık Bilimleri Enstitüsü.
- VAN DER BURG, E. (1988), *Nonlinear Canonical Correlation and Some Related Techniques*, Leiden: Dswo Press.
- VAN DER HEIJDEN, PGM., De Leeuw, J. (1989), *Correspondence Analysis, with Special Attention to the Analysis of Panel Data and Event History Data, Sociological Methodology*, Basil Blackwell, Oxford, 43-87.
- VAN DER HEIJDEN, PGM., GİLULA, Z. and VAN DER ARK, LA. (1999), *An Extended Study into the Relationship between Correspondence Analysis and Latent Class Analysis, Sociological Methodology*, Cambridge: Blackwell.

Investigation of the Effect of Risk Factors on Narrowness Degree of Blood Vessel by Using Homogeneity Analysis

ABSTRACT

The aim of our study is to investigate by using Homogeneity Analysis (HA) of factors (sex, age, smoking, hypertension, diabetes mellitus, and family history of coroner heart disease) which could be effected on narrowness degree of atherosclerosis blood vessel determined by angiography.

As a result of HA, narrowness degree of blood vessel of patients who have family history of coroner heart disease, hypertension and diabetes mellitus at aged higher than 55 is found 70% or higher, in addition narrowness degree of blood vessel of patients who have not got family story of coroner heart disease, hypertension and diabetes mellitus at aged less than 54 is found 50% or lesser. Also the men were found smoker and the women were found no smoker.

Finally we can say that HA is a useful method to analyse categorical data set.

Key words: *Homogeneity Analysis, Optimal Scaling, Degree of Blood Vessel*

Yapay Sinir Ağı Performansına Etki Eden Faktörlerin Analizinde Taguchi Yöntemi: Hisse Senedi Fiyat Tahmini Uygulaması

Alperen ÖZALP*

A. Sermet ANAGÜN**

ÖZET

İlgilenilen problemin yapısına bağlı olarak istatistiksel pek çok tahminleme yöntemi geliştirilmiştir. Bir yapay zeka tekniği olan Yapay Sinir Ağı (YSA); tanıma, sınıflandırma, tahminleme ve eniyileme konularında kullanılan etkili bir tekniktir. YSA'da model belirleme problemi literatürde önemli bir konu olarak ele alınmaktadır. Bu çalışmada, YSA performansını etkileyen faktörlerin analizi ve performansını iyileştiren uygun değerlerin belirlenmesinde Taguchi yöntemi kullanılmıştır. Uygun faktör değerleri belirlenmiş YSA, İstanbul Menkul Kıymetler Borsası'nda (İMKB) işlem gören Koç Holding hisse senedinin fiyat tahmini problemine, Ocak96-Aralık01 dönemi için derlenen veriler dikkate alınarak uygulanmıştır. Yaklaşımın etkinliğini belirlemek ve YSA'da model seçiminin önemini vurgulamak amacıyla, elde edilen sonuç rassal olarak tasarlanmış bir YSA ve çoklu doğrusal regresyon modeli ile karşılaştırılmıştır.

Anahtar kelimeler: Taguchi Yöntemi, Yapay Sinir Ağı, Regresyon Analizi, Tahminleme, Hisse Senedi.

1. GİRİŞ

İnsan sinir hücrelerinin öğrenme ve sinyal transferi fonksiyonlarının modellenmesi gayreti ile ortaya çıkmış olan YSA'lar, doğrusal olmayan veriler arasındaki ilişkileri öğrenebilme ve çıkarsama yapabilme yetenekleri nedeniyle, klasik yöntemlere göre daha çok alanda uygulanmakta ve daha başarılı sonuçlar elde edilmektedir. YSA'nın performansı, ilgilenilen problemin yapısına bağlı olarak değişiklik gösteren; öğrenme katsayısı, momentum terimi, gizli katman sayısı, gizli katmandaki işlem elemanı sayısı, eğitim algoritması, transfer fonksiyonu ve hata düzeyi gibi faktörlerin bir fonksiyonudur (Enke vd, 2000).

Her problem için geçerli olabilecek evrensel bir YSA modeli kurulamamıştır. Bu nedenle her problem için, performansı etkileyen faktörlerin uygun değerlerinin deneysel olarak belirlenmesi gerekmektedir. Benzer şekilde, diğer bir yapay zeka tekniği olan

* University of Wisconsin, Production and Manufacturing Eng. Department, Madison, WI 53711.

** Osmangazi Üniversitesi, Endüstri Mühendisliği Bölümü, Bademlik, 26030 Eskişehir.

genetik algoritmalarda, problemin uygun bir kromozom yapısı ile modellenmesi ve uygun faktör değerlerinin belirlenmesi probleme bağlı olarak değişiklik göstermekte ve ancak deneysel yaklaşımlar ile çözümlenebilmektedir (Grefenstette, 1986). Bu tür faktörlerin uygun bileşimlerinin istatistiksel yöntemlerle belirlenmesinde, başarısı bir çok çalışma ile kanıtlanmış etkili bir teknik olan deney tasarımı kullanılmaktadır (Enke vd, 2000).

Fiziksel bir sistem matematiksel olarak modellenirken, sistem özellikleri denklemlerle ifade edilmekte ve gelecekte sistemin durumu ve içinde bulunacağı şartlar tahminlenmeye çalışılmaktadır. Ancak, çok sayıda faktörün söz konusu olduğu doğrusal olmayan sistem modellerinde, bir takım kabuller ve sadeleştirmeler yapmadan bütün denklem sistemini çözmek pratik olarak mümkün değildir (Oliveria vd, 2000). Güncel yaşamda bu derece önemli ve gerekli oluşunun yanında, tutarlı tahminleme yapmanın zorluğu, bu konuda bir takım özel yöntemlerin geliştirilmesine sebep olmuştur.

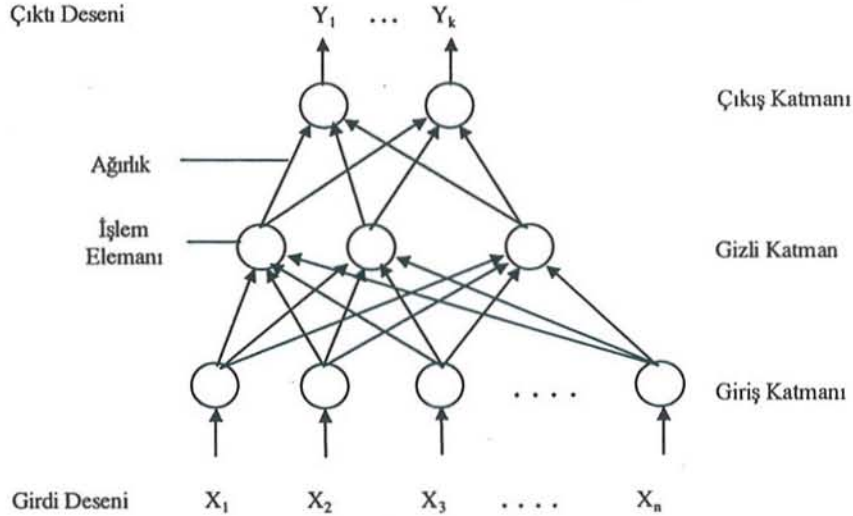
Geleneksel yöntemler, yanlış sonuçların elde edilmesi riski nedeniyle eksik ve/veya aşırı sapma içeren veriler için uygun değildir. Öte yandan, YSA yaklaşımı ise, verilere bağlı olmayıp; eksik, kısmen hatalı veya aşırı sapmalı verileri değerlendirebilir, hatta karmaşık ilişkileri öğrenebilir, genelleyebilir ve bu sayede daha önce hiç karşılaşmadığı sorulara kabul edilebilir bir hatayla cevap bulabilir. Bu özellikleri nedeniyle YSA tahminlemede etkili bir yöntem olarak kullanılmaktadır. YSA ile kaynak yönetimi (Raggad, 1996), pazar belirleme (Venugopal ve Beats, 1994) gibi konularda yapılan çalışmalara ek olarak, otellerdeki oda devir hızının tahmini (Law, 1998), uluslararası hava taşımacılığındaki yolcu sayısının tahmini (Faraway ve Chatfield, 1998), finansal analiz ve ev fiyat tahmini (Daniels ve Kamp, 1999), enflasyon oranının tahmini (Aiken, 1999), öğrenci akademik başarı ortalamasının tahmini (McLauchlan vd., 1999), zaman serileri tahmini (Oliveira vd., 2000), firma iflaslarının tahmini (Nasir vd, 2001) gibi bir dizi başarılı tahminleme çalışmaları gerçekleştirilmiştir.

Bu çalışmada, dikkate alınan faktör sayısına bağlı olarak hem tam faktöriyel hem de kesirli faktöriyel tasarımlar için uygun ve klasik deney tasarımına alternatif olan Taguchi yöntemi kullanılarak, YSA performansını etkileyen faktörlerin uygun değerleri belirlenmiştir. Geliştirilen YSA, İMKB'de işlem gören Koç Holding hisse senedinin fiyatının tahmini amacıyla kullanılmıştır. Seçilen hisse senedinin fiyatını tahmin etmek amacıyla; döviz kuru, enflasyon oranları, İMKB-100 bileşik endeksi, işlem adedi ve işlem hacmi değerleri gibi bir takım ekonomik göstergelerin Ocak96-Aralık01 dönemine ilişkin değerleri derlenmiştir.

En iyi faktör değerleri Taguchi yöntemi ile belirlenen YSA kullanılarak gerçekleştirilen tahminlemeden elde edilen sonuçlar; hem rassal olarak tasarlanmış bir YSA modelinin sonuçları hem de uzun dönem tahmin yapmaya imkan veren istatistiksel yöntemlerden olan regresyon analizinin sonuçları ile karşılaştırılmıştır.

2. YAPAY SİNİR AĞLARI

İnsan beyninin düşünme, problem çözme ve tanıma gibi yeteneklerini sergileyecek biçimde geliştirilmiş bir matematiksel model olan YSA, insan beynindeki işlem elemanlarını temsil eden birimlerin bir araya gelmesi ile oluşur (Simpson, 1990). Bir YSA; bağlantı mimarisi (ağ topoloji), hesaplama dinamiği (etkinlik fonksiyonları) ve öğrenme dinamiği (öğrenme kuralları) olmak üzere üç elemandan oluşmaktadır (Masson ve Wang, 1990). Çok katmanlı bir YSA Şekil 1'de verilmiştir.



Şekil 1. Çok Katmanlı Bir YSA

Ağ yapısı, Şekil 1'de gösterildiği gibi, işlem elemanlarının yer aldığı katmanları ve katmanlar ve/veya katmanlarda yer alan işlem elemanları arasındaki bağlantıları içerir. Her YSA'da dış kaynaklardan gelen uyarıları alan bir giriş katmanı ve ağda türetilen uyarıları dışarıya göndermeyi sağlayan bir çıkış katmanı bulunur. Çözümü araştırılan probleme bağlı olarak belli sayıda işlem elemanının yer aldığı girdi ve çıktı katmanları arasında bulunan ve dış ortamla hiçbir etkileşimi bulunmayan ve gizli katman olarak tanımlanan bir katman yer alabilir (Simpson, 1990). İşlem elemanları arasındaki bağlantıların her biri, bağlantının gerilimini veya gücünü ve etkileşimini gösteren, pozitif veya negatif değer alabilen ağırlıklar ile temsil edilir.

Hesaplama dinamiği, çok sayıda birbiri ile bağlantılı işlem elemanlarının dışarıdan aldığı girdi desenlerini (uyarıları) işleyerek bu girdilere cevap niteliğinde bir çıktı deseni türetilmesine dayanır. Her bir işlem elemanının çıktısı, kendisine gelen uyarıların ağırlıklı toplamının belirli veya rassal nitelikli transfer fonksiyonlarından geçirilmesi ile elde edilir. Transfer fonksiyonu uygulamaya bağlı olarak seçilir (Caudill, 1988). Öğrenme dinamiği ise, sistemin bir bütün olarak istenen işlevi yerine getirecek şekilde bilginin depolandığı işlem elemanları ve/veya katmanlar arasındaki ağırlıklarının girdi-çıkış ilişkilerini en iyi düzeyde öğrenecek biçimde ayarlanması sürecidir (Simpson, 1990).

Bir YSA'nın performansı, sözü edilen öğrenme kurallarında yer alan ve kullanılan ağın yapısına bağlı olan bir dizi faktörden etkilenmektedir. Çok katmanlı bir YSA'daki katman sayısı ve katmanlarda yer alan işlem elemanı sayısı, ağın eğitilmesinde kullanılan genelleştirilmiş delta kuralında yer alan ve toplam hatayı en küçükleyecek biçimde öğrenme sırasındaki ilerleme hızını temsil eden öğrenme katsayısı ile öğrenme sırasında ilerleme yönünü veren momentum terimi bu tür faktörlerdir (Zurada, 1992; Anagün ve Liou, 1993). Öğrenme katsayısının düzgün bir yakınsamayı garanti edecek şekilde 0 ile 1 arasında seçilmesi önemli ve gereklidir. Öğrenme katsayısının büyük olması, ağın bütünsel en iyiye yakınsaması yerine bir yerel en iyi noktasına takılmasına neden olmaktadır. Öte yandan küçük öğrenme katsayısı yakınsamayı yavaşlatacağından ardıştırma sayısı artmaktadır (Fu, 1994).

Çok katmanlı bir YSA'da bir gizli katman kullanılması durumunda her türlü ilişki incelenebilmektedir (Burr, 1988). Sahip olduğu katman ve işlem elemanı sayısı yönüyle yeterince büyük olan bir ağ, rassal olarak seçilmiş bir karmaşık ilişkiyi temsil edebilmektedir. Ancak, ağın verilen problem için en iyi sonuçları verecek büyüklükte veya yapıda belirlenmesi gerekmektedir. Gizli katmandaki işlem elemanlarının sayısının az veya fazla olması, ağın ya bir iyiye yakınsamasını engelleyerek salınım yapmasına, dolayısıyla girdi-çıktı desenleri arasındaki ilişkiyi öğrenememesine ya da girdi-çıktı desenlerini ezberlemesine ve dolayısıyla başarının düşmesine neden olmaktadır (Bailey ve Thompson, 1990; Klimasauskas, 1989). Bir başka ifadeyle, YSA, veri yapısını kavrayacak ve sonuçta probleme ilişkin modeli oluşturacak bir büyüklükte olmalıdır.

Çok katmanlı YSA'da, en hızlı iniş algoritması mantığında geliştirilmiş geriyayılım algoritması kullanılmaktadır. Geriyayılım algoritmasında, adım büyüklüğüne ek olarak, ağın bir önceki ardıştırmadaki adım değişimlerini dikkate alarak uygun yönde bir yerel en iyi noktada takılmadan hata yüzeyinin en küçük noktasına doğru inilmesine imkan veren momentum terimi yer almaktadır. Öğrenme katsayısında olduğu gibi, 0 ile 1 arasında değer alan momentum terimi de çözümü araştırılan probleme uygun olarak belirlenmek durumundadır (Zurada, 1992).

Bütün bu faktörler, ağın genelleme yeteneği üzerinde etkilidirler. Genelleme, ağın, eğitim esnasında kullanılan veriler ile test aşamasında ağa girilen daha önce görmediği verilere karşı ürettiği sonuçların tutarlılığı veya uygunluğu ile ilişkilidir ve ağın yapısına büyüklüğüne, öğrenme algoritmasına, problemin karmaşıklığına, eğitim verilerinin kalitesine ve sayısına bağlıdır (Fu, 1994). Dolayısıyla, iyi bir genelleme için; veri sayısının ne olması, nasıl bir öğrenme algoritmasının kullanılması, ilgili faktörlerin değerlerinin ne olması, ağın nasıl bir yapı ve büyüklükte olması gerektiği çözümü aranan probleme uygun olarak araştırılmalıdır (Anagün, 1999).

Her biri kendisine ait ağ yapısına, hesaplama ve öğrenme dinamiğine sahip olmakla birlikte, YSA'lar kullandıkları öğrenme kurallarına bağlı olarak iki grupta incelenirler ve desen tanıma, sınıflandırma ve tahminlemeden, görüntü ve konuşma işleme, robot uygulamaları ve eniyileme konulara kadar bir çok alanda yoğun biçimde kullanılmaktadırlar (Siyahi ve Anagün, 1998).

İlgilenilen problem(ler)in çözümünde klasik yöntemlerin kullanılması, kimi durumlarda; zaman, para, donanım veya personel yönüyle pahalı ve elde edilen hassasiyet yönüyle yetersiz olabilir. YSA'lar, klasik yöntemlerin olmadığı, uygulanmasının uygun olmadığı veya uygulanması halinde elde edilen sonuçların yeterli olmadığı durumlarda, ilgilenilen problem(ler)in çözümünde etkilidirler (Caudill, 1993). Problemin yapısına olan bağımlılığın yanı sıra, YSA'lar aşağıda verilen yetenekleri nedeniyle klasik yöntemlere göre tercih edilmektedirler (Klimasauskas, 1989; Fu, 1994; Bose ve Liang, 1996):

- YSA'lar verilerden hareketle bilinmeyen ilişkileri veya veriler arasındaki içsel dinamikleri öğrenebilmekte ve kendilerini organize ederek yenileyebilmektedirler.
- Bellek ve işlem elemanı iç içe olup, paralel ve asenkronize bir çalışma söz konusudur.
- Bilgiler işlem elemanları arasındaki ağırlıklarda saklanır ve eğitim süreci uzun olmakla birlikte, test aşamasında reaksiyon zamanı oldukça kısadır.
- YSA'lar doğrusal olmadıkları için, karmaşık problemlere doğrusal tekniklerden daha iyi ve etkin biçimde çözüm üretirler.
- YSA'lar, işlem elemanları arasında yer alan ağırlıklardaki bozulma nedeniyle oluşacak hatayı tolere edecek, verilerde eksiklik olduğunda uygun sınıflandırmayı yapacak ve karar aşamasında bilginin yetersiz olması halinde birikimini kullanarak sonuç üretecek yetenektedirler.
- YSA'ların hafızasında yer alan verilerin özellikleri ile yeni karşılaşılan bir duruma ilişkin verinin ilişkilendirilmesi olarak tanımlanan genelleme yetenekleri vardır.

3. HİSSE SENEDİ FİYATI TAHMİN PROBLEMİ

Hisse senetlerinin fiyatları rekabetçi bir ortamda arz ve talebe göre belirlenir. Hisse senetlerinin fiyatları; ekonomik faktörler, psikolojik faktörler ve diğer faktörler olmak üzere üç tür faktörden etkilenmektedir (Harrington, 1987).

Ekonomik faktörler içinde en önemlisi bileşik endekstir. Hisse senedi fiyatını etkileyen diğer ekonomik faktörlerin belirlenmesi amacıyla çeşitli çalışmalar yapılmıştır. Berberoğlu vd (1992) tarafından yapılan araştırmaların sonucunda dolar kuru, mark kuru ve enflasyon oranının hisse senedi fiyatını etkilediği ortaya çıkmıştır. Hisse senedi fiyatını etkileyen diğer bir faktör de psikolojik faktördür. Firmalarla ilgili çıkan söylentiler, hükümet bunalımları, liderlerin vefatı, firmaların mali ve idari yapılarıyla ilgili çeşitli söylentilerin varlığı ve kişilerin tepkileri arz ve talep üzerinde olumlu ya da olumsuz etki yapabilmektedir (Sarı, 1992).

Bunların dışında hisse senedi fiyatları üzerinde takvim etkileri de söz konusu olabilmektedir. Takvim etkileri, sadece bir sektöre ait hisse senedi fiyatlarını etkileyen

mevsimsel etkiler veya bütün hisse senetlerinin fiyatlarını etkileyen belirli aylar veya günler olabilir. İMKB' de Ocak ayında görülen dalgalanma, diğer aylarda görülen dalgalanmaya göre daha yüksektir (Balaban vd., 1996). Benzer şekilde, hafta başı (Pazartesi Salı günleri) artış oranları, hafta sonu (Perşembe ve Cuma günleri) artış oranlarından genellikle daha düşüktür (Seler, 1996). İçeriden bilgilendirme de (firma içinden bazı bilgilerin sızdırılması) hisse senedi fiyatlarını etkileyen diğer bir faktördür.

Hisse senedi fiyatının gelecekteki seyrini ve böylece bir hisse senedinin uygun alım satım zamanını tahmin etme çabası, yatırımcıların en ısrarlı çabalarından biri olmuştur. Bu arayış, bilimsel olandan gizemli ve büyüleyici olana kadar birçok yöntem ortaya çıkarmıştır. Bugün, özellikle gelişmiş ülkelerde birçok ekonomik ve sosyal değişkeni dikkate alan oldukça kompleks ekonometrik modellerle hisse senedi fiyatını tahmin etmeye çalışan yatırımcılar bulunmaktadır (Özçam, 1996).

Hisse senedinin fiyatını ekonomik faktörlerden etkilenen arz ve talep belirlemektedir. Bu nedenle hisse senedinin fiyatını; işlem adedi, işlem hacmi, dolar kuru, mark kuru, bileşik endeks, TEFE, TÜFE gibi ekonomik göstergelerin değerlerindeki değişimlerin hisse senedi fiyatı hakkında bilgi verebileceği kabul edilmektedir.

4. KOÇ HOLDİNG HİSSE SENEDİ FİYAT TAHMİNİNE İLİŞKİN VERİLERİN DÜZENLENMESİ

Koç Holding hisse senedi fiyatı ve tahmini için tercih edilen faktörlerin Ocak96-Aralık01 dönemine ilişkin değerleri; işlem adedi ve işlem hacmi için www.analiz.com internet sitesi; dolar-mark kurları için Yapı Kredi Bankası Yatırım Danışmanlığı Servisi; TEFE-TÜFE oranları için Garanti Bankası Yatırım Danışmanlığı Servisi; bileşik endeks için ise T.C. Merkez Bankası aracılığıyla elde edilmiştir.

Amaç geçmiş aylara ait verileri kullanarak izleyen ayda hisse senedinin fiyat değerini tahmin etmektir. Literatürde bu şekildeki tahminlerde, hisse senedinin fiyatını tahmin etmek için önceki üç aya ilişkin fiyat değerleri de alınmaktadır (Altuğ, 1994). Böylece tahmin yapılırken, belirtilen ekonomik göstergelere ek olarak, hisse senedinin sadece bir önceki aydaki değil, son üç aydaki fiyat seyrine de bakılmakta ve fiyatın o dönemdeki eğilimleri değerlendirmeye katılmaktadır. Buna göre; (t), (t+1), (t+2) aylarındaki hisse senedi fiyat verileri de girdi grubuna dahil edilerek, (t+3) ayındaki fiyat çıktı olarak incelenmektedir. Bu düşüncenin etkisiyle, Koç Holding için dikkate alınan ekonomik göstergelere ek olarak fiyat değişkenlerini içerecek üç sütun oluşturulmuştur. Her yeni sütun bir öncekinden bir ay ilerisini gösterecek şekilde kaydırma yaparak elde edilmiştir. Hisse senedinin (t) ayı fiyatının yanına, (t+1) ve (t+2) sütunları, en sona (t+3) ayı sütunları eklenerek, hisse senedi fiyatını etkileyecek toplam değişken sayısı 10'a çıkarılmıştır.

Derlenen verilerdeki eğilim ve şiddetli dalgalanmaların etkisini azaltmak, aynı zamanda da YSA'ya uygun veri biçimini elde edebilmek için, orijinal veriler çeşitli

dönüşümlere tabi tutulmuştur. Transfer fonksiyonu olarak seçilen hiperbolik tanjant fonksiyonun $[-1, 1]$ aralığında değer alıyor olmasından dolayı mevcut verilerin değişim oranları;

$$C_i = \frac{[b_i - b_{(i-1)}]}{b_{(i-1)}}; \quad i = 1, 2, 3, \dots, n \quad (1)$$

eşitliği kullanılarak hesaplanmış ve böylece derlenen tüm verilerin değerleri transfer fonksiyonu için uygun olan tanım aralığına indirgenmiştir. Hesaplama kullanılan yöntem nedeni ile Ocak96 ve Aralık01 dönemlerine ilişkin değişim oranları hesaplanamamıştır. Değişim oranlarına göre düzenlenmiş verileri örneklemek amacıyla Şubat96-Haziran96 dönemi verileri Tablo 1’de verilmiştir.

Tablo 1. Şubat96-Haziran96 Dönemi İçin Değişim Oranlarına Göre Düzenlenmiş Veriler

Tarih	(t)	(t+1)	(t+2)	İ. Adedi	İ. Hacmi
Şubat96	0.032079	-0.12625	0.012601	0.575525	0.781621
Mart96	-0.12625	0.012601	0.093867	-0.32058	-0.23225
Nisan96	0.012601	0.093867	0.195018	1.906775	3.291835
Mayıs96	0.093867	0.195018	0.370742	0.374172	0.941644
Haziran96	0.195018	0.370742	0.45854	-0.23846	-0.40805
B. Endeks	Dolar	Mark	TÜFE	TEFE	(t+3)
0.203814	0.009569	0	0.024691	0.021053	0.093867
0.212395	0.052125	0.075089	-0.45783	-0.40206	0.195018
0.037645	0.077744	0.063193	0.244444	0.224138	0.370742
-0.0562	0.05	0.010574	0.089286	0.140845	0.45854
0.042641	0.051108	0.049714	-0.2623	-0.49383	-0.20269

Tablo 1’de görüldüğü gibi, ekonomik göstergeleri ve ilgili dönemlere ilişkin fiyat değişimlerini içeren 10 ve tahmin edilmek istenen (t+3)’üncü döneme karşı gelen bir değişkenden oluşan veri dosyası elde edilmiştir. Girdi katmanında 10 girdi ve çıktı katmanında 1 işlem elemanına sahip bir YSA ile yapılacak denemelerde kullanılmak üzere, toplam 70 veri deseninin; ilk 59 tanesi (Şubat96-Aralık00 dönemi) eğitim, son 11 tanesi de (Ocak01-Kasım01 dönemi) test verisi olarak iki gruba ayrılmıştır.

5. TAGUCHI YÖNTEMİ İLE YAPAY SINIR AĞI PERFORMANSINA ETKİ EDEN FAKTÖRLERİN UYGUN DEĞERLERİNİN BELİRLENMESİ

Taguchi, çok sayıda değişkenin az sayıda deney ile süreç performansı üzerindeki etkisinin araştırılmasında, diğer tasarımlara kıyasla daha etkin olan ortogonal dizilerin kullanımını önermektedir (Antony ve Roy, 1999). Öte yandan, Taguchi tarafından geliştirilen ve hem tam hem de kesirli faktöriyel tasarımlarda kullanılan Taguchi yöntemi; özellikle yüksek mertebeden etkileşimlerin incelenmesi, rassallaştırma ve bloklara ayırma gibi konularda yetersizliği nedenleriyle daha ziyade üç düzeyli tasarımlarda kullanılmaktadır (Fowlkes ve Creveling, 1995).

Gerek kalite tasarımında ve gerekse hem ürünün imalatı aşamasında hem de kullanımında söz konusu olan değişkenliği azaltacak şekilde mükemmel (robust) tasarım elde edilmesi için, Ross (1988), Taguchi yöntemi mantığına uygun yedi adımdan oluşan bir yaklaşım önermektedir. YSA performansını etkileyen faktörlerin uygun değerlerinin araştırıldığı bu çalışma, ilgili yaklaşıma uygun olarak gerçekleştirilmiştir.

Adım 1 : Değerlendirilecek Faktör ve/veya Etkileşimlerin Seçilmesi

Bu adımda, söz konusu olan problem ve bu problemin çözümüne ilişkin amaç ortaya konulduktan sonra; beyin fırtınası, süreç akış şeması, süreç haritası ve sebep-sonuç diyagramı gibi yöntem ve teknikler kullanılarak, ilgilenilen performans karakteristiğine etkisi olan faktör ve/veya bileşik etkiler seçilmektedir.

Çalışmada; YSA'nın eğitim aşamasında farklı faktör düzeyleri için türettiği ortalama hata değeri performans karakteristiği olarak seçilmiş ve seçilen karakteristiğin en küçülenmesi amaçlanmıştır. YSA'larda böyle bir performans karakteristiği bir dizi faktörden etkilenmektedir. Çalışma kapsamında; kabul edilebilir hata düzeyi denemeler boyunca 0.05 değerinde tutularak, verilere uygunluğu nedeniyle transfer fonksiyonu için hiperbolik tanjant seçilerek ve bir gizli katman içeren YSA oluşturularak dört faktör dikkate alınmıştır.

Adım 2 : Faktör Düzeylerinin Seçilmesi

YSA'nın performansına etki eden faktörlere ilişkin olarak literatürde yapılan açıklamalar ışığında; öğrenme katsayısı ve momentum terimi için [0.1, 0.5], gizli katmandaki işlem elemanı sayısı için [3, 7] aralığında değerler atanmıştır. Eğitim algoritması faktörü için ise, denemelerde kullanılan Qwicknet programında yer alan ve öğrenme katsayısı, momentum terimi gibi YSA'nın hesaplama dinamiğini oluşturan faktörlerin kullanıcı tarafından belirlendiği üç algoritma seçilmiş ve faktörlerin düzeyleri Tablo 2'de verilmiştir:

Tablo 2. Seçilen Faktörler ve Düzeyleri

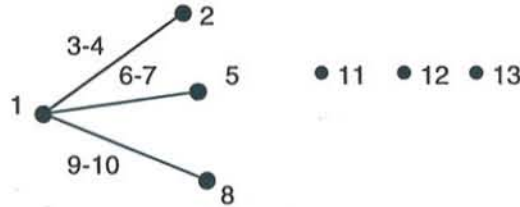
Faktör	1. Düzey	2. Düzey	3. Düzey
A Öğrenme Katsayısı	0.1	0.3	0.5
B Momentum Terimi	0.1	0.3	0.5
C Gizli Katman İşlem Elemanı Sayısı	3	5	7
D Eğitim Algoritması	Geriyayılım	Rassal geriyayılım	Parti geriyayılım

Adım 3 : Uygun Ortogonal Dizinin Seçilmesi

Bu adımda, Taguchi tarafından tasarlanmış ortogonal dizilerden, ilgilenilen performans karakteristiğine etki eden faktörler ve düzeyleri dikkate alınarak bir ortogonal dizi seçilmektedir. Ortalama hata değerine üç-düzeyle dört faktörün etkili olduğu dikkate alındığında; $L_{81}(3^{40})$ ve $L_{27}(3^{13})$ şeklinde iki farklı tasarım alternatifleri söz konusu olmaktadır. Bu çalışmada, yüksek mertebeden bileşik etkiler göz önüne alınmadığından, $L_{27}(3^{13})$ ortogonal dizisi seçilmiştir.

Adım 4 : Faktör ve/veya Bileşik Etkilerin Sütunlara Atanması

Faktör ve/veya bileşik etkilerin seçilen ortogonal diziyeye göre sütunlara atanmasında, Taguchi tarafından geliştirilen doğrusal grafikler (linear graphs) veya üçgen tablolar (triangular tables) kullanılmaktadır. İncelenmek istenen faktör ve/veya bileşik etkilerin sayısına bağlı olarak kesirli faktöriyel tasarıma yönelik farklı tasarımlar için geliştirilmiş doğrusal grafikler söz konusudur. Seçilen doğrusal grafik, etkisi araştırılmak istenen faktör ve/veya bileşik etkiler dikkate alınarak uygun şekilde düzenlenebilmektedir. $L_{27}(3^{13})$ ortogonal dizisi için düzenlenen doğrusal grafik Şekil 2’de verilmiştir.



Şekil 2. $L_{27}(3^{13})$ Ortogonal Dizisi İçin Kullanılan Doğrusal Grafik

İlgilenilen performans karakteristiği için, Şekil 2’de verilen doğrusal grafik dikkate alınarak; asıl faktörler (A, B, C ve D) sırasıyla; 1., 2., 5. ve 8. sütunlara, bileşik etkiler (AB, AC, AD) sırasıyla; 3-4, 6-7, 9-10 numaralı sütunlara karşılık gelecek şekilde tasarım matrisi oluşturulmuş ve 11-13 numaralı sütunlar hata terimine aktarılmak üzere boş bırakılmıştır.

Adım 5 : Testlerin Yapılması

Bu adımda, belirlenen tasarım temelinde deneyler gerçekleştirilmekte ve seçilen performans karakteristiğine ilişkin gözlem değerleri elde edilmektedir. Belirlenen faktör ve düzeyleri dikkate alınarak toplam 27 deney yapılmış ve YSA ile eğitim süreci tamamlandığında elde edilen eğitim hatası değerleri kaydedilmiştir.

Adım 6 : Sonuçların Analiz Edilmesi

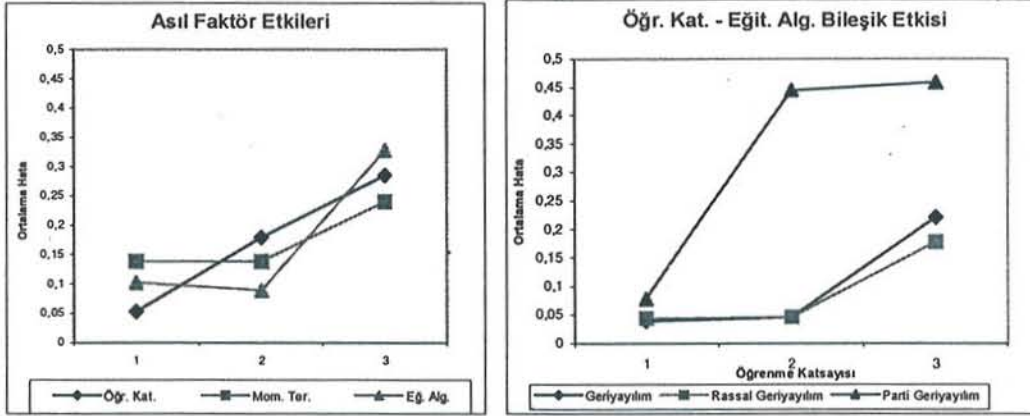
Elde edilen gözlem değerlerinin analizinde, varyans analizi ve asıl faktör ve/veya bileşik etkilerin düzeylere göre ortalama etkilerini gösteren grafikler kullanılmaktadır. Veriler için ANOVA programı ile elde edilen varyans analizi sonuçları Tablo 3'de yer almaktadır.

Tablo 3. Varyans Analizi Sonuçları

Kaynak	Birleştirildi?	SS	sd	MS	F	% katkı
A	H	0.24284	2	012142	16.49728	24.71
B	H	0.06274	2	0.03137	4.26223	5.20
AB	H	0.07111	4	0.01778	2.41576	4.51
C	E	0.00378	2			
AC	E	0.03088	4	0.00772		
D	H	0.32300	2	0.16150	21.94293	32.40
AD	H	0.13503	4	0.03376	4.58696	11.44
HATA	-	0.08836	12	0.00736		20.73
TOPLAM	-	0.92309	26			

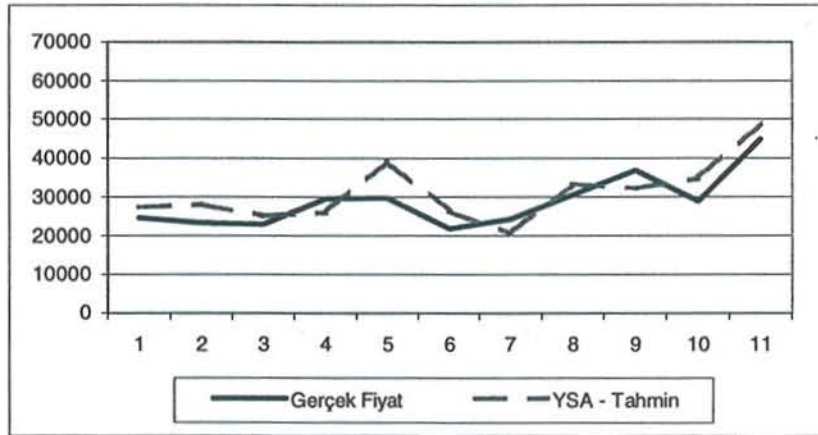
Elde edilen F değerleri, ilgili serbestlik dereceleri dikkate alınarak %5 anlam düzeyinde test edilecek olursa, $F_{0,05;2;12}=3.89$ ve $F_{0,05;4;12}=3.26$ için; A, B ve D faktörleri ile AD bileşik etkisinin kritik olduğu sonucuna ulaşılmaktadır. Öte yandan, Tablo 3'ün son sütununda yer alan ve varyans analizi tablosundaki her bir bileşenin toplam varyans içindeki payını ifade eden % katkı değerleri dikkate alınır, kritik olan faktörler ve bileşik etkinin uygun değerlerinin belirlenmesi ile toplam değişkenlikte %74.75 oranında azalmanın sağlanabileceği görülmektedir.

Kritik olduğuna karar verilen faktör ve bileşik etkilerin uygun değerlerinin belirlenmesi için ortalama etki grafikleri dikkate alınmıştır. Varyans analizine göre kritik olan asıl faktör ve bileşik etkilere ilişkin ortalama etki grafikleri Şekil 3'de verilmiştir.



Şekil 3. Kritik Olan Faktör ve Bileşik Etkilerin Ortalama Etki Grafikleri

Yapılan deneyler ve ilgili grafiklerin analizi sonucunda; asıl faktörlerden öğrenme katsayısı, momentum terimi ve eğitim algoritması ile öğrenme katsayısı-eğitim algoritması bileşik etkisinin performans karakteristiği olan ortalama hata üzerinde kritik etkiye sahip olduğu belirlenmiştir. Kritik faktörlerin uygun değerleri dikkate alınarak belirlenen A1B1C3D2 bileşimi (öğrenme katsayısı: 0,1, momentum terimi: 0,1, işlem elemanı sayısı: 7, algoritma: rassal geri yayılım) kullanılarak yapılan tahminleme sonuçları (ortalama tahmin hatası %15,13) ile gerçekleşen fiyat değerleri karşılaştırmalı olarak Şekil.4'de verilmiştir.



Şekil 4. Gerçek ve Tahmini Fiyatların Karşılaştırılması (YSA)

Adım 7 : Doğrulama Denei(ler)inin Yapılması

Gerçekleştirilen uygulamanın bir ürünün imalatı aşamasındaki değişkenliği azaltarak mükemmel bir tasarım elde edilmesi şeklinde olmaması nedeniyle, bu adımda, uygun faktör değerleri belirlenmiş YSA kullanılarak ilgili döneme ilişkin olarak yapılan tahminler ile gerçekleşen fiyat değerleri karşılaştırılmıştır. Tahminlerdeki ortalama %15.13 tahmin hatası ve gerçekleşen fiyat değerlerine uygun bir seyrin elde edilmiş olması YSA'nın performansını etkileyen faktörlerin değerlerinin uygun olarak belirlendiğini ve seçilen hisse senedinin izleyen dönemlerdeki fiyat değerlerinin tahmininde kullanılabileceğini göstermektedir.

YSA'nın performansına etki eden faktörlerin açıklandığı gibi sistematik bir şekilde belirlenmesi yerine, rassal olarak tasarlanmış YSA ile fiyat tahminlerinin yapılması halinde, tahmin sonuçlarının faktör değerlerinden etkilenip etkilenmediğini araştırmak amacıyla; öğrenme katsayısı 0.3, momentum terimi 0.1, gizli katmandaki işlem elemanı sayısı 7 ve eğitim algoritması parti geriyayılım olan bir YSA ile aynı döneme ilişkin tahminler yapılmış ve ortalama tahmin hatası %34.42 olarak elde edilmiştir.

Taguchi yöntemi kullanılarak, YSA performansına etki eden faktörlerin en iyi değerlerini belirlemek amacıyla gerçekleştirilen çalışmanın, tahmin sonuçları üzerindeki etkisi, hisse senedi fiyat tahmin problemi için %1-2 mertebesinde bir iyileşmenin bile önemli olduğu dikkate alınır, daha net anlaşılmaktadır.

6. ÇOKLU DOĞRUSAL REGRESYON MODELİ İLE KOÇ HOLDİNG HİSSE SENEDİ FİYAT TAHMİNİ

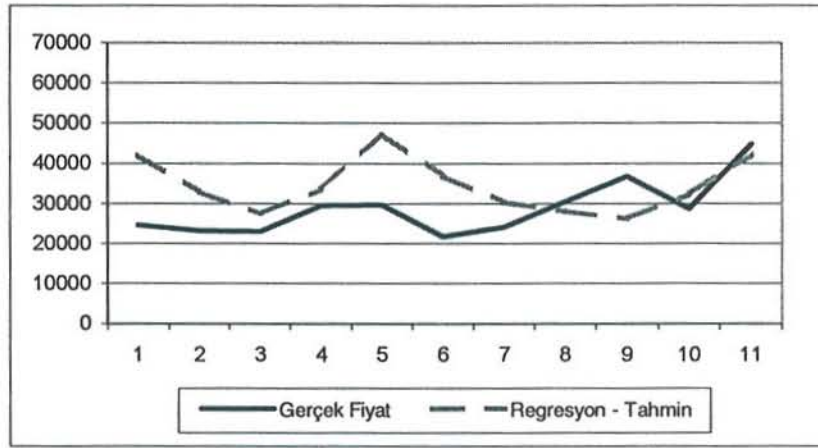
Hisse senedi fiyatı üzerinde ekonomik göstergeler olarak adlandırılan; işlem adedi ve hacmi, dolar ve mark kuru, TEFE-TÜFE oranları, bileşik endeks değeri ve senedin önceki üç aya ilişkin fiyat değerleri gibi göstergelerin etkisini incelemek ve YSA ile elde edilmiş tahmin sonuçlarıyla karşılaştırmak amacıyla çoklu doğrusal regresyon modeli oluşturulmuştur.

Çoklu doğrusal regresyon modeli ile seçilen hisse senedi fiyatının tahmini işlemi için hazırlanan veri gruplarından, fiyatın izleyen ay gerçekleşecek değeri ile arasındaki korelasyon ilişkisi en yüksek olan değişkenlerin belirlenebilmesi için stepwise regresyon uygulanmış ve Minitab paket programıyla yapılan analiz sonrasında; değişken adları ve katsayı değerleri Tablo.4'de belirtilen regresyon modeli elde edilmiştir.

Tablo 4. Çoklu Doğrusal Regresyon Analizi Sonuçları

Bağımsız Değişken	Değişken	Katsayılar
Sabit		2365.1
(t+2)	X ₃	0.38934
İŞLEM ADEDİ	X ₄	-0.00001551
İŞLEM HACMİ	X ₅	0.00000115
B.ENDEKS	X ₆	1.2197

Elde edilen modelin belirlilik katsayısı %98.56 olarak gerçekleşmiştir. Regresyon denkleminin tutarlılık testi (p=0.000) denklemin bağımlı ve bağımsız değişkenler arasındaki ilişkiyi açıklayabilir nitelikte olduğunu göstermektedir. Belirlenen regresyon modeli ile yapılan tahminlemede ortalama hata %32.30 olarak gerçekleşmiştir. Çoklu doğrusal regresyon modelini kullanılarak ilgili döneme (Ocak01-Kasım01) ilişkin elde edilen tahmin sonuçları ile gerçekleşen fiyat değerleri Şekil 5’de verilmiştir.



Şekil 5. Gerçek ve Tahmini Fiyatların Karşılaştırılması (Regresyon Analizi)

7. SONUÇ VE ÖNERİLER

Taguchi yöntemi kullanılarak YSA performansını etkileyebilecek faktörlerin uygun değerleri belirlenmiş ve İMKB’de işlem gören Koç Holding hisse senedinin fiyatı, Ocak96-Aralık01 dönemine ilişkin olarak derlenmiş ekonomik göstergeler dikkate alınarak, YSA yaklaşımıyla tahmin edilmeye çalışılmıştır.

Performansı etkileyen faktörlerin değerleri rassal olarak belirlenmiş bir YSA ile yapılan tahminleme sonucunda %34.42 düzeyinde bir tahmin hatası söz konusu olmakla birlikte, ilgili faktörlerin uygun değerleri Taguchi yöntemiyle belirlenmiş YSA modeli ile yapılan tahminlerde %15.13 düzeyinde bir tahmin hatası elde edilmiştir. Belirlilik katsayısı %98.56 olan bir regresyon modeli kullanılarak, aynı döneme ilişkin olarak yapılan tahminlerde %32.30 düzeyinde tahmin hatası gerçekleşmiştir. Her bir uygulamadan elde

edilen tahmin hataları, verilerin eksik veya aşırı sapmalar gösterdiği durumlardaki performansı ve doğrusal olmayan ilişkileri öğrenebilme yeteneği sayesinde uygun faktör değerleri belirlenmiş YSA'nın, regresyon analizine kıyasla daha tutarlı tahminler yapabileceğini göstermektedir.

Taguchi yöntemi kullanılarak belirlenen kritik faktör ve bileşik etkileri ile farklı transfer fonksiyonunu ve gizli katman sayısını içerecek şekilde yeni bir tasarım oluşturularak, faktör düzeylerinin değer alabildikleri değişim aralığı daraltılıp daha hassas değerler bulunabilir ve YSA'nın performansı iyileştirilebilir. Bu şekilde YSA'nın daha gerçekçi değerler türeterek tahmin yapması sağlanabilir. Uygun olan göstergeler için aylık veriler yerine günlük veya haftalık veri kullanılması, veriler arasındaki aşırı dalgalanmalar azalacağından, tahmin hatasını azaltıcı yönde etki yaratabilir. Ayrıca, hisse senedi fiyatının diğer hisse senetlerinin fiyatlarındaki değişimden etkilenme durumu, çalışmaya eklenecek yeni veriler ile incelenebilir. Geçmiş dönemlerdeki veriler incelenerek büyük dalgalanmaların sebepleri araştırılabilir ve mümkün olduğu ölçüde önemli olaylar değerlendirilip sayısallaştırılarak bu veriler de girdi desenlerine dahil edilebilir.

Özellikle hisse senedi fiyatını etkileyen ekonomik ve/veya psikolojik faktörlerin değerlendirilmesindeki yetersizliğin etkisini azaltacak biçimde, YSA ile etkileşimli bir şekilde çalışacak bir uzman sistem tasarlanabilir. Bir anlamda melez (hybrid) sistemler olarak ifade edilen böylesi bir yapı ile, uzman sistemlerin sahip olabildiği uzman bilgisiyile, YSA'ların doğrusal olmayan ilişkileri öğrenebilme ve çıkarsama yapabilme yetenekleri birleştirilerek daha başarılı sistemler oluşturulabilir.

KAYNAKLAR

- AIKEN, M. (1999), *Using A Neural Network to Forecast Inflation, Industrial Management and Data Systems*, 99: 7, 296-301.
- ALTUĞ, S. (1994), *Price Prediction in İMKB Using Neural Networks*, MBA Thesis, Bilkent University, Ankara.
- ANAGÜN, A.S. (1999), *Bilgi Güvelliğinin Sağlanmasında Kullanıcı Özelliklerine Dayalı Bir Yapay Sinirsel Ağ Yaklaşımı*, *Endüstri Mühendisliği*, 10: 4, 3-11.
- ANAGÜN, A.S., LIOU, Y.H.A. (1993), *A Neural Network Application for Apnea Recognition: A Preliminary Study*, *ASME Intelligent Engng System through Artificial Neural Networks*, 3, 321-326.
- ANTONY J., ROY, R.K. (1999), *Improving The Process Quality Using Statistical Design of Experiments: A Case Study*, *Quality Assurance*, 6, 87-95.
- BAILEY, D., THOMPSON D. (1990), *How to Develop Neural Network Applications*, *AI Expert*, 38-47.

- BALABAN, E., CANDEMİR, H.B., KUNTER, K. (1996), *İstanbul Menkul Kıymetler Borsasında Aylık Dalgalanma Tahmini, Sermaye Piyasası ve İMKB Üzerine Çalışmalar*, İşletme ve Finans Yayınları, Ankara.
- BERBEROĞLU, N., ARSLAN, S., AFŞAR, M. (1992), *Hisse Senetlerinde Değerleme Yöntemleri ve Türkiye'de Hisse Senetlerinin Fiyatlarını Belirleyen Faktörlerin Analizi*, Anadolu Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 10, 1-34.
- BOSE, N.K., LIANG, P. (1996), *Neural Network Fundamentals with Graphs, Algorithms, and Applications*, McGraw-Hill, New York.
- BURR, D. J. (1988), *Experiments on Neural Net Recognition of Spoken and Written Text*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 36: 7, 1162-1168.
- CAUDILL, M. (1993), *What is a Neural Network?*, AI Expert, 7-14.
- CAUDILL, M. (1988), *Neural Network Premier, Part-I*, AI Expert, 2-7.
- DANIELS, H., KAMP, B. (1999), *Application of MLP Networks to Bond Rating and House Pricing*, Neural Computation and Applications, 8, 226-234.
- ENKE, D. , DIWE, P., VAITIANATHASAMY, S. (2000), *Factorial Design For Developing Feed Forward Neural Network Architectures*, ASME Intelligent Engng System through Artificial Neural Networks, 10, 109-114.
- FARAWAY, J., CHATFIELD, C. (1998), *Time Series Forecasting with Neural Networks: A Comparative Study Using the Airline Data*, Application Statistics, 47: 2, 231-250.
- FOWLKES, W.Y, CREVELING, C.M. (1995), *Engineering Methods for Robust Product Design Using Taguchi Methods in Technology and Product Development*, Canada: Addison-Wesley Publishing Company.
- FU, L. M. (1994), *Neural Networks in Computer Intelligence*, New York: McGraw-Hill.
- GREFENSTETTE, J. (1986), *Optimization of Control Parameters for Genetic Algorithms*, IEEE Transactions on Systems, Man, and Cybernetics, 16, 122-128.
- HARRINGTON, D. (1987), *Modern Portfolio Theory*, New Jersey: Prentice Hall.
- KLIMASAUSKAS, C. C. (1989), *An Introduction to Neural Networks*, Part III: Training a Neural Network, PC AI, 20-24.
- LAW, R. (1998), *Room Occupancy Rate Forecasting: A Neural Network Approach*, International Journal of Contemporary Hospitality Management, 10: 6, 234-239.
- MASSON, E., WANG, Y. J. (1990), *Introduction to Computing and Learning in Artificial Neural Networks*, European Journal of Operational Research, 47: 1, 1-28.

- MCLAUCHLAN, R.A., WECKMAN, G.R., PALLERLA, S., VEGALETI, V. (1999), *Predicting Student Academic Success in the Engineering Curriculum at Texas A&M University Kingville Using Neural Networks*, ASME Intelligent Engng System through Artificial Neural Networks, 9, 1183-1188.
- NASIR, M.L., JOHN, R.I., BENNETT, S.C., RUSSELL, D.M. (2001), *Selecting the Neural Network Topology for Student Modelling of Prediction of Corporate Bankruptcy*, Campus Wide Information Systems, 18: 1, 13-22.
- OLIVERIA, K.A., VANNUCCI, A, SILVA, E.C. (2000), *Using Artificial Neural Networks to Forecast Chaotic Time Series*, Physica A, 284, 393-404.
- ÖZÇAM, F. (1996), *Teknik Analiz ve İstanbul Menkul Kıymetler Borsası, Sermaye Piyasası Kurulu, Ankara.*
- RAGGAD, B.G. (1996), *Neural Network Technology for Knowledge Resource Management, Management Decision*, 34: 2, 20-24.
- ROSS, P.J. (1988), *Taguchi Techniques for Quality Engineering*, New York: McGraw-Hill.
- SARI, Y. (1992), *Borsada Teknik Analiz*, İstanbul: Scala Yayıncılık & Tanıtım A.Ş.
- SELER, İ.T. (1996), *Haftanın Günleri: İMKB' ye Etkileri Üzerine Bir İnceleme, Sermaye Piyasası ve İMKB Üzerine Çalışmalar*, İşletme ve Finans Yayınları, Ankara.
- SIMPSON, P.K. (1990), *Artificial Neural Systems: Foundations, Paradigms, Applications, and Implementations*, New York: Pergamon Press.
- SİYAHİ, B., ANAGÜN, A.S. (1998), *Sismik Sıvılaşma Potansiyelinin Yapay Sinir Ağları ile Belirlenmesi Zemin Mekaniği ve Temel Mühendisliği Yedinci Ulusal Kongresi Bildiriler Kitabı*, 552-561.
- VENUGOPAL, V., BAETS, W. (1994), *Neural Networks and Statistical Techniques in Marketing Research: A Conceptual Comparison*, Marketing Intelligence and Planning, 12: 7, 30-38.
- ZURADA, J. M. (1992), *Introduction to Artificial Neural Systems*, St. Paul: West Publishing.

Analyzing Performance of Artificial Neural Networks by Taguchi Methods: Forecasting Stock Market Prices

ABSTRACT

A wide variety of statistically based forecasting methods have been developed depending on the nature of the problem concerned. An Artificial Neural Network (ANN), one of the artificial intelligence techniques, is considered to be a powerful tool for recognition, classification, forecasting, and optimization. The design of an ANN architecture has been widely recognized as an important issue in the literature. In this paper, Taguchi method was used to analyze the factors that affect the performance of an ANN and to determine appropriate values of the factors regarding to improving the performance of an ANN. A stock, namely Koç Holding, has been chosen from IMKB for an application purpose and the optimized ANN has been applied to forecast prices of the selected stock using data in the period of Jan96-Dec01. In order to determine efficiency of the approach and to emphasize the necessity of performance optimization in ANN, the result obtained was compared with a randomly designed ANN and the multiple regression model.

Key Words: Taguchi Method, Artificial Neural Network, Regression Analysis, Forecasting, Stock.

(α)

İki ve Üç Yönlü Tabloların Gözlerinde 5'den Küçük Beklenen Frekans Olması Durumunda, I. Tip Hata Olasılığının (α) Durumu

Mehmet AKYOL*

Fikret GÜRBÜZ**

ÖZET

Simulasyon çalışmasıyla; 2 ve 3 yönlü tabloların gözlerine genellikle 5'den az ve 5'den fazla frekans düşecek şekilde, eşit ve birbirinden farklı olarak verilen marjinal toplamalar sabit tutularak, uniform tesadüfi sayıları üretilip χ^2 analizi 100000 defa yapılmış ve gerçekleşen I'inci tip hata olasılıkları bulunmuştur. Yapılan hesaplamalar sonucunda gözlerle teorik olarak 2 ve 4 frekans düşecek şekilde belirlenen eşit marjinal toplamalar durumunda, gerçekleşen I'inci tip hata olasılığında büyük bir sapma olmuştur.

Anahtar Kelimeler: İki Yönlü Tablolar, Üç Yönlü Tablolar, Çok Yönlü Tablolar, χ^2 Analizi, Simulasyon.

1.GİRİŞ

Hipotez testlerinde kullanılan olasılık dağılımları, değişkenlerin sürekli veya kesikli olmasına göre değişmektedir. Her zaman sürekli değişkenleri bulmak veya verileri bu şekilde ifade etmek mümkün olmayabilir. Böyle durumlarda özellikleri kategorilerine göre sınıflandırmak, her sınıf içine düşen beklenen frekansı hesaplamak suretiyle ki-kare testi uygulanır.

Yönlü tabloların analizinde kullanılan χ^2 fonksiyonu sürekli bir fonksiyondur. Oysa yönlü tablolarda hesaplanabilecek beklenen frekansların sayısı sınırlıdır. Bu sebeple kesikli bir fonksiyon daha uygun düşer. Örneklem hacmi yeterince büyük olduğunda, kesikli bir dağılımın sürekli bir dağılıma yaklaşık olduğunu kabul etmek, genellikle sonuçlarda fazla hata meydana getirmemektedir. Ancak güvenilir yorumlar yapabilmek için, beklenen frekansların hiç olmazsa 5 olması gerekir. Beklenen frekanslar arasında 5 den küçük olan değerlere rastlandığında, birbirine yakın sınıflar birleştirilerek bu frekanslar büyütülebilir (Serper, 1986).

Beklenen frekansların küçük olduğu, fakat sınıfları birleştirmenin uygun görülmediği hallerde χ^2 formülüne Yates düzeltmesi uygulanır. Sınıfların birleştirmenin uygun görülmediği ve Yates düzeltmesinin yapılmadığı durumda, I.tip hata olasılığının durumunu incelemek amacıyla bu çalışma yapılmıştır.

* (Haberleşme Adresi)Devlet İstatistik Enstitüsü, Ankara, Türkiye

**Ankara Üniversitesi, Ziraat Fakültesi, Zootečni Bölümü, Biyometri ve Genetik Anabilim Dalı, Ankara, Türkiye

2 ve 3 yönlü tabloların gözelerine genellikle 5'den az ve 5'den fazla frekans düşecek şekilde eşit ve birbirinden farklı olarak verilen marjinal toplamlar sabit tutularak, üniform tesadüf sayıları üretilip χ^2 analizi 100000 defa yapılmış ve gerçekleşen I. tip hata olasılıkları bulunmuştur. Bulunan I. tip hata olasılıkları $\alpha=0.05$ ile karşılaştırılmıştır.

2. İKİ VE ÜÇ YÖNLÜ TABLOLAR

2.1.İki (rxc) Yönlü Tablolar

İki özelliğe sahip anakütleden çekilmiş örnek, iki yönlü tabloda, genel olarak Tablo 1'deki gibi gösterilir. N gözlem iki nicel değişkene göre sınıflandırılmıştır. Birinci değişken satırda, r (row) sınıftan ikinci değişken de sütunda, c (column) sınıftan oluşur.

Tablo 1. İki Yönlü Tablonun Genel Formu

		Sütun (değişken 2)				Toplam
		1	2	...	c	
Satır (değişken 1)	1	n_{11}	n_{12}	...	n_{1c}	$n_{1.}$
	2	n_{21}		$n_{2.}$
	
	r	n_{r1}	n_{rc}	$n_{r.}$
Toplam		$n_{.1}$	$n_{.2}$...	$n_{.c}$	$n_{..}=N$

i'inci satır j'inci sütundaki gözlenmiş frekans n_{ij} ($i=1,2,\dots,r$, $j=1,2,\dots,c$) olarak gösterilir. i'inci satırdaki gözlenmiş frekansların toplamı, $n_{i.}$, j'inci sütundaki gözlenmiş frekansların toplamı $n_{.j}$ ile gösterilir. Bunlar marjinal toplamlar olarak ifade edilir. Marjinal toplamları şu şekilde bulmak mümkündür:

$$n_{i.} = n_{i1} + n_{i2} + \dots + n_{ic}$$

$$= \sum_{j=1}^c n_{ij}$$

$$n_{.j} = n_{1j} + n_{2j} + \dots + n_{rj}$$

$$= \sum_{i=1}^r n_{ij}$$

Benzer şekilde genel toplam;

$$n_{..} = \sum_{i=1}^r \sum_{j=1}^c n_{ij}$$

$$= \sum_{i=1}^r n_{i.} = \sum_{j=1}^c n_{.j}$$

şeklinde gösterilir. $n_{..}$ örnekdeki gözlenen frekansların toplamını verir ki genellikle N ile gösterilir.

Anakütlenin iki yönlü olarak verilmesi durumunda beklenen frekanslar F_{ij} ; şu şekilde bulunabilir:

$$F_{ij} = N p_{ij}$$

i 'inci satırdaki olasılık p_i ve j 'inci sütundaki olasılık p_j ile gösterilirse, iki değişken arasındaki bağımsızlık eşitliği şu şekilde verilebilir:

$$p_{ij} = p_i \cdot p_j$$

Bu eşitlik $F_{ij} = N p_{ij}$ eşitliğinde p_{ij} yerine konulursa, eşitlik

$$F_{ij} = N p_i \cdot p_j$$

şekline dönüşür.

İki değişken arasında bağımsızlık testi yapılırken anakütledeki birey sayısı bilinmemektedir. Bundan dolayı gözlenmiş frekanslardan hareketle p_i ve p_j olasılıkları yerine \hat{p}_i ve \hat{p}_j olasılıkları kullanılabilir. Bu olasılıklar gözlenen frekanslar yardımıyla şu şekilde hesaplanabilir.

$$\hat{p}_i = \frac{n_i}{N} \quad \text{ve} \quad \hat{p}_j = \frac{n_j}{N}$$

İki özelliğe sahip anakütleden çekilmiş örneğin ij 'inci gözündeki beklenen frekans değeri en çok olabirlik tahmin yöntemine göre;

$$\begin{aligned} E_{ij} &= N \hat{p}_i \hat{p}_j \\ &= N \frac{n_i}{N} \frac{n_j}{N} = \frac{n_i \cdot n_j}{N} \end{aligned}$$

şeklinde. İki değişkenin bağımsızlığı sözkonusu olduğunda, yukarıda tahmin edilen beklenen frekanslarla gözlenmiş frekansların farklılığı incelenir. Bu, n_{ij} ile E_{ij} frekansları arasındaki farklılıktır.

Bağımsızlığın ileri sürüldüğü H_0 hipotezine göre hesaplanan beklenen frekanslar (E_{ij}) ve gözlem yolu ile elde edilen gözlenmiş frekanslardan (n_{ij}) hareketle Pearson tarafından ileri sürülen χ^2 istatistiği şu şekilde hesaplanır:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

Serbeslik derecesi; s.d. = $(r - 1) (c - 1)$ dir. Karşılıklı bağımsızlık hipotezi için beklenen marjinal toplamların gözlenen marjinal toplamlara ($E_{i.} = n_{i.}$ ve $E_{.j} = n_{.j}$) eşitliliği gibi kısıtlamalar sözkonusudur.

2. Üç (rxcxl) Yönlü Tablolar

Üç yönlü tabloların analizi, iki yönlü tabloların analizi ile karşılaştırılarak verilebilir. rxc yönlü tabloları için kullanılan terminoloji rxcxl yönlü tabloları için de genişletilebilir. Genel olarak üç yönlü tablolar Tablo 2'deki gibi gösterilebilir.

Tablo 2. Üç Yönlü Tabloların Genel Formu

Satır (Değişken 1)	Tabaka (Değişken 3)	Sütun (Değişken 2)			
		1	2	...	c
1	1	n_{111}	n_{121}	...	n_{1c1}
	2	n_{112}	n_{122}	...	n_{1c2}
	⋮	⋮	⋮	⋮	⋮
	l	n_{11l}	n_{12l}	...	n_{1cl}
2	1	n_{211}	n_{221}	...	n_{2c1}
	2	n_{212}	n_{222}	...	n_{2c2}
	⋮	⋮	⋮	⋮	⋮
	l	n_{21l}	n_{22l}	...	n_{2cl}
⋮	1			...	
	2			...	
	⋮			...	
	l			...	
r	1	n_{r11}	n_{r21}	...	n_{rc1}
	2	n_{r12}	n_{r22}	...	n_{rc2}
	⋮	⋮	⋮	⋮	⋮
	l	n_{r1l}	n_{r2l}	...	$n_{rc l}$

Üç yönlü tablolarda r satırları, c sütunları ve l tabakaları gösterir. ijk gözündeki gözlenmiş frekans n_{ijk} olarak verilir. $i=1,2,\dots,r$, $j=1,2,\dots,c$ ve $k=1,2,\dots,l$. Çeşitli alt indisler dikkate alınarak yapılan toplamlar marjinal toplamlar olarak adlandırılır. Örnek olarak i ve j'in tüm değerleri için yapılan toplam, k'ncü tabakanın marjinal toplamını verir. Benzer şekilde her j ve k, her i ve k indislerine ait n_{ijk} değerlerinin toplanmasıyla i'inci satırın ve j'inci sütunun marjinal toplamları elde edilir. Bu toplamlar tek değişkenli marjinal toplamlar olarak bilinir. Bu toplamlar şu şekilde verilebilir:

$$n_{i.} = \sum_{j=1}^c \sum_{k=1}^l n_{ijk},$$

$$n_{.j.} = \sum_{i=1}^r \sum_{k=1}^l n_{ijk},$$

$$n_{.k.} = \sum_{i=1}^r \sum_{j=1}^c n_{ijk}.$$

Gözlenmiş frekansların genel toplamı ($n_{...}$) ise şu şekilde hesap edilir;

$$n_{..} = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^l n_{ijk}$$

ve genellikle N olarak gösterilir.

İki yönlü tablolardaki iki değişkenin bağımsızlığı hipotezine karşılık gelen hipotez, üç yönlü tablolarda üç değişkenin bağımsızlığı şeklindedir. Üç yönlü tabloda değişkenlerin karşılıklı bağımsızlık hipotezi aşağıdaki gibi gösterilebilir. Bu hipoteze göre satırlar, sütunlar ve tabakalar birbirinden tamamen bağımsızdır.

$$H_0: p_{ijk} = p_{i..} \cdot p_{.j.} \cdot p_{..k}$$

p_{ijk} ; ijk gözündeki gözlemlerinin gözlenme olasılığını, $p_{i..}$, $p_{.j.}$ ve $p_{..k}$ satır, sütun ve tabaka değişkenlerinin marjinal olasılıklarını gösterir. Verilen bu hipotez iki yönlü tablolar için kurulan bağımsızlık hipotezinin üç yönlü tablolardaki karşılığıdır. Bu hipotezi test etmek için, iki yönlü tablolarda kullanılan yaklaşım kullanılır. İlk olarak H_0 hipotezinin geçerliliğinde, beklenen frekanslar hesaplanır. Daha sonra, bu frekanslar bilinen χ^2 istatistiğindeki gibi gözlenmiş frekanslarla eşleştirilir. Son olarak serbestlik derecesi ve önemlilik düzeyi belirlenir. Üç değişkenli karşılıklı bağımsızlık hipotezi durumunda örnek beklenen frekansların bulunmasında da iki yönlü tablolar için kullanılan yöntem benzer yöntem kullanılır.

$$E_{ijk} = N \hat{p}_{i..} \hat{p}_{.j.} \hat{p}_{..k}$$

eşitliğindeki $\hat{p}_{i..}$, $\hat{p}_{.j.}$ ve $\hat{p}_{..k}$ olasılıkları $p_{i..}$, $p_{.j.}$ ve $p_{..k}$ olasılıkların tahminleridir. Bu olasılık tahminleri, ilgili değişkenlerin marjinal toplamlarından hareketle hesap edilir.

$$\hat{p}_{i..} = \frac{n_{i..}}{N}, \hat{p}_{.j.} = \frac{n_{.j.}}{N}, \hat{p}_{..k} = \frac{n_{..k}}{N}$$

Bunlar birçok olabilirlik tahminleridir. Marjinal olasılıklar eşitlik $E_{ijk} = N \hat{p}_{i..} \hat{p}_{.j.} \hat{p}_{..k}$, eşitliğinde yerine konulduğunda,

$$E_{ijk} = N \frac{n_{i..}}{N} \frac{n_{.j.}}{N} \frac{n_{..k}}{N}$$

$$= \frac{n_{i..} \cdot n_{.j.} \cdot n_{..k}}{N^2}$$

ijk gözünün beklenen frekansı elde edilmiş olur. Bu eşitlikten yararlanarak beklenen frekanslar hesaplanır. Daha sonra test istatistiği şu şekilde hesap edilir (Everitt, 1994).

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^l \frac{(n_{ijk} - E_{ijk})^2}{E_{ijk}}$$

Üç yönlü tablolar için χ^2 istatistiğinin serbestlik derecesi bilinen yöntem olan aşağıdaki genel eşitlik yardımıyla hesap edilebilir.

$$\begin{aligned} s.d. &= (\text{tablodaki göze sayısı} - 1) - (\text{hipotez testi için verilerden yararlanılarak} \\ &\quad \text{tahmin edilen olasılıkların sayısı}) \\ &= r(c-1) - (r-1) - (c-1) - (l-1) - 1 \\ &= r(c-1) - r - c + l + 2 \end{aligned}$$

Karşılıklı bağımsızlık hipotezi için tek değişkenli beklenen marjinal toplamaların tek değişkenli gözlenen marjinal toplamlara ($E_{i.}=n_{i.}$, $E_{.j}=n_{.j}$ ve $E_{.k}=n_{.k}$) eşitliliği gibi kısıtlamalar sözkonusudur.

3. SİMULASYON ÇALIŞMASI VE SONUÇLAR

İki ve üç yönlü tabloların gözlemlerinde 5'den az frekansların olması halinde küçük örneklerde gerçekleşen I'inci tip hatanın nasıl etkilendiğini irdelemek amacıyla marjinal toplamları birbirine eşit ve birbirinden çok farklı (extrem) olan üçer tablo ele alınmıştır (Akyol,1999).

İki ve üç yönlü tablolarda karşılaştırılan marjinal toplamlar dikkate alınarak, 100000'er defa üretilip analize tabi tutulmuş ve gerçekleşen I'inci tip hata olasılıkları belirlenmiştir.

Tesadüfi sayı üretimi ve hesaplamalar (Ek 1) için gerekli program FORTRAN programlama dilinde yazılmıştır (Akyol,1999). Bu işle ilgili program yapılırken, Copas and Fisk (1979), Patefield (1981) tarafından yapılan iki yönlü tablolar için simülasyon programlarından yararlanılmıştır.

İki yönlü tablo olarak 3x2 düşünülmesi durumunda toplam gözlem adedi 12, 24 ve 60 olarak karşılaştırılmıştır. Bu toplam gözlemler satır ve sütun marjinal toplamlarına ilk olarak eşit şekilde, daha sonra da ekstrem olarak bölünmüştür. İki yönlü tabloda karşılaştırılan marjinal toplamlar Tablo 3'deki gibidir.

Eşit marjinal toplamı durumunda, göz başına teorik olarak beklenen frekanslar 2, 4 ve 10 olmaktadır. Ekstrem marjinal toplamı durumunda ise gözlemlere düşen teorik olarak beklenen frekanslar ise haliyle farklı olmaktadır.

Karşılaştırılan toplam gözlem adedi ve marjinal toplamlar çerçevesinde her bir tablodaki gözlenen frekanslar tesadüf sayılar üretim tekniği ile 100000'er defa üretilmiş ve bağımsızlığa ilişkin H_0 hipotezi $\alpha=0.05$ düzeyinde kontrol edilmiştir. Bu simülasyon sonuçlarında, eşit marjinal toplamı durumunda I'inci tip hata olasılıkları sırasıyla 0.14334, 0.07046 ve 0.04814 olarak gerçekleşmişlerdir. Ekstrem marjinal toplamı durumunda ise I'inci tip hata olasılıkları sırasıyla 0.05995, 0.05161 ve 0.05538'dir. Dikkat edileceği üzere gözlemlere düşen teorik olarak beklenen frekanslar azalma yönünde 5'den çok uzaklaştığında gerçekleşen I'inci tip hata olasılıkları artmaktadır.

Tablo 3. 3x2 Yönlü Tablolar

a)Eşit marjinal toplamı 3x2 yönlü tablolar

1'inci 3x2 yönlü tablo		2'inci 3x2 yönlü tablo		3'üncü 3x2 yönlü tablo	
Satır marj. toplamları i=1,2,3	Sütun marj. toplamları j=1,2	Satır marj. toplamları i=1,2,3	Sütun marj. toplamları j=1,2	Satır marj. toplamları i=1,2,3	Sütun marj. toplamları j=1,2
4	6	8	12	20	30
4	6	8	12	20	30
4		8		20	
Genel toplam 12		Genel toplam 24		Genel toplam 60	

b)Extrem marjinal toplamı 3x2 yönlü tablolar

1'inci 3x2 yönlü tablo		2'inci 3x2 yönlü tablo		3'üncü 3x2 yönlü tablo	
Satır marj. toplamları i=1,2,3	Sütun marj. toplamları j=1,2	Satır marj. toplamları i=1,2,3	Sütun marj. toplamları j=1,2	Satır marj. toplamları i=1,2,3	Sütun marj. toplamları j=1,2
2	2	4	4	10	20
3	10	6	20	20	40
7		14		30	
Genel toplam 12		Genel toplam 24		Genel toplam 60	

Üç yönlü tabloların 3x2x3 olarak düşünülmesi durumundaki toplam gözlem adetleri 36, 72 ve 180 olarak karşılaştırılmıştır. Bu tablolarda satır, sütun ve tabaka marjinal toplamları Tablo 4'deki gibi karşılaştırılmıştır. Bu tablolarda eşit marjinal toplamlar durumunda teorik olarak beklenen frekanslar 2, 4 ve 10 olmaktadır. Extrem marjinal toplamlar durumunda ise gözlemlere düşen teorik olarak beklenen frekanslar farklı olmaktadır.

Kararlaştırılan toplam gözlem adedi ve marjinal toplamlar çerçevesinde her bir tablodaki gözlenen frekanslar tesadüf sayılar üretim tekniği ile 100000'er defa üretilmiş ve bağımsızlığa ilişkin H_0 hipotezi $\alpha=0.05$ düzeyinde kontrol edilmiştir. Bu simülasyon sonuçlarında eşit marjinal toplamlar durumunda I'inci tip hata olasılıkları sırasıyla 0.03600, 0.04143 ve 0.04688 olarak gerçekleşmişlerdir. Farklı marjinal toplamlar durumunda gerçekleşen I. tip hata olasılıkları 0.04713, 0.04141 ve 0.05719 'dır.

Dikkat edileceği üzere gözlemlere genellikle 5'den az gözlem düşecek şekilde verilen eşit marjinal toplamlar durumunda I'inci tip hata olasılıkları 0.05'in altına düşmektedir. Extrem marjinal toplamlar da benzer durum gözlenmektedir. Toplam gözlem sayısı 180 olarak karşılaştırıldığında eşit marjinal toplamlar çerçevesinde gerçekleşen I. tip hata $\alpha=0.05$ 'e oldukça yakın (0.04688) olarak gerçekleşmekte, extrem marjinal toplamlar durumunda ise gerçekleşen I. tip hata $\alpha=0.05$ 'in bir miktar üstünde (0.05719) bulunmaktadır. Bunun sebebi olarak da alınan üç yönlü tabloda extrem marjinal toplamlar durumunda gözlemlere teorik olarak beklenen frekansların oldukça geniş bir aralıkta (0.28-50.9) değiştiği ve hesaplanan χ^2 istatistiğinin uygun serbestlik dereceli teorik χ^2 dağılımına uyumunun bir miktar etkilendiği düşünülebilir. Nitekim

benzer durum toplam gözlem adedi 60 olan extrem marjinal toplamı 3x2 iki yönlü tabloda da gözlenmektedir.

Tablo 4. 3x2x3 Yönlü Tablolar

a) Eşit marjinal toplamı 3x2x3 yönlü tablolar

1'inci 3x2x3 yönlü tablo			2'inci 3x2x3 yönlü tablo		
Satır marj. toplamları i=1,2,3	Sütun marj. toplamları j=1,2	Tabaka marj. toplamları k=1,2,3	Satır marj. toplamları i=1,2,3	Sütun marj. toplamları j=1,2	Tabaka marj. toplamları k=1,2,3
12	18	12	24	36	24
12	18	12	24	36	24
12		12	24		24
Genel toplam			Genel toplam		
36			72		

3'üncü 3x2x3 yönlü tablo		
Satır marj. toplamları i=1,2,3	Sütun marj. toplamları j=1,2	Tabaka marj. toplamları K=1,2,3
60	90	60
60	90	60
60		60
Genel toplam		
180		

b) Extrem marjinal toplamı 3x2x3 yönlü tablolar

1'inci 3x2x3 yönlü tablo			2'inci 3x2x3 yönlü tablo		
Satır marj. toplamları i=1,2,3	Sütun marj. toplamları j=1,2	Tabaka marj. toplamları k=1,2,3	Satır marj. toplamları i=1,2,3	Sütun marj. toplamları j=1,2	Tabaka marj. toplamları k=1,2,3
6	10	8	12	20	15
10	26	10	20	52	20
20		18	40		37
Genel toplam			Genel toplam		
36			72		

3'üncü 3x2x3 yönlü tablo		
Satır marj. toplamları i=1,2,3	Sütun marj. toplamları j=1,2	Tabaka marj. toplamları k=1,2,3
20	30	15
60	150	55
100		110
Genel toplam		
180		

KAYNAKLAR

- AKYOL, M. (1999),. *Çok Yönlü Tablolarda İstatistiksel Analizler*, Ankara Üniversitesi, Fen Bilimleri Enstitüsü , Doktora Tezi. Ankara.
- COPAS, J.B. and FİSK, P.R. (1979), *Applied Statistics*, Journal of the Royal Statistical Societyl. Series C. Volume:28. No:3.
- EVERİTT, B.S. (1994), *The Analysis of Contingency Tables*, Monographs Applied Probability and Statistics 45, Chapman & Hall, New York.
- PATEFIELD, W. M. (1981), *Algorithm As 159, An Efficient Method of Generating Random RxC Tables with Given Row and Column Totals*, Applied Statistics. Volume: 30, Pages: 91-97.
- SERPER, Ö. (1986), *Uygulamalı İstatistik 2*, Filiz Kitabevi, Beyazıt İstanbul.

The Case of Cells of Two and Three Dimensional Tables Having Less Than 5 Frequencies, The Case of First Type Error (α) Probabilities

ABSTRACT

2 and 3 dimensional tables are set as the frequencies of the table cells will be less than 5 and 5 or more, and marginal totals will be constant (equal to each other) and different from each other by simulation study. In these conditions, uniform random numbers were produced and χ^2 analyze has been done 100000 times by using same marginal totals. At the end of the simulation study, α (first type error) probabilities have been found. For the condition of constant (equal to each other) marginal totals in the case of having theoretical frequency as 2 and 4, a big difference has occurred for α probability realized.

Key Words: Two-Dimensional Tables, Three-Dimensional Tables, Multidimensional Tables, The Analysis of χ^2 , Simulation.

Ek - 1 Marjinal toplamlardan hareketle iki ve üç yönlü tablolarda veri türetimi

```
*****UC.FOR*****
*****MARJİNAL TOPLAMLARDAN HAREKETLE*****
*****İKİ VE ÜÇ YÖNLÜ TABLOLARDAN VERİ TÜRETİR*****
*****100 000 DEFA TÜRETİLEN BU TABLOYU BAĞIMSIZLIK*****
*****TESTİ UYGULAR VE H0 RED VE KABUL SAYILARINI VERİR*
  DIMENSION NROWT(15),NCOLT(15),NTAT(15),BB(15,15)
  DIMENSION A(15,15,15),T(15),S(15),C(15),G(15,15,15),B(15,15)
  INTEGER NROW,NCOL,NTA,SD
  REAL KIKARE,TKIKAR
  OPEN(3,FILE='A:\UC17.DAT',STATUS='NEW')
  WRITE(*,*)'SATIR, SÜTUN VE TABAKA SAYISINI GİRİNİZ'
  WRITE(3,*)'SATIR, SÜTUN VE TABAKA SAYISINI GİRİNİZ'
  READ(*,*)NROW,NCOL,NTA
  WRITE(3,*)NROW,NCOL,NTA
  WRITE(*,*)'SATIR MARJİNAL TOPLAMLARINI GİRİNİZ'
  WRITE(3,*)'SATIR MARJİNAL TOPLAMLARINI GİRİNİZ'
  TT=0
  DO 101 I=1,NROW
    READ(*,*)NROWT(I)
101  WRITE(3,*)NROWT(I)

    WRITE(*,*)'SÜTUN MARJİNAL TOPLAMLARINI GİRİNİZ'
    WRITE(3,*)'SÜTUN MARJİNAL TOPLAMLARINI GİRİNİZ'
    DO 102 J=1,NCOL
      READ(*,*) NCOLT(J)
      WRITE(3,*)NCOLT(J)
102  TT=TT+NCOLT(J)

    IF (NTA.GT.1)THEN
      WRITE(*,*)'TABAKA MARJİNAL TOPLAMLARINI GİRİNİZ'
      WRITE(3,*)'TABAKA MARJİNAL TOPLAMLARINI GİRİNİZ'
      DO 103 K=1,NTA
        READ(*,*) NTAT(K)
103  WRITE(3,*)NTAT(K)
      ENDIF

    IF (NTA.LT.2)THEN
      SD=(NROW-1)*(NCOL-1)
      WRITE(*,*)SD
      WRITE(*,*)'SERBEST DERECELİ Kİ-KARE TABLO DEĞERİNİ GİRİNİZ'
      READ(*,*)TKIKAR
      WRITE(*,*)'LÜTFEN BEKLEYİNİZ'
    ELSE
      SD=(NROW*NCOL*NTA)-NROW-NCOL-NTA+2
      WRITE(*,*)SD
      WRITE(*,*)'SERBEST DERECELİ Kİ-KARE TABLO DEĞERİNİ GİRİNİZ'
      READ(*,*)TKIKAR
      WRITE(*,*)'LÜTFEN BEKLEYİNİZ'
    ENDIF
    MC=1
    NN=0
    MM=0
  *

  DO 998 L=1,100000
    CALL RCONT(A,BB,NROW,NCOL,NROWT,NCOLT,NTA,NTAT)
    IF (NTA.LT.2)THEN
```

Ek 1 (devam)

```
IF (MC.LT.6)THEN
WRITE(3,*)MC
WRITE(3,*)'. TABLO DEĞERİ'
WRITE(3,*)'      I,      J'
ENDIF
DO 980 I=1,NROW
DO 980 J=1,NCOL
B(I,J)=NROWT(I)*NCOLT(J)/TT

IF (MC.LT.6)THEN
WRITE(3,*)BB(I,J),I,J
ENDIF

980 CONTINUE
KIKARE=0
DO 970 I=1,NROW
DO 970 J=1,NCOL
970 KIKARE=KIKARE+(BB(I,J)-B(I,J))**2/B(I,J)

IF (MC.LT.6)THEN
WRITE(3,*)'HESAPLANAN KİKARE DEĞERİ'
WRITE(3,*)KIKARE
MC=MC+1
ENDIF

IF (KIKARE.LE.TKIKAR)THEN
NN=NN+1
ELSE
MM=MM+1
ENDIF

ELSE

IF(MC.LT.6)THEN
WRITE(3,*)MC
WRITE(3,*)'. ÜRETİLEN TABLO DEĞERİ'
WRITE(3,*)'A(I,J,K), I, J, K'
DO 601 I=1,NROW
DO 601 J=1,NCOL
DO 601 K=1,NTA
WRITE(3,*)A(I,J,K),I,J,K
601 CONTINUE
ENDIF
N=0
DO 201 I=1,NROW
S(I)=0
DO 202 J=1,NCOL
DO 202 K=1,NTA
S(I)=S(I)+A(I,J,K)
202 CONTINUE
N=N+S(I)
* WRITE(3,*)I,S(I)
201 CONTINUE
* WRITE(3,*)'SÜTUN TOPLAMLARI'
DO 203 J=1,NCOL
C(J)=0
```

Ek 1 (devam)

```
DO 204 I=1,NROW
DO 204 K=1,NTA
C(J)=C(J)+A(I,J,K)
204 CONTINUE
* WRITE(3,*)J,C(J)
203 CONTINUE

* WRITE(3,*)'TABAKA TOPLAMLARI'
DO 205 K=1,NTA
T(K)=0
DO 206 I=1,NROW
DO 206 J=1,NCOL
T(K)=T(K)+A(I,J,K)
206 CONTINUE
* WRITE(3,*)K,T(K)
205 CONTINUE

DO 207 I=1,NROW
DO 207 J=1,NCOL
DO 207 K=1,NTA
G(I,J,K)=(S(I)*C(J)*T(K))/(N*N)
* WRITE(3,*)I,J,K,G(I,J,K)
207 CONTINUE
* WRITE(3,*)'KI-KARE DEĞERİ VE SERBESTLİK DERESESİ'
KIKARE=0
DO 209 I=1,NROW
DO 209 J=1,NCOL
DO 209 K=1,NTA
KIKARE=KIKARE+(A(I,J,K)-G(I,J,K))**2/G(I,J,K)
209 CONTINUE

IF (MC.LT.6)THEN
WRITE(3,*)'HESAPLANAN KİKARE DEĞERİ'
WRITE(3,*)KIKARE
MC=MC+1
ENDIF

IF (KIKARE.LE.TKIKAR)THEN
NN=NN+1
ELSE
MM=MM+1
ENDIF
* WRITE(3,*)KIKARE,SD
ENDIF
*
998 CONTINUE
WRITE(3,*)'100 000 DEFA TEKRARLAMAMIN SONUCU'
WRITE(3,*)'Ho KABUL SAYISI, Ho RED SAYISI, TABLO DEĞERİ'
WRITE(3,*)NN,MM,TKIKAR

STOP
END
```

İki ve Üç Yönlü Tabloların Gözelerinde 5'den Küçük Beklenen Frekans Olması Durumunda, I. Tip Hata Olasılığının (α) Durumu

Ek 1 (devam)

```
SUBROUTINE RCONT(A,BB,NROW,NCOL,NROWT,NCOLT,NTA,NTAT)
DIMENSION MATRIX(15,15,15),NNTAT(15),A(15,15,15),NSUBT(15)
DIMENSION MAT(15,15),BB(15,15),NROWT(15),NCOLT(15),KVECT(1500)
DIMENSION NVECT(1500),NNVECT(1500),NTAT(15),MSUBT(15),KNVECT(1500)
INTEGER NROW,NCOL,NTA
REAL MATRIX
IFALT=0
*****HATALI GİRİŞLERİ KONTROL ETME
IF(NROW .LE. 0) GOTO 212
IF(NCOL .LE. 1) GOTO 213

IF (NROWT(1) .LE. 0) GOTO 214
MSUBT(1)=NROWT(1)
DO 200 I=2,NROW
IF (NROWT(I) .LE. 0) GOTO 214
MSUBT(I)=MSUBT(I-1)+NROWT(I)
200 CONTINUE

IF (NCOLT(1) .LE. 0) GOTO 215
NSUBT(1)=NCOLT(1)
DO 201 J=2,NCOL
IF (NCOLT(J) .LE. 0) GOTO 215
NSUBT(J)=NSUBT(J-1)+NCOLT(J)
201 CONTINUE
NTOTAL=NSUBT(NCOL)

IF(NTOTAL .GT. 1500) GOTO 216

*****DEĞİŞİM İÇİN BAŞLANGIÇ VEKTÖRÜ
DO 202 I=1,NTOTAL
NVECT(I)=I
202 CONTINUE
*****DEĞİŞİM İÇİN BAŞLANGIÇ VEKTÖRÜ
DO 204 I=1,NTOTAL
NNVECT(I)=NVECT(I)
204 CONTINUE
*****DEĞİŞİM VEKTÖRÜ
NTEMP=NTOTAL
5 CALL RANDOM_SEED()

DO 205 I=1,NTOTAL
CALL RANDOM_NUMBER(X1)
IF (I.EQ.1)THEN
IF (X2.EQ.X1)THEN
GOTO 5
ENDIF
X2=X1
ENDIF
NOCT=X1*FLOAT(NTEMP)+1.0
NVECT(I)=NNVECT(NOCT)
NNVECT(NOCT)=NNVECT(NTEMP)
NTEMP=NTEMP-1
205 CONTINUE

*****RASGELE MATRİSİN ELDE EDİLMESİ
IF (NTA.LT.2)THEN
```

Ek 1 (devam)

```
DO 226 I=1,NROW
DO 226 J=1,NCOL
MAT(I,J)=0
226 CONTINUE

II=1
DO 221 I=1,NROW
LIMIT=NROWT(I)
DO 222 K=1,LIMIT

DO 223 J=1,NCOL
IF (NVECT(II) .LE. NSUBT(J)) GOTO 208
223 CONTINUE

208 II=II+1
MAT(I,J)=MAT(I,J)+1
222 CONTINUE
221 CONTINUE

DO 960 I=1,NROW
DO 960 J=1,NCOL
960 BB(I,J)=MAT(I,J)

ELSE
*
NNTAT(1)=NTAT(1)
DO 229 K=2,NTA
229 NNTAT(K)=NNTAT(K-1)+NTAT(K)

*****DEĞİŞİM İÇİN BAŞLANGIÇ VEKTÖRÜ
DO 402 I=1,NTOTAL
KVECT(I)=I
402 CONTINUE
*****DEĞİŞİM İÇİN BAŞLANGIÇ VEKTÖRÜ
DO 404 I=1,NTOTAL
KNVECT(I)=KVECT(I)
404 CONTINUE
*****DEĞİŞİM VEKTÖRÜ
NTEMP=NTOTAL
15 CALL RANDOM_SEED()

DO 405 I=1,NTOTAL
CALL RANDOM_NUMBER(X1)
IF (I.EQ.1)THEN
IF (X2.EQ.X1)THEN
GOTO 15
ENDIF
X2=X1
ENDIF
NOCT=X1*FLOAT(NTEMP)+1.0
KVECT(I)=KNVECT(NOCT)
KNVECT(NOCT)=KNVECT(NTEMP)
NTEMP=NTEMP-1
405 CONTINUE
```

İki ve Üç Yönlü Tabloların Gözelerinde 5'den Küçük Beklenen Frekans Olması Durumunda, I. Tip Hata Olasılığının (α) Durumu

Ek 1 (devam)

```
DO 230 I=1,NROW
DO 230 J=1,NCOL
DO 230 K=1,NTA
MATRIX(I,J,K)=0
230 CONTINUE
```

```
II=1
DO 300 I=1,NROW
LIMIT=NROWT(I)
DO 301 K=1,LIMIT
DO 302 J=1,NCOL
IF (NVECT(II).LE.NSUBT(J))THEN
DO 303 L=1,NTA
IF (KVECT(II).LE.NNTAT(L))GOTO 281
303 CONTINUE
ELSE
ENDIF
302 CONTINUE
281 II=II+1
MATRIX(I,J,L)=MATRIX(I,J,L)+1
301 CONTINUE
300 CONTINUE
```

```
DO 291 I=1,NROW
DO 291 J=1,NCOL
DO 291 K=1,NTA
A(I,J,K)=MATRIX(I,J,K)
291 CONTINUE
```

ENDIF

RETURN

*****SATIR SAYISININ 0 VE DAHA AZ OLMASI DURUMU

```
212 IFAULT=1
WRITE(*,*)'SATIR SAYISININ 0 VE DAHA AZ OLMASI DURUMU'
RETURN
```

*****SÜTUN SAYISININ 1 VE DAHA AZ OLMASI DURUMU

```
213 IFAULT=2
WRITE(*,*)'SÜTUN SAYISININ 1 VE DAHA AZ OLMASI DURUMU'
RETURN
```

*****SATIR MARJİNAL TOPLAMININ 0 VE DAHA AZ OLMASI DURUMU

```
214 IFAULT=3
WRITE(*,*)'SATIR MARJİNAL TOPLAMININ 0 VE DAHA AZ OLMASI DURUMU'
RETURN
```

*****SÜTUN MARJİNAL TOPLAMININ 0 VE DAHA AZ OLMASI DURUMU

```
215 IFAULT=4
WRITE(*,*)'SÜTUN MARJİNAL TOPLAMININ 0 VE DAHA AZ OLMASI DURUMU'
RETURN
```

*****BÜYÜK ÖRNEKLEM DURUMU

```
216 IFAULT=5
WRITE(*,*)'BÜYÜK ÖRNEKLEM DURUMU'
RETURN
END
```

Yaşam Çözümlemesinde Yarışan Riskler ve Bir Uygulama

Durdu SERTKAYA*

M. Tekin SÖZER**

ÖZET

Bu çalışmada, yaşam çözümlemesinde birden fazla başarısızlık sebebinin olduğu durum ele alınmıştır. Yarışan riskler kavramı, sebebe-özel hazard fonksiyonu ve yaşam fonksiyonu gibi ilgili diğer fonksiyonlar incelenmiş ve yarışan riskler durumunda Cox regresyon modeline değinilmiştir. İş makinelerine ait lastik verileri ile bir uygulama yapılmıştır.

Anahtar kelimeler: Yarışan Riskler, Cox Regresyon, Yaşam Çözümlemesi

1. GİRİŞ

Canlı ya da cansız her birim birçok ölüm (başarısızlık) riskine maruz kalmaktadır. Ölüm, tekrar etmeyen bir olaydır ve bir tek sebebe bağlıdır. Yarışan riskler (competing risks), sebebe-özel ölümlülük (cause-specific mortality) çözümlemesinde ele alınmaktadır. Örneğin, ölüm riski olarak kanserin incelendiği bir çalışmada bazı kişiler, çalışma periyodu boyunca başka sebeplerden ölebilirler. Bu kişiler kanserden ölmemişlerdir ancak çalışma periyodunun sonunda da yaşamıyorlardır. Bu durumda yarışan risklerden (competing risks) bahsedilmektedir. Literatürde bu konuda pekçok çalışma yapılmıştır.

Kalbfleisch ve Prentice (1980), Kuk (1992), Marubini ve Volsecchi (1994), Lunn ve McNell (1995) ile Fine ve Gray (1999) yarışan riskler durumunda orantılı hazard modelini incelemişlerdir.

Green ve Byar (1980), Kay (1986) yarışan riskler çözümlemesini kullanarak prostat kanserli hastalar için tedavi etkilerini değerlendirmişlerdir.

Fusaro, Bacchetti ve Jewell (1996) AIDS, Krongrad, Lai ve Lai (1997) ile Cheng, Fine ve Wei (1998) prostat kanseri, Chapman, Fish ve Link (1999) ile Wohlfahrt, Andersen ve Melbye (1999) meme kanseri verilerini kullanarak yarışan riskler çözümlemesi yapmışlardır.

Goetghebeur ve Ryan (1995), Kundu ve Basu (2000) yaptıkları kuramsal çalışmada bazı bireyler için başarısızlık süreleri gözlemlendiği halde başarısızlık sebeplerinin gözlenemediği durumunu ele almışlardır.

* Dr., Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Beytepe-Ankara. durdu@hacettepe.edu.tr (Haberleşme adresi)

** Prof. Dr., Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Beytepe-Ankara.

2. YAŞAM ÇÖZÜMLEMESİNDE YARIŞAN RİSKLER

Yarışan riskler probleminde birden fazla başarısızlık türü ya da farklı başarısızlık sebebi vardır. Klinik çalışmalarda, epidemiyolojide, nüfusbilimde, temel bilimlerde ve endüstriyel alanlarda bu konuda çalışmalar yapılabilmektedir. Herbir çalışma alanındaki veriler, $T \geq 0$ başarısızlık süresini (başarısızlık süresi gözlenememiş ise durdurulmuş (censored) olarak ifade edilir) ve eğer T durdurulmuşsa bilinmeyen, $J \in \{1, 2, \dots, m\}$ başarısızlık türünü içermektedir. Ayrıca, başarısızlığı etkileyen nedenleri belirlemek için $z = (z_1, \dots, z_p)$ regresyon vektörü vardır (Prentice ve Kalbfleisch, 1978).

2.1. Sebebe-Özel Hazard Fonksiyonu ve İlgili Diğer Fonksiyonlar

Bir bireyin $J \in \{1, 2, \dots, m\}$ ile m tane başarısızlık sebebine maruz kaldığı varsayalım. Bir başarısızlık oluştuğunda T süresi ve J başarısızlık sebebi gözlenmektedir.

Homojen bir kitlede, sebebe-özel hazard fonksiyonu

$$h_j(t) = \lim_{\Delta t \rightarrow 0^+} P\{t \leq T < t + \Delta t, J = j | T \geq t\} / \Delta t, \quad j = 1, \dots, m$$

biçiminde tanımlanmaktadır. Bu hazard fonksiyonu, tüm sebeplerin varlığında, t süresinde j sebebi için anlık (instantaneous) başarısızlık hızı olarak tanımlanmaktadır. $(t, t+dt)$ aralığında j sebebi için koşullu başarısızlık olasılığı $h_j(t)dt$ 'dir. Sebebe-özel birikimli hazard (cause-specific cumulative hazard) fonksiyonu

$$H_j(t) = \int_0^t h_j(u) du$$

biçiminde ifade edilmektedir. Ayrıca,

$$S_{cs,j}(t) = \exp[-H_j(t)]$$

biçimindeki sebebe-özel yaşam fonksiyonları (cause-specific survival functions) da tanımlanabilmektedir.

Genel hazard (overall hazard) fonksiyonu

$$\lambda(t) = \sum_{j=1}^m h_j(t)$$

biçiminde olmaktadır. Böylece birikimli genel hazard (cumulative overall hazard) fonksiyonu

$$\Lambda(t) = \int_0^t \lambda(u)du = \sum_{j=1}^m H_j(t)$$

olarak yazılabilmektedir. Yaşam fonksiyonu

$$S(t) = \exp\left[-\sum_{j=1}^m H_j(t)\right] = \prod_{j=1}^m S_{cs,j}(t)$$

biçiminde ifade edilebilmektedir (Mann vd, 1974; Johnson ve Johnson, 1980; Marubini ve Volsecchi, 1994). $S(t) = S(t, t, \dots, t)$ $S(0) = 1$ ve $S(\infty) = 0$ ifadelerini sağlamakta ve monoton artmayan, sağdan sürekli bir fonksiyon olmaktadır.

Başarısızlık süresi T sürekli varsayıldığında, z regresyon vektörü ile bir bireyin genel başarısızlık hızı ya da hazard fonksiyonu ise,

$$\lambda(t; z) = \lim_{\Delta t \rightarrow 0^+} P\{t \leq T < t + \Delta t | T \geq t, z(t)\} / \Delta t$$

biçiminde verilmektedir. Bu durumda sebebe-özel hazard fonksiyonları,

$$\lambda_j(t; z) = \lim_{\Delta t \rightarrow 0^+} P\{t \leq T < t + \Delta t, J = j | T \geq t; z(t)\} / \Delta t, \quad j = 1, \dots, m$$

ile tanımlanmaktadır. $\lambda_j(t; z)$ fonksiyonu, diğer başarısızlık türlerinin de varlığında ve $z(t)$ regresyon vektörü verildiğinde, t süresinde j sebebinden anlık başarısızlık hızını vermektedir. Genel hazard fonksiyonu sebebe-özel hazard fonksiyonlarıyla,

$$\lambda(t; z) = \sum_{j=1}^m \lambda_j(t; z)$$

biçiminde ifade edilebilmektedir.

n bireyin verilerinin $(t_i, j_i, \delta_i, z_i^*)$ $i=1, \dots, n$ olduğunu varsayalım. Burada t_i başarısızlık süresini, j_i başarısızlık sebebini, δ_i durdurma göstergesini (censoring indicator) ve $z_i^* = z_i^*(t_i)$ i 'inci birey için regresyon vektörünü göstermektedir. Eğer başarısızlık oluşursa durdurma göstergesi 1 değerini, tersi durumda 0 değerini almaktadır. $\delta_i = 0$ ise, j_i başarısızlık sebebi bilinmiyor demektir. Bu durumda olabilirlik fonksiyonu

$$\prod_{i=1}^n \left\{ [\lambda_{j_i}(t_i; z_i)]^{\delta_i} S(t_i; z_i^*) \right\} = \left(\prod_{i=1}^n [\lambda_{j_i}(t_i; z_i)]^{\delta_i} \prod_{j=1}^m \exp\left\{-\int_0^{t_i} \lambda_j[u; z(u)] du\right\} \right)$$

biçiminde olmaktadır. Olabilirlik fonksiyonu tamamen $\lambda_j(t; z)$ $j=1, \dots, m$ sebebe-özel hazard fonksiyonlarıyla belirlenmektedir. Sebebe-özel hazard fonksiyonları (t, j, δ, z^*) biçimindeki verilerden doğrudan tahmin edilebilmektedir. $\lambda_j(t; z)$, j 'nin dışındaki sebeplere ait veriler durdurulmuş olarak alınarak elde edilmektedir (Prentice ve Kalbfleisch, 1978).

2.2. Yarışan Riskler Durumunda Cox Regresyon Modeli

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

Cox (1972)'un incelediği model

$$\lambda(t; \mathbf{z}) = \lambda_0(t) \exp(\beta' \mathbf{z}) \quad (1)$$

biçimindedir. Burada β , regresyon katsayıları vektörü, $\lambda_0(t)$ ise $\mathbf{z} = 0$ olan bir birimin temel hazard fonksiyonu olmaktadır. $\lambda_0(t)$ için özel bir biçim varsayılmamaktadır (Cox, 1972; Cox ve Oakes, 1984).

$\mathbf{z} = (x_1, \dots, x_p)$ açıklayıcı değişken vektörü ile bir bireyin hazard fonksiyonu $\lambda(t; \mathbf{z})$, (1) eşitliğinde verildiği biçimde olmaktadır. Sebebe-özel hazardın (1) modeli ile ifadesi

$$\lambda_j(t; \mathbf{z}) = \lambda_{0j}(t) \exp(\beta_j' \mathbf{z}) \quad j = 1, \dots, m \quad (2)$$

biçimindedir. Burada, temel hazard ve açıklayıcı değişkenler katsayılarının her ikisi de sebebe-özeldir (Kay, 1986). Cox'un orantılı hazard modeli, sebebe-özel hazard fonksiyonlarında regresyon değişkenlerinin etkilerini modellemede yardımcı olmaktadır. (2) eşitliğinde $\lambda_{0j}(\cdot) \geq 0$ dir ve keyfidir. β_j , $j = 1, \dots, m$, verilerden tahmin edilmekte ve sebebe-özel regresyon katsayılarının kolon vektörü olmaktadır (Prentice ve Kalbfleisch, 1978; Cheng vd, 1998).

$t_{j1} < \dots < t_{jk_j}$ $j = 1, \dots, m$ için j sebepli başarısızlıkların k_j sürelerini gösterebilir. z_{ji} , t_{ji} 'de başarısız olan birey için regresyon fonksiyonu olsun. Bu durumda kısmi olabilirlik,

$$L(\beta_1, \dots, \beta_m) = \prod_{j=1}^m \prod_{i=1}^{k_j} \left(\frac{\exp[z_{ji}(t_{ji})\beta_j]}{\sum_{l \in R(t_{ji})} \exp[z_{li}(t_{ji})\beta_j]} \right) \quad (3)$$

biçiminde verilmektedir. Burada $R(t_{ji})$, z_{ji} ve t_{ji} 'den önce riskte olan bireylerin kümesini göstermektedir. β_j 'lerin tahmini için standart asimtotik olabilirlik yöntemleri (3) ifadesine uygulanabilmektedir (Kalbfleisch ve Prentice, 1980).

Lunn ve McNell (1995) yaşam çözümlemesinde yarışan riskler modelindeki parametrelerin tahmini için iki yöntem vermişlerdir. Her iki durumda da veri tekraralama yöntemi kullanılarak Cox'un orantılı hazard regresyon modeli uygulanmıştır. İki yöntem de riskler bağımsız varsayılarak farklı başarısızlık türleri için kullanılmaktadır.

Goetghebeur ve Ryan (1995) bazı bireyler için başarısızlık türleri kayıp olduğunda yarışan riskli yaşam verilerini çözümlmek için bir yöntem önermişlerdir. Yaklaşımları, başarısızlık türlerinin herbiri için standart orantılı hazard yapısına dayanmaktadır (Goetghebeur ve Ryan, 1995).

Fine ve Gray (1999) açıklayıcı değişkenlerle yarışan riskler verileri için standart çözümlemenin, orantılı hazard varsayımı yoluyla sebebe-özel hazard fonksiyonlarını modellemeyi içerdiğini ifade etmişlerdir.

3. LASTİK VERİLERİ İLE İLGİLİ BİR UYGULAMA

Ülkemizde madencilik ve inşaat sektörlerinde istatistiksel uygulamalar pek yapılmamaktadır. Özellikle bu alanlarda yarışan riskler çözümlemesine ise raslanmamıştır. Her alanda, herhangi bir nedenle başarısızlık ve başarısızlık süresi olduğu durumda yaşam çözümlemesi yapılabilmektedir. Eğer başarısızlığı etkileyen nedenler birden fazla ise bu durumda yarışan riskler çözümlemesi uygulanabilmektedir.

Bu çalışmada, özel bir sektör tarafından kullanılan iş makinelerine ait lastik verileri ile uygulama yapılmıştır.

Uygulamada kullanılan 539 lastikten 264'ü çeşitli nedenlerle ömrünü tamamlamış ve hurdaya ayrılmış, 275 lastiğin ise kullanımı devam etmektedir. Lastikler üç değişik markadan (A, B, C) ve yedi değişik tipten (VMTS, VRLS, VRLS-LS, BIAS, RL 4J/4S, X KD1A, R 24 JE42S) oluşmaktadır. Lastiklerin çalışma ortamı dekapaj, kömür nakli ve dekapaj + kömür nakli olmak üzere üçe ayrılmıştır. Lastik iş makinelerine altı değişik pozisyonda (sol ön, sol arka iç, sol arka dış, sağ ön, sağ arka iç, sağ arka dış) takılabilmektedir. Lastiğin hurdaya ayrılma sebepleri; yanakta taş kesigi, sırtta taş kesigi, sırt atması ve diğer sebepler olmak üzere dört grupta toplanmıştır. Bir lastik bir pozisyondan bir başka pozisyona değiştirilebilmektedir. Bir lastiğin çalışma saati süresince kaç farklı pozisyonda kullanıldığı, bilgilerin işlendiği kartlardan elde edilmiş ve dört farklı kategoride toplanmıştır.

Çalışmada başarısızlık, lastiğin hurdaya ayrılması olarak tanımlanmıştır. Bir lastiğin kullanımına başlandığı andan hurdaya ayrılıncaya kadar geçen süre (saat olarak) lastiğin başarısızlık süresi olarak alınmıştır. Halen kullanımı devam eden lastikler ise durdurulmuş (censored) olarak tanımlanmıştır.

Uygulamada kullanılan değişkenler ve düzeyleri Tablo 1'de verilmiştir.

Çalışmada yarışan riskler durumunda herbir başarısızlık sebebi için lastiklerin başarısızlıklarını etkileyen faktörlerin belirlenmesi amaçlanmıştır. Yarışan riskler

durumunda herbir başarısızlık sebebinde başarısızlığı etkileyen risk faktörlerinin belirlenmesi için Cox regresyon çözümlemesi yapılmıştır. Cox regresyon çözümlemesinde, herbir değişken kodlarındaki ilk kod referans kategorisi olarak alınmıştır. Bu ise değişken düzeylerini yorumlamada önemli olmaktadır.

Herbir başarısızlık sebebinde başarısızlığı etkileyen risk faktörlerini belirlemek için öncelikle tek değişkenli Cox regresyon modelleri incelenmiştir. Modeldeki değişken için β parametresi, p-değeri, β parametresinin risk oranı ($\exp(\beta)$) ile önemli bulunan değişken düzeyleri için risk oranının alt ve üst sınırları verilmiştir. β parametresinin pozitif değer olması bu düzeyin referans kategorisine göre daha fazla riskli olduğunu, β parametresinin negatif değer olması ise bu düzeyin referans kategorisine göre daha az riskli olduğunu göstermektedir. Risk oranı olan $\exp(\beta)$ değeri ise önemli bulunan düzeyin, referans kategorisine göre kaç kat (ya da % ne kadar) daha riskli olduğu yorumunu getirmektedir.

Tablo 1. Kullanılan Değişkenler ve Düzeyleri

Değişken	Değişken Düzeyleri	n	%
Marka	1. A	391	72.5
	2. B	94	17.5
	3. C	54	10.0
Onarım	1. Onarım yok	505	93.7
	2. Onarım var	34	6.3
Ortam	1. Dekapaj	300	55.7
	2. Kömür	116	21.5
	3. Dekapaj + Kömür	80	14.8
Pozisyon sayısı	1. 1 pozisyon	323	59.9
	2. 2 pozisyon	127	23.6
	3. 3 pozisyon	57	10.6
	4. 4+ pozisyon	32	5.9
Pozisyon	1. Sol ön	90	16.7
	2. Sol arka iç	79	14.7
	3. Sol arka dış	81	15.0
	4. Sağ ön	97	18.0
	5. Sağ arka iç	96	17.8
	6. Sağ arka dış	88	16.3
Tip	1. VMTS	111	20.6
	2. VRLS	175	32.5
	3. VRLS-LS	73	13.5
	4. BIAS	27	5.0
	5. RL 4J/4S	43	8.0
	6. X KD1A	54	10.0
	7. R 24 JE42S	27	5.0
Sebep	1. Yanakta taş kesiği	172	65.2
	2. Sırtta taş kesiği	30	11.4
	3. Sırt atması	31	11.7
	4. Diğer	31	11.7
Durdurma	0. Hurda (başarısız)	264	49.0
	1. Kullanılıyor (yaşıyor)	275	51.0

Yanakta taş kesigi sebebi için elde edilen sonuçlar Tablo 2’de, sırtta taş kesigi sebebi için elde edilen sonuçlar Tablo 3’de, sırt atması sebebi için elde edilen sonuçlar Tablo 4’de ve diğer sebepler için elde edilen sonuçlar Tablo 5’de verilmiştir.

Tablo 2’deki p değerleri incelendiğinde MARKA, ONARIM, ORTAM ve TİP değişkenlerinin yanakta taş kesigi sebebi için önemli risk faktörleri olduğu söylenebilmektedir ($p < 0.05$). Önemli bulunan bu değişkenlerin herbir düzeyine karşılık gelen p değerlerine bakılarak önemli değişken düzeyleri belirlenebilmektedir. Herbir değişken için ilk düzeyler referans kategorisi olarak alındığından çizelgede yer almamaktadır. MARKA(2) (B) ve MARKA(3) (C) düzeyleri önemli olarak bulunmuştur. MARKA(2), MARKA(1)’e (A) göre 2 kat ($\exp(\beta) = 2,1639$) daha fazla riskli olmaktadır. MARKA(3) ise MARKA(1)’e göre %15 ($\exp(\beta) = ,1557$) daha az riskli olduğu görülmektedir. ONARIM değişkeni için ise, onarım yapılan lastiklerin başarısızlık riskinin yapılmayanlara göre yaklaşık 2 kat daha fazla olduğu söylenebilmektedir. ORTAM(1)’e (Dekapaj) göre ORTAM(2) (Kömür) 1,7 kat, ORTAM(3) (Dekapaj+Kömür) ise 2 kat daha fazla riskli olmaktadır. TİP değişkeni için önemli bulunan düzeylere bakıldığında TİP(2)’nin (VRLS) TİP(1)’e (VMTS) göre %54 daha az riskli olduğu, TİP(4)’ün (BIAS) TİP(1)’e göre 2.4 kat daha fazla riskli olduğu, TİP(6)’nın (X KD1A) TİP(1)’e göre %10 daha az riskli olduğu görülmektedir.

Tablo 2. Yanakta Taş Kesigi Sebebi İçin Tek Değişkenli Cox Regresyon Modeli

Değişken	β	p	Exp(β)	Altsınır-Üst sınır
MARKA		,0000		
MARKA (2)	,7719	,0033	2,163	1,2924 - 3,6229
MARKA (3)	-1,8597	,0002	,1557	,0576 - ,4208
ONARIM	,6282	,0044	1,8743	1,2160 - 2,8887
ORTAM		,0047		
ORTAM (2)	,5132	,0063	1,6706	1,1556 - 2,4150
ORTAM (3)	,7303	,0243	2,0756	1,0994 - 3,9186
POZİSYON SAYISI		,6677		
POZİSYON SAYISI (2)	,1914	,2823	1,2110	-
POZİSYON SAYISI (3)	-,0491	,8372	,9521	-
POZİSYON SAYISI (4)	,0520	,8420	1,0534	-
POZİSYON		,2097		
POZİSYON (2)	-,0331	,9386	,9675	-
POZİSYON (3)	,4711	,2332	1,6017	-
POZİSYON (4)	,3050	,4393	1,3566	-
POZİSYON (5)	,7553	,0516	2,1282	-
POZİSYON (6)	,7256	,0684	2,0660	-
TİP		,0000		
TİP (2)	-,6204	,0030	,5377	,3569 - ,8102
TİP (3)	-,4055	,0615	,6667	-
TİP (4)	,8789	,0111	2,4082	1,2225 - 4,7441
TİP (5)	,1790	,6465	1,1960	-
TİP (6)	-2,2775	,0000	,1025	,0365 - ,2880
TİP (7)	-10,8664	,9621	1,909E-05	-

Tablo 3 incelendiğinde sırtta taş kesigi sebebi için MARKA ve TİP değişkenleri önemli risk faktörleri olarak bulunmuştur ($p < 0.05$). Önemli bulunan MARKA(2) (B),

Tablo 3. Sırtta Taş Kesiği Sebebi İçin Tek Değişkenli Cox Regresyon Modeli

Değişken	β	p	Exp(β)	Altsınır-Üst sınır
MARKA		,0208		
MARKA (2)	1,2411	,0247	3,4593	1,1712 - 10,2172
MARKA (3)	-1,5370	,1323	,2150	-
ONARIM	-,9274	,3621	,3956	-
ORTAM		,2008		
ORTAM (2)	,6464	,1532	1,9086	-
ORTAM (3)	1,0370	,1784	2,8207	-
POZİSYON SAYISI		,8973		
POZİSYON SAYISI (2)	-,2220	,6221	,8009	-
POZİSYON SAYISI (3)	,1789	,7179	1,1959	-
POZİSYON SAYISI (4)	-,1298	,8397	,8783	-
POZİSYON		,5910		
POZİSYON (2)	,8215	,3436	2,2740	-
POZİSYON (3)	-12,7832	,9785	2,808E-06	-
POZİSYON (4)	,1264	,8995	1,1348	-
POZİSYON (5)	1,3573	,1081	3,8857	-
POZİSYON (6)	,9611	,2955	2,6145	-
TİP		,0005		
TİP (2)	-,2761	,5811	,7587	-
TİP (3)	-,8106	,1865	,4446	-
TİP (4)	1,9745	,0014	7,2027	2,1413 - 24,2279
TİP (5)	-12,5823	,9825	3,432E-06	-
TİP (6)	-1,9046	,0797	,1489	-
TİP (7)	-12,8329	,9951	2,671E-06	-

Tablo 4. Sırt Atması Sebebi İçin Tek Değişkenli Cox Regresyon Modeli

Değişken	β	p	Exp(β)	Altsınır-Üst sınır
MARKA		,0000		
MARKA (2)	1,8603	,0000	6,4259	2,7750 - 14,8801
MARKA (3)	-1,1748	,2521	,3089	-
ONARIM	-,9759	,3374	,3769	-
ORTAM		,0490		
ORTAM (2)	,2175	,6497	1,2430	-
ORTAM (3)	1,3834	,0141	3,9886	1,3219 - 12,0342
POZİSYON SAYISI		,1007		
POZİSYON SAYISI (2)	,8197	,0442	2,2698	1,0214 - 5,0442
POZİSYON SAYISI (3)	,1473	,8041	1,1587	-
POZİSYON SAYISI (4)	-,8870	,3990	,4119	-
POZİSYON		,3752		
POZİSYON (2)	1,9645	,0693	7,1311	-
POZİSYON (3)	,2394	,8657	1,2705	-
POZİSYON (4)	1,1883	,3035	3,2814	-
POZİSYON (5)	-11,4706	,9769	1,043E-05	-
POZİSYON (6)	1,3030	,2896	3,6803	-
TİP		,0000		
TİP (2)	-,2330	,6800	,7921	-
TİP (3)	-,0359	,9503	,9647	-
TİP (4)	2,6033	,0000	13,5078	4,2916 - 42,5158
TİP (5)	-12,4337	,9821	3,982E-06	-
TİP (6)	-1,3368	,2260	,2627	-
TİP (7)	-12,3478	,9930	4,339E-06	-

MARKA(1)'e (A) göre 3.5 kat daha fazla riskli olmaktadır. TİP değişkeninin önemli bulunan düzeyi TİP(4) (BIAS) ise TİP(1)'e (VMTS) göre 7.2 kat daha fazla risk taşımaktadır.

Tablo 4 incelendiğinde sırt atması sebebi için MARKA, ORTAM ve TİP değişkenleri başarısızlık riskinde önemli faktörler olarak bulunmuştur ($p < 0.05$). MARKA değişkeninin önemli bulunan MARKA(2) (B) düzeyinin, MARKA(1)'e (A) göre 6.4 kat daha fazla riskli olduğu görülmektedir. ORTAM değişkeninin önemli bulunan ORTAM(3) (Dekapaj + Kömür) düzeyi, ORTAM(1)'e (Dekapaj) göre 4 kat, TİP değişkeninin önemli olan TİP(4) (BIAS) düzeyi ise TİP(1)'e (VMTS) göre 13,5 kat daha fazla riskli olmaktadır. Burada POZİSYON SAYISI(2) düzeyinin de önemli olduğu görülmektedir. Bu durumda POZİSYON SAYISI(2)'nin POZİSYON SAYISI(1)'e göre 2 kat daha fazla riskli olduğu söylenebilmektedir.

Tablo 5. Diğer Sebepler İçin Tek Değişkenli Cox Regresyon Modeli

Değişken	β	p	Exp(β)	Altsınır-Üst sınır
MARKA		,0592		
MARKA (2)	1,3168	,0182	3,7316	1,2509 - 11,1319
MARKA (3)	,0217	,9683	1,0220	-
ONARIM	,5292	,3252	1,6976	-
ORTAM		,0111		
ORTAM (2)	1,1781	,0039	3,2481	1,4591 - 7,2306
ORTAM (3)	1,1181	,1444	3,0589	-
POZİSYON SAYISI		,6581		
POZİSYON SAYISI (2)	-,5633	,2153	,5693	-
POZİSYON SAYISI (3)	-,1525	,7677	,8586	-
POZİSYON SAYISI (4)	-,3300	,6019	,7189	-
POZİSYON		,8886		
POZİSYON (2)	,0406	,9526	1,0415	-
POZİSYON (3)	-1,2347	,2631	,2909	-
POZİSYON (4)	-,4065	,5779	,6660	-
POZİSYON (5)	-,0362	,9612	,9644	-
POZİSYON (6)	-,2865	,7358	,7509	-
TİP		,0114		
TİP (2)	-1,1075	,0232	,3304	,1270 - ,8593
TİP (3)	-1,4176	,0183	,2423	,0746 - ,7869
TİP (4)	1,1094	,1010	3,0325	-
TİP (5)	-,4175	,6933	,6587	-
TİP (6)	-1,1124	,0983	,3288	-
TİP (7)	-10,4464	,9832	2,905E-05	-

Tablo 5 incelendiğinde diğer sebepler için ORTAM ve TİP değişkenlerinin önemli risk faktörleri olduğu ifade edilebilmektedir. Önemli bulunan ORTAM(2) (Kömür) düzeyi, ORTAM(1)'e (Dekapaj) göre 3 kat daha fazla riskli olmaktadır. TİP değişkeninin önemli bulunan TİP(2) (VRLS) düzeyinin TİP(1)'e (VMTS) göre %33, TİP(3) (VRLS-LS) düzeyinin ise TİP(1)'e (VMTS) göre %24 daha az riskli olduğu söylenebilmektedir. Burada MARKA(2) (B) düzeyinin de önemli olduğu ve MARKA(2)'nin (B), MARKA(1)'e (A) göre 3.7 kat daha riskli olduğu görülmektedir.

Bundan sonra ise tüm değişkenler ve etkileşimleri birlikte çözümlenmeye alınarak her bir başarısızlık sebebinde başarısızlığı etkileyen risk faktörlerini belirlemek için adimsal Cox regresyon çözümlenmesi yapılmıştır.

Yanakta taş kesici sebebi için adimsal Cox regresyon modeli sonucu Tablo 6’da verilmiştir.

Tablo 6. Yanakta Taş Kesici Sebebi İçin Adimsal Cox Regresyon Modeli

Değişken	β	p	Exp(β)	Altsınır-Üst sınır
ONARIM*TİP				
ONARIM*TİP (2)	2,1300	,0019	8,4146	2,2004 - 32,1781
ONARIM*TİP (3)	1,1529	,0398	3,1674	1,0554 - 9,5059
ONARIM*TİP (4)	3,0350	,9562	20,8007	-
ONARIM*TİP (5)	-8,6314	,9593	1,784E-04	-
ONARIM*TİP (6)	9,9870	,6489	21741,418	-
ONARIM*TİP (7)	-2,9287	,9870	,0535	-
POZİSYON SAYISI*TİP				
POZİSYON SAYISI (2)*TİP (2)	-1,9279	,0001	,1454	,0541 - ,3913
POZİSYON SAYISI (3)*TİP (2)	-2,6160	,0010	,0731	,0154 - ,3474
POZİSYON SAYISI (4)*TİP (2)	-1,7956	,1608	,1660	-
POZİSYON SAYISI (2)*TİP (3)	-,9190	,1090	,3989	-
POZİSYON SAYISI (3)*TİP (3)	-,8190	,2595	,4409	-
POZİSYON SAYISI (4)*TİP (3)	-1,2622	,1498	,2830	-
POZİSYON SAYISI (2)*TİP (4)	-1,3271	,2203	,2652	-
POZİSYON SAYISI (3)*TİP (4)	-,4504	,6967	,6373	-
POZİSYON SAYISI (4)*TİP (4)	-10,2892	,9258	3,400E-05	-
POZİSYON SAYISI (2)*TİP (5)	-2,2733	,0551	,1030	-
POZİSYON SAYISI (3)*TİP (5)	13,5578	,9680	772853,27	-
POZİSYON SAYISI (4)*TİP (5)	-1,5195	,2070	,2188	-
POZİSYON SAYISI (2)*TİP (6)	-8,1849	,7660	2,788E-04	-
POZİSYON SAYISI (3)*TİP (6)	-8,1096	,8124	3,006E-04	-
POZİSYON SAYISI (4)*TİP (6)	,9233	,4912	2,5175	-
POZİSYON SAYISI (2)*TİP (7)	6,3453	,8178	569,8062	-
POZİSYON SAYISI (3)*TİP (7)	-9,3602	,9780	8,608E-05	-
POZİSYON SAYISI (4)*TİP (7)	9,2871	,9331	10797,529	-

Üsteki model incelendiğinde ONARIM*TİP(2), ONARIM*TİP(3), POZİSYON SAYISI(2)*TİP(2) ve POZİSYON SAYISI(3)*TİP(2) etkileşim terimleri anlamlı bulunmuştur. ONARIM*TİP(2) için, onarım yapılan VRLS tipi lastiklerin başarısızlık riskinin, referans kategorisi olan onarım yapılmayan VMST tipi lastiklerden 8,4 kat daha fazla olduğu yorumu yapılabilmektedir. ONARIM*TİP(3) için, onarım yapılan VRLS-LS tipi lastiklerin başarısızlık riskinin, referans kategorisine göre 3 kat daha fazla olduğu söylenebilmektedir. POZİSYON SAYISI(2)*TİP(2) için, iki kez pozisyon değiştiren VRLS tipi lastiklerin başarısızlık riskinin, referans kategorisi olan sadece bir pozisyonda kullanılan VMST tipi lastiklere göre %15 daha az olduğu yorumlanabilmektedir. Son olarak, POZİSYON SAYISI(3)*TİP(2) için, üç kez pozisyon değiştiren VRLS tipi lastiklerin başarısızlık riskinin, referans kategorisine göre %7 daha az olduğu söylenebilmektedir.

Sırtta taş kesliği sebebi için Cox regresyon modeli sonuçları Tablo 7’de verilmiştir.

Tablo 7. Sırtta Taş Kesliği Sebebi İçin Adımsal Cox Regresyon Modeli

Değişken	B	p	Exp(β)	Altsınır-Üst sınır
TİP				
TİP (2)	-,6823	,6576	,5055	-
TİP (3)	-1,5659	,3411	,2089	-
TİP (4)	,6764	,5822	1,9667	-
TİP (5)	-13,2979	,9810	1,678E-06	-
TİP (6)	-3,2031	,0368	,0406	,0020 - ,8214
TİP (7)	-15,4621	,9978	1,927E-07	-
MARKA*ORTAM		,0107		
MARKA (2)*ORTAM (2)	-1,5070	,4523	,2216	-
MARKA (3)*ORTAM (2)	2,3492	,4805	10,4776	-
MARKA (2)*ORTAM (3)	-5,3393	,0013	,0048	1,837E-04 - ,1254
MARKA (3)*ORTAM (3)	-5,2655	,0335	,0052	4,030E-05 - ,6624

Sırtta taş kesliği sebebi için TİP değişkeni ve MARKA*ORTAM etkileşim terimi önemli risk faktörleri olarak modelde yer almaktadır. Önemli bulunan TİP(6)’nın (X KD1A), TİP(1)’e (VMTS) göre %4 daha az riskli olduğu ifade edilebilmektedir. MARKA*ORTAM etkileşim teriminin önemli bulunan MARKA(2)*ORTAM(3) düzeyi için Dekapaj + Kömür ortamında kullanılan B marka lastiklerin, Dekapaj ortamında kullanılan A marka lastiklere göre daha az riskli olduğu, önemli bulunan MARKA(3)*ORTAM(3) için ise Dekapaj + Kömür ortamında kullanılan C marka lastiklerin Dekapaj ortamında kullanılan A marka lastiklere göre daha az riskli olduğu yorumu yapılabilmektedir.

Sırt atması sebebi için Cox regresyon modeli sonuçları Tablo 8’de verilmiştir.

Tablo 8. Sırt Atması Sebebi İçin Adımsal Cox Regresyon Modeli

Değişken	β	p	Exp(β)	Altsınır-Üst sınır
POZİSYON SAYISI*TİP				
POZİSYON SAYISI (2)*TİP (2)	-,8730	,6607	,4177	-
POZİSYON SAYISI (3)*TİP (2)	3,8128	,7576	45,2782	-
POZİSYON SAYISI (4)*TİP (2)	,9596	,9428	2,6107	-
POZİSYON SAYISI (2)*TİP (3)	-2,9773	,0679	,0509	-
POZİSYON SAYISI (3)*TİP (3)	-1,0757	,9334	,3411	-
POZİSYON SAYISI (4)*TİP (3)	-1,3103	,8867	,2698	-
POZİSYON SAYISI (2)*TİP (4)	-6,6243	,0014	,0013	2,258E-05 - ,0781
POZİSYON SAYISI (3)*TİP (4)	3,9286	,7490	50,8365	-
POZİSYON SAYISI (4)*TİP (4)	-6,8218	,7324	,0011	-
POZİSYON SAYISI (2)*TİP (5)	9,1264	,7610	9194,8011	-
POZİSYON SAYISI (3)*TİP (5)	6,6856	,8586	800,8001	-
POZİSYON SAYISI (4)*TİP (5)	13,6061	,6853	811043,47	-
POZİSYON SAYISI (2)*TİP (6)	-2,8624	,6187	,0571	-
POZİSYON SAYISI (3)*TİP (6)	2,3239	,8654	10,2151	-
POZİSYON SAYISI (4)*TİP (6)	6,4652	,4871	642,3761	-
POZİSYON SAYISI (2)*TİP (7)	-16,0628	,5914	1,057E-07	-

Tablo 8 incelendiğinde, POZİSYON SAYISI(2)*TİP(4) düzeyleri sırt atması sebebi için önemli risk faktörü olmaktadır. Buradan, iki kez pozisyon değiştiren BIAS tipi lastiklerin bir tek pozisyonda kullanılan VMTS tipi lastiklere göre daha az riskli olduğu söylenebilmektedir.

Diğer sebepler için Cox regresyon modeli sonuçları Tablo 9’da verilmiştir.

Tablo 9. Diğer Sebepler İçin Adımsal Cox Regresyon Modeli

Değişken	β	p	Exp(β)	Altsınır-Üst sınır
ORTAM				
ORTAM (2)	1,1781	,0039	3,2481	1,4591 - 7,2306
ORTAM (3)	1,1181	,1444	3,0589	-

Tüm değişkenler ve etkileşimler çözümlenmeye alındığında, en küçük p-değerine sahip olan ORTAM değişkeni model için anlamlı bulunan değişken olmuştur. Bu değişkenin anlamlı bulunan düzeyi olan ORTAM(2) (Kömür), ORTAM(1)’e (Dekapaj) göre 3.2 kat daha fazla risk taşımaktadır.

4. SONUÇ

Bu çalışmada, yarışan riskler çözümlenmesi tanıtıldıktan sonra, birden fazla başarısızlık sebebi olduğunda, lastik başarısızlığını etkileyen faktörleri belirlemek için Cox regresyon çözümlenmesi yapılmıştır. Bu çözümlenme ile değişik sebeplerde başarısızlığı etkileyen önemli açıklayıcı değişkenler belirlenebilmekte, açıklayıcı değişkenlerin düzeyleri karşılaştırılabilmekte ve başarısızlık riskinin hangi düzeyde daha fazla olduğu yorumu yapılabilmektedir.

Lastik başarısızlığı üzerinde etkili olan değişkenleri belirlemek için yapılan tek değişkenli Cox regresyon çözümlenmesi sonucunda, yanakta taş kesigi sebebi için marka, onarım, ortam ve tip, sırtta taş kesigi sebebi için marka ve tip, sırt atması sebebi için marka, ortam ve tip, diğer sebepler için ise ortam ve tip değişkenleri önemli bulunmuştur. Bu sonuçlar Tablo 10’da özet biçimde verilmiştir.

Tablo 10. Tek Değişkenli Cox Regresyon Modeli Sonuçları

Sebepler	Değişken	p
Yanakta Taş Kesigi	MARKA	,0000
	ONARIM	,0044
	ORTAM	,0047
	TİP	,0000
Sırtta Taş Kesigi	MARKA	,0208
	TİP	,0005
Sırt Atması	MARKA	,0000
	ORTAM	,0490
	TİP	,0000
Diğer	ORTAM	,0111
	TİP	,0114

Başarısızlık sebepleri üzerinde etkili olan değişkenleri belirlerken marka, onarım, ortam, pozisyon sayısı, tip ve tüm ikili etkileşim terimlerinin olduğu model düşünülerek adimsal Cox regresyon çözümlemesi yapıldığında, yanakta taş kesigi sebebi için onarım ile tip etkileşimi ve pozisyon sayısı ile tip etkileşimi, sırtta taş kesigi sebebi için tip değişkeni ve marka ile ortam etkileşimi, sırt atması sebebi için pozisyon sayısı ile tip etkileşimi, diğer sebepler için ise sadece ortam değişkeni başarısızlığı etkileyen faktörler olarak bulunmuştur. Bu sonuçlar Tablo 11'de özet olarak verilmiştir.

Tablo 11. Adimsal Cox Regresyon Modeli Sonuçları

Sebeup	Değişken	p
Yanakta Taş Kesigi	ONARIM*TİP (2)	,0019
	ONARIM*TİP (3)	,0398
	POZİSYON SAYISI (2)*TİP (2)	,0001
	POZİSYON SAYISI (3)*TİP (2)	,0010
Sırtta Taş Kesigi	TİP (6)	,0368
	MARKA*ORTAM	,0107
Sırt Atması	POZİSYON SAYISI (2)*TİP (4)	,0014
Diğer	ORTAM (2)	,0039

KAYNAKLAR

- CHAPMAN, J.W., FISH, E.B. and LINK, M.A. (1999), *Competing Risks Analyses for Recurrence from Primary Breast Cancer*, British Journal of Cancer, 79, (9/10), 1508-1513.
- CHENG, S. C., FINE, J.P. and WEI L.J. (1998), *Prediction of Cumulative Incidence Function under the Proportional Hazards Model*, Biometrics, 54, 219-228.
- COX, D.R. (1972), *Regression Models and Life Tables*, Journal of Royal Statistical Society, Series B, 34, 187-202.
- COX, D.R. and OAKES, D. (1984), *Analysis of Survival Data*, Chapman and Hall.
- FINE, J.P. and GRAY, R.J. (1999), *A Proportional Hazards Model for the Subdistribution of a Competing Risk*, JASA, 94, 446, 496-509.
- FUSARO, R.E., BACCHETTI, P. and JEWELL, N.P. (1996), *A Competing Risks Analysis of Presenting AIDS Diagnoses Trends*, Biometrics, 52, 211-225.
- GOETGHEBEUR, E. and RYAN L. (1995), *Analysis of Competing Risks Survival Data when some Failure Types are Missing*, Biometrika, 82, 4, 821-833.
- GREEN, S.B. and BYAR, D.P. (1980), *The Choice of Treatment for Cancer Patients Based on Covariate Information: Application to Prostate Cancer*, Bulletin Cancer, 67, 477-488.

- JOHNSON, R.E. and JOHNSON, N.L. (1980), *Survival Models and Data Analysis*, John Wiley and Sons.
- KALBFLEISCH, J. D. and PRENTICE, R.L. (1980), *The Statistical Analysis of Failure Time Data*, John Wiley and Sons.
- KAY, R. (1986), *Treatment Effects in Competing-Risks Analysis of Prostate Cancer Data*, *Biometrics*, 42, 203-211.
- KRONGRAD, A., LAI, H. and LAI, S. (1997), *Competing Risks of Mortality in Prostate Cancer*, *Journal of Urology*, 158, 865-868.
- KUK, A.Y.C. (1992), *A Semiparametric Mixture Model for the Analysis of Competing Risks Data*, *Austral. J. Statist.*, 34, 2, 169-180.
- KUNDU, D. and BASU, S. (2000), *Analysis of Incomplete Data in Presence of Competing Risks*, *Journal of Statistical Planning and Inference*, 87, 221-239.
- LAWLESS, J.F. (1982), *Statistical Models and Methods for Lifetime Data*, John Wiley and Sons.
- LUNN, M. and MCNEIL, D. (1995), *Applying Cox Regression to Competing Risks*, *Biometrics*, 51, 524-532.
- MANN, N.R., SCHAFER, R.E. and SINGPURWALLA, N.D. (1974), *Methods for Statistical Analysis of Reliability and Life Data*, John Wiley and Sons.
- MARUBINI, E. and VOLSECCHI, M.G. (1994), *Analysing Survival Data from Clinical Trials and Observational Studies*, John Wiley and Sons.
- PRENTICE, R.L. and KALBFLEISCH, J. D. (1978), *The Analysis of Failure Times in the Presence of Competing Risks*, *Biometrics*, 34, 541-554.
- WOHLFAHRT, J., ANDERSEN, P.K. and MELBYE, M. (1999), *Multivariate Competing Risks*, *Statistics in Medicine*, 18, 1023-1030.

Competing Risks in the Survival Analysis and an Application

ABSTRACT

In this study, survival analysis is investigated when there are more than one distinct cause of death. The concept of competing risks, cause-specific hazard function and other related functions such as survival function is explained. In the presence of competing risks, the Cox regression model is examined. An application is carried out by using the tire data of big trucks.

Key words: Competing Risks, Cox Regression, Survival Analysis

Sürekli Değişken İçeren Grafikselle Modeller

Hülya BAYRAK* Fikri GÖKPINAR* Berrin ÖZKAYA**

ÖZET

Çok değişkenli normal dağılıma sahip değişkenler için elde edilen grafikselle modeller kovaryans seçimli modeller olarak adlandırılır. Kovaryans seçimli modeller değişken çiftlerinin koşullu bağımsızlığına dayanarak belirlenir. Bu çalışmada bir grafikselle modeldeki test işleminde kullanılan sapma istatistiğinin kovaryans seçimli modellerde irdelenmesi verildi. Uygulama olarak 30 farklı un örneği kullanıldı. Un örneklerinin protein miktarı, gluten miktarı ve sedimantasyon değerleri ile bu unlardan yapılan ekmeklerin, hacim verimi ve Dallman değerleri tesbit edildi. Elde edilen sonuçlar, kovaryans seçimli model ile irdelendi.

Anahtar kelimeler: Kovaryans Seçimli Modeller, Sapma, Markov Özellikleri, Koşullu Bağımsızlık

1. GİRİŞ

Çok değişkenli normal dağılıma sahip değişkenler için elde edilen grafikselle modeller kovaryans seçimli modeller olarak adlandırılır. Kovaryans seçimli modellerde, kovaryans yapısı, kovaryans matrisinin tersinin elemanlarını sıfıra götürerek basitleştirilebilir. Kovaryans matrisinin tersinde sıfırların nasıl oluşturulacağını ortaya koyan bir kural Dempster(1972) tarafından verilmiştir.

Bir grafikte, V köşe kümesini, E köşeler arası kenar kümesini temsil eder. Eğer bir köşe çifti arası birden fazla kenar yoksa grafik basittir.

Bir kenar hem $(\alpha, \beta) \in E$ hem de $(\beta, \alpha) \in E$ şeklinde ise E yön verilmemiş kenar, $(\alpha, \beta) \in E$ şeklinde olup $(\beta, \alpha) \notin E$ ise yön verilmiş bir kenardır.

Bir grafiğin tüm köşeleri arasında çizgi ya da ok varsa bu grafiğe tam grafik denir. V'nin bir alt kümesine, eğer buna ilişkin grafik tam ise, tamdır denir. V'nin tam alt kümelerine klik denir.

* Gazi üniversitesi Fen Edebiyat Fakültesi İstatistik bölümü, Ankara, Türkiye, hbayrak@gazi.edu.tr-fikri@gazi.edu.tr

**Ankara üniversitesi Ziraat Fakültesi, Gıda Mühendisliği bölümü, Ankara, Türkiye. bozkaya@agri.ankara.edu.tr

α 'dan β 'ya bir ok varsa α 'ya β 'nin ailesi, β 'ya da α 'nın çocuğu denir. β 'nin aile kümesini $pa(\beta)$, α 'nın çocuklar kümesini $ch(\alpha)$ şeklinde gösterilebilir.

α ile β arasında bir çizgi varsa α ve β 'ya bitişik ya da komşu denir. α köşesinin komşu kümesi $ne(\alpha)$ şeklinde gösterilebilir.

Grafiksel modellerin temeli rasgele değişkenlerin koşullu bağımsızlığına dayanır. Bu modellerin grafikleri koşullu bağımsızlık ilişkilerini gösterir.

X, Y ve Z kesikli rasgele değişkenler olsun. $X \perp Y/Z$ koşullu bağımsızlık ifadesi aşağıdaki gibi yazılır.

$$P(X=x, Y=y/Z=z) = P(X=x/Z=z) P(Y=y/Z=z)$$

Burada tüm z 'ler için $P(Z=z) > 0$ 'dır. X, Y, Z sürekli rasgele değişkenleri için $X \perp Y/Z$ koşullu bağımsızlığı

$$X \perp Y/Z \Leftrightarrow f_{XY/Z}(x, y/z) = f_{X/Z}(x/z) f_{Y/Z}(y/z) \quad (1)$$

biçiminde ifade edilir.

$X \perp Y/Z$ ilişkisi aşağıdaki özelliklere sahiptir.

$$\left. \begin{array}{l} X \perp Y/Z \Rightarrow Y \perp X/Z \text{ dir.} \\ X \perp Y/Z \text{ ve } U=h(x) \Rightarrow U \perp Y/Z \text{ dir.} \\ X \perp Y/Z \text{ ve } U=h(x) \Rightarrow X \perp Y/(Z,U) \text{ dir.} \\ X \perp Y/Z \text{ ve } X \perp W/(Y,Z) \Rightarrow X \perp (W,Y)/Z \text{ dir.} \end{array} \right\} \quad (2)$$

Rasgele değişkenler $(\chi_\alpha)_{\alpha \in V}$ koleksiyonu ve yön verilmemiş grafik için değişik markov özellikleri vardır. Bu özellikler aşağıdaki gibidir;

χ üzerindeki olasılık ölçümü P olmak üzere; (P) bitişik olmayan herhangi bir köşe çifti (α, β) için aşağıdaki koşul sağlanıyorsa ikili markov özelliğine sahiptir.

$$\alpha \perp \beta / \forall \{ \alpha, \beta \}$$

(L) herhangi bir köşe $\alpha \in V$ için aşağıdaki koşul sağlanıyorsa yerel markov özelliğine sahiptir.

$$\alpha \perp V \setminus cl(\alpha) / bd(\alpha)$$

(G) V 'nin bağlantısız altkümelerin herhangi bir üçlüsü (A, B, S) aşağıdaki koşulu sağlıyorsa global markov özelliğine sahiptir.

$$A \perp B/S$$

$X=(X_1, X_2, \dots, X_p)$ vektörünün yoğunluk fonksiyonu $f_v(X_v)=f(X_1, \dots, X_p)$ mevcut ise, $f_v(x_v)$ yoğunluğu, $G=(V,E)$ grafiğine göre aşağıdaki gibi faktörize edilebilir.

$$f_v(x_v) = \prod_{c \in C} \Psi_c(x_c) \quad (3)$$

Burada C , G 'deki kliklerin kümesi ve $\Psi_c(x_c)$ x üzerinden x_c 'ye bağlı negatif olmayan fonksiyonlardır. Bileşik yoğunluk $f_v(x_v)$ kesin pozitif ise faktörizasyon özelliği ve global markov özelliği denktir (Lauritzen, 1996).

2 SÜREKLİ MODELLER

Grafiksel modeller sürekli değişkenler için kovaryans seçimli modelleri temel alır. Kovaryans seçimli modeller ilk önce Dempster(1972) tarafında kullanılmıştır. Bu modeller Whittaker(1990) tarafından geliştirilmiştir.

$Y=(Y_1, \dots, Y_q)$ q boyutlu rassal değişkenler vektörü olsun.

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \mu_q \end{pmatrix} \quad (4)$$

ortalama vektörü ve

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdot & \cdot & \sigma_{1q} \\ \sigma_{21} & \sigma_{22} & \cdot & \cdot & \sigma_{2q} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{q1} & \sigma_{q2} & \cdot & \cdot & \sigma_{qq} \end{pmatrix} \quad (5)$$

kovaryans matrisine sahip çok değişkenli normal dağılıma sahiptir. Burada özellikle ilgilenilecek olan kovaryans matrisinin tersidir.

$$K = \Sigma^{-1} = \begin{pmatrix} w^{11} & w^{12} & \cdot & \cdot & w^{1q} \\ w^{21} & w^{22} & \cdot & \cdot & w^{2q} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ w^{q1} & w^{q2} & \cdot & \cdot & w^{qq} \end{pmatrix} \quad (6)$$

Bu matris kesinlik matrisi ya da konsantrasyon matrisi olarak adlandırılır(Lauritzen and Wermuth, 1989).

(Y_3, \dots, Y_q) verilmişken (Y_1, Y_2) 'in koşullu dağılımı 'Eş.7'deki konsantrasyon matrisine sahip iki değişkenli normal dağılımdır.

$$\begin{pmatrix} w^{11} & w^{12} \\ w^{21} & w^{22} \end{pmatrix}^{-1} = \frac{1}{w^{11}w^{22} - (w^{12})^2} \begin{pmatrix} w^{22} & -w^{21} \\ -w^{12} & w^{11} \end{pmatrix} \quad (7)$$

bu iki değişkenli dağılıma ilişkin korelasyon katsayısı 'Eş 8'deki gibidir.

$$\rho^{12,34\dots q} = \frac{-w^{12}}{(w^{11}w^{22})^{\frac{1}{2}}} \quad (8)$$

Ayrıca

$$\rho^{12,34\dots q} = 0 \Leftrightarrow w^{12} = 0 \quad (9)$$

denkliği 'Eş 8''den açıkça görülür.

Bir başka deyişle, geri kalan değişkenler verilmişken iki değişkenin bağımsız olması için gerek ve yeter koşul konsantrasyon matrisinde karşılık gelen elemanların sıfır olmasıdır(Cox and Wermuth 1993). Bu şekilde konsantrasyon matrisinin elemanları, log lineer modeldeki iki faktörlü etkileşimlerle aynı rolü üstlenir.

Bu sonucu bir başka yönden geliştirmek faydalı olur. Y'nin yoğunluğu

$$f(y) = |2\pi\Sigma_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right\} \quad (10)$$

şeklinde yazılır. Bu yoğunluk düzenlenirse,

$$f(y) = \exp(g + h'y - \frac{1}{2}y'Ky) \quad (11)$$

ile ifade edilir. Burada $K=\Sigma^{-1}$, $h=\Sigma^{-1}\mu$ ve g normalleştirme sabiti olur ve

$$g = -\frac{1}{2}\ln|\Sigma| - \frac{1}{2}\mu'\Sigma\mu - \frac{q}{2}\ln(2\pi) \quad (12)$$

eşitliği ile verilir. Üstel aile terminolojisinde h ve K kanonik parametreler olarak adlandırılır(Edwards,2001). 'Eş. 11 tekrar dikkate alınarak,

$$f(y) = \exp\left(g + \sum_{j=1}^q h^j y_j - \frac{1}{2} \sum_{j=1}^q \sum_{k=1}^q w^{jk} y_j y_k\right) \quad (13)$$

şeklinde düzenlenebilir ve

$$Y_j \perp Y_k / (\text{geri kalan}) \Leftrightarrow w^{jk} = 0 \quad (14)$$

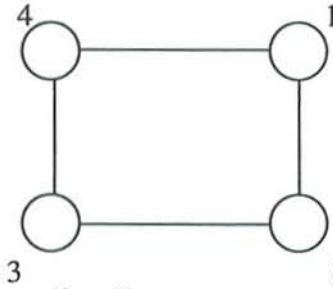
biçiminde ifade edilir.

Grafiksel Gauss modelleri, konsantrasyon matrisinin elemanları yani kısmi korelasyonların sıfır olmasıyla belirlenebilir. Örneğin $q=4$ iken modelde $w^{13}=w^{24}=0$ olduğu düşünölsün. Böylece konsantrasyon matrisi

$$K = \Sigma^{-1} = \begin{pmatrix} w^{11} & w^{12} & 0 & w^{14} \\ w^{21} & w^{22} & w^{23} & 0 \\ 0 & w^{23} & w^{33} & w^{34} \\ w^{41} & 0 & w^{43} & w^{44} \end{pmatrix} \quad (15)$$

şeklinde yazılır.

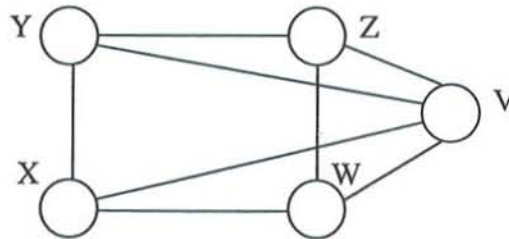
Bu modelin grafiđi, sıfır olmayan kısmi korelasyonlara karşılık gelen çiftler arasında birer kenar konmasıyla edilir. Böyle bir modele ilişkin grafik aşğıdaki gibidir.



Şekil 1. $w^{13}=w^{24}=0$ Durumundaki Grafik

Grafiksel Gauss modellerin formüllendirilmesi için deđişkenleri sayı yerine harf ile ve deđişkenler kümesini Γ ile gösterilecektir.

Kesikli deđişkenler içeren grafiksel modeller için olduđu gibi, model formölü, grafiđin klikleriyle verilen deđişkenler kümesinin (üreteçler) listesini içerir. Örneđin Şekil 2'deki grafik göz önüne alınsın.



Şekil 2. VWX, VWZ, VZY, VXY Kiklerine Sahip Olan Grafik

Kayıp kenarlar WY ve XY'dir. Yani w^{wy} ve w^{xz} sıfır olarak belirlenir. Grafiklerin klikleri $\{V,W,X\}, \{V,W,Z\}, \{V,Z,Y\}, \{V,X,Y\}$ 'dir. Böylece modelin formölü

$$VWX, VWZ, VZY, VXY$$

şeklindedir.

Grafiksel gauss modellerde log lineer modellerde olduğu gibi hiyerarşik ya da grafiksel olmayan model ayrımı yoktur. Tüm modeller grafiksel ve grafiklerle modeller arasında birebir ilişki vardır.

2.1 Sürekli Modeller için Olabilirlik

N gözlemlik $y^{(1)}, \dots, y^{(N)}$ bir örnek alınsın. $\bar{y} = \sum_{k=1}^N y^{(k)} / N$ örnek ortalaması vektörüdür. $S = \sum_{k=1}^N (y^{(k)} - \bar{y})(y^{(k)} - \bar{y})' / N$

örnek kovaryans matrisi olsun. log-yoğunluk

$$\ln(f(y)) = -Nq \ln(2\pi) / 2 - N \ln|\Sigma| / 2 - (y - \mu)' \Sigma^{-1} (y - \mu) / 2$$

şeklinde yazılabilir. Ayrıca log-olabilirliği

$$l(\mu, K) = -Nq \ln(2\pi) / 2 - N \ln|\Sigma| / 2 - \sum_{k=1}^N (y^{(k)} - \mu)' K (y^{(k)} - \mu) / 2$$

dir. Son terim

$$\sum_{k=1}^N (y^{(k)} - \mu)' K (y^{(k)} - \mu) = \sum_{k=1}^N (y^{(k)} - \bar{y})' K (y^{(k)} - \bar{y}) + N(\bar{y} - \mu)' K (\bar{y} - \mu)$$

şeklinde yazılabilir. Bu ifade iz fonksiyonu kullanarak aşağıdaki gibi basitleştirilebilir.

$$\sum_{k=1}^N (y^{(k)} - \bar{y})' K (y^{(k)} - \bar{y}) = Ntr(KS)$$

Böylece log-olabilirliği aşağıdaki eşitlik ile verilir.

$$l(\mu, K) = -Nq \ln(2\pi) / 2 - N \ln|\Sigma| / 2 - Ntr(KS) / 2 - N(\bar{y} - \mu)' K (\bar{y} - \mu) / 2 \quad (16)$$

$a \in \Gamma$ olan değişken altkümesi için, Σ^{aa} , S^{aa} a 'ya karşılık gelen Σ ve S 'in alt matrisleri olsun. q_1, q_2, \dots, q_t üreteçleriyle belirlenen model için, minimal yeterli istatistikler kümesinin, \bar{y} örnek ortalaması ve üreteçlere karşılık gelen örnek kovaryans matrisinin marjinal alt kümelerinin kümesi (S^{aa} , $a=q_1, q_2, \dots, q_t$ için) olduğu gösterilebilir. Olabilirlik eşitlikleri model altında beklenen değerleriyle birlikte bunları (minimal yeterli istatistik) eşitleyerek oluşturulur. Böylece $a=q_1, q_2, \dots, q_t$ için $\hat{\mu} = \bar{y}$ ve $\hat{\Sigma}^{aa} = S^{aa}$ eşitlikleri elde edilir.

Bu sonuçları kullanarak, model altında olabilirliği maksimize etmek için ifade basitleştirilebilir. $\hat{\mu} = \bar{y}$ olduğunda 'Eş. 16'daki son terim yok olur ve $\hat{\Sigma}$ ve S , $w^{ij}=0$ olan elemanlar için kesin olarak farklılık gösteriyorsa $tr(\hat{K}S) = tr(\hat{K}\hat{\Sigma}) = q$ olur. Böylece model altında maksimize edilen log-olabilirlik

$$l_m = -Nq \ln(2\pi) / 2 - N \ln|\hat{\Sigma}| / 2 - Nq / 2$$

ifadesine basitleştirilebilir. Tam model M_f altında $\hat{\Sigma} = S$ 'dir. Böylece bu modelin maksimize edilmiş log-olabilirliği

$$l_m = -Nq \ln(2\pi) / 2 - N \ln|S| / 2 - Nq / 2$$

dir.

Grafiksel modellerde bir kenarın çıkarılıp çıkarılmayacağına ilişkin teste sapma istatistiğinden faydalanılır. Bir M_0 modelinin sapması M_f kısıtsız(doymuş) modeline karşı M_0 'ın olabilirlik oranı testidir. Böylece modelin sapması

$$\begin{aligned} G^2 &= 2(\hat{l}_f - \hat{l}_m) \\ &= N \ln(|\hat{\Sigma}| / |S|) \end{aligned} \quad (17)$$

olur. $M_0 \subseteq M_1$ için sapma farkı

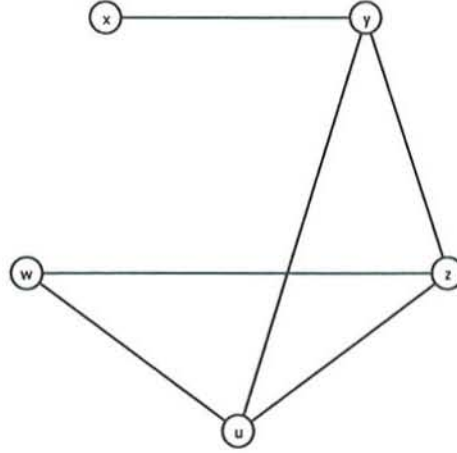
$$d = \ln \left(\frac{|\hat{\Sigma}_0|}{|\hat{\Sigma}_1|} \right)$$

dir (Edwards,2001). $\hat{\Sigma}_0$ ve $\hat{\Sigma}_1$, sırasıyla M_0 ve M_1 altında Σ 'ın tahminleridir. M_0 altında d , M_0 ile M_1 arasında parametre sayılarındaki (köşe sayısı) fark kadar serbestlik dereceli asimptotik χ^2 dağılımıdır.

3 UYGULAMA

Ülkemizde günlük kalorinin sağlanmasında ve halkın beslenmesinde ekmeğin çok önemli bir yeri olmasına karşın ülkemizde üretilen ekmeklerin kalitesi istenen düzeyde değildir. Bunun en önemli nedenlerinden birisi kullanılan buğdayların teknolojik kalitesinin yeterli olmamasıdır. Buğday ve bunlardan elde edilen unların teknolojik kalitelerini belirleyen faktörlerin başında protein ve gluten miktarları ile bunların kalitesi gelir. Protein miktarları birbirine eşit olan buğdayların teknolojik özellikleri arasında da önemli farklar olabilir. Bu çoğu kez gluten miktar ve kalitesindeki farklılıktan kaynaklanır. Bu uygulamada 30 farklı un örneği kullanılmıştır. Un örneklerinin protein miktarı (x), gluten miktarı (y) ve sedimentasyon değerleri (z) ile bu unlardan yapılan ekmeklerin, hacim verimi (w) ve Dallman değerleri (u) tesbit edilmiştir(Özkaya,B. ve diğerleri).

Un ve ekmeğin kalitesini etkileyen kriterler arasında nasıl bir ilişki olduğu ve bu ilişkinin özelliği Kovaryans seçimli model ile incelenmiş ve bu kriterler arasındaki ilişkileri gösteren grafiksel model Şekil 3’de verilmiştir.



Şekil 3. Unun Ekmeklik Kalitesini Etkileyen Faktörler Arası İlişkileri Gösteren Model

Şekil 3 incelendiğinde unun gluten miktarı ve sedimentasyon değeri ile bu undan yapılan ekmeğin Dallman değeri arasında tam bir ilişki bulunmaktadır.

Un örneklerinin protein miktarı (x), gluten miktarı (y) ve sedimentasyon değerleri (z) ile bu unlardan yapılan ekmeklerin, hacim verimi (w) ve Dallman değerleri (u) değişkenlerine ilişkin koşullu bağımsızlıkları gösteren Markov özellikleri Tablo 1’deki gibi yazılabilir.

Tablo 1. Grafiğe İlişkin Koşullu Bağımsızlıklar		
İkili Markov Özelliği	Yerel Markov Özelliği	Global Markov Özelliği
$x \perp w / (u, y, z)$	$x \perp (w, u, z) / y$	$(x, y) \perp w / (u, z)$
$x \perp z / (u, y, w)$	$y \perp w / (x, z, u)$	
$x \perp u / (w, y, z)$		
$y \perp w / (x, z, u)$		

Tablo 1’e göre ikili, yerel ve global markov özellikleri dikkate alındığında şu yorumlar yapılabilir: Unun protein miktarının, gluten miktarı verildiğinde, sedimentasyon değerinden ve bu undan yapılan ekmeğin hacim veriminden koşullu olarak bağımsız olduğu, aynı zamanda unun protein miktarının, gluten miktarı verildiğinde, bu undan yapılan ekmeğin Dallman değerinden de koşullu olarak bağımsız olduğu görülmektedir. Yani unun protein miktarı verildiğinde sedimentasyon değeri hakkında bir değerlendirme yapabilmek için unun gluten miktarının bilinmesine gereksinim bulunmaktadır. Global markov özelliğinden dalman değeri ve sedimentasyon değeri verildiğinde unun protein miktarının ve gluten miktarının bu

unlardan yapılan ekmeklerin hacim veriminden koşullu olarak bağımsız olduğu söylenebilir.

Bu çalışma sonuçlarının kovaryans seçimli model ile irdelenmesinin, un ve ekmek kalitesini gösteren kriterler arasında ilişki olup olmadığını belirlemek bakımından olduğu kadar unun basit yöntemlerle tespit edilen bazı kimyasal ve fizikokimyasal özelliklerine bakılarak ekmeğin kalitesini tahmin edebilme açısından da yararlı olacağı düşünülmektedir.

KAYNAKLAR

- DEMPSTER, A.P., (1972), *Covariance selection*, Biometrics, 28, 157-75
- LAURITZEN S.L., (1996), *Graphical models*, Clarendon press, London
- WHITTAKER, J., (1990), *Graphical models in applied multivariate statistics*, John Wiley and Sons, Chichester
- LAURITZEN, S.L. and WERMUTH, N., (1989), *Graphical models for associations between variables*, some of which are qualitative and quantitative, Annals of Statistics, 17, 31-57
- COX, D.R. and WERMUTH N., (1993), *Linear dependencies represented by chain graph (with discussion)*. Statistical Science, 8, 204-218
- EDWARDS, D., (2001), *Introduction to graphical modelling*. 2nd Edition, Springer-Verlag, New York
- ÖZKAYA, B., BAYRAK, H., GÖKPINAR F., (2002), *Unun Ekmeklik Kalitesini Etkileyen Bazı Faktörler Arasındaki İlişkilerin Kovaryans Seçimli Model İle Belirlenmesi*, Hububat 2002 Hububat işleme teknolojileri kongre ve sergisi, bildiriler kitabı, 549-555 Gaziantep

Graphical Models for Continuous Variables

ABSTRACT

Graphical models for variables, which was distributed multinormal is called covariance selection models. Covariance selection models were determined by conditional independences. In this study, deviance, which is used to test conditional independences, was given. Thirty different flour samples were used in this study. Protein, wet gluten and sedimentation values of flour samples along with the loaf volume and Dallman values of bread made of these flours were investigated. Covariance selection models evaluated the obtained results

Keywords: *Covariance Selection Model, Deviance, Markov Properties, Conditional Independence*

Bayes Ağlarda Koşullu Bağımsızlıkların İncelenmesi Üzerine Bir Çalışma

Hülya OLMUŞ*

Semra ORAL ERBAŞ**

ÖZET

Bir Bayes ağ, koşullu bağımsızlık özelliklerine sahip yön verilmiş döngüsel olmayan bir grafiktir. Bayes ağ, değişkenler ve değişkenler arası yön verilmiş kenarların kümesinden oluşur. Kenarlar, değişkenler arası olasılık bağımlılıkları gösterir. Bu bağımlılıklar koşullu olasılıkların kümesinden oluşur. Her bir değişkenin ebeveynleri verildiğinde değişkenin koşullu olasılığı belirlenir. Bir düğümün ebeveynleri olmadığı zaman, bir değişken koşulsuz (marjinal) bir olasılığa sahiptir. Bu çalışmada, Bayes ağlarda, koşullu bağımsızlıklar aşağıdaki farklı üç yoldan araştırılmıştır. Bu yollardan ilki, yön verilmiş Markov özelliğidir. İkincisi, moral ve üçgen grafik yardımıyla elde edilebilen koşullu bağımsızlıktır. Moral ve üçgen grafikten elde edilebilen takımlar sayesinde birleşme ağacı kurulur. Birleşme ağacından, verilen Bayes ağ modeline ilişkin koşullu bağımsızlıklar elde edilir. Üçüncüsü ise, koşullu bağımsızlık kavramının yönsel-ayırılma kriteri ile verilmesidir. Verilen Bayes ağ modeli için, üç farklı şekilde koşullu bağımsızlık özellikleri gösterilmiş ve bu yollar arasındaki ilişkiler incelenmiştir.

Anahtar Sözcükler: Yön Verilmiş Döngüsel Olmayan Grafik, Yön Verilmiş Grafik, Yön Verilmemiş Grafik, Koşullu Bağımsızlık, Markov Bağımsızlık, Moral Grafik, Üçgen Grafik, Birleşme Grafiği.

1. GİRİŞ

Grafik; $G=(V,E)$, düğümlerin (köşeler) sonlu kümesi V ve bu düğümler arası kenarların (bağlantılar) sonlu kümesi E 'den oluşan bir yapıdır (Edwards, 1995). E kümesi V 'den alınan değişken çiftlerinden oluşur. Yani; E , $\{a,b\}$ düzenlenmiş çiftler kümesi olup her $\{a,b\}$ de V 'nin elemanlarıdır.

Grafik modelleri; düğümlerin rasgele değişkenlerle gösterildiği ve kenarların koşullu bağımsızlık özellikleri gösterdiği grafiktir. Yön verilmiş ve yön verilmemiş grafik modeller olarak iki şekilde incelenir. Bayes ağlar, yön verilmiş grafik modelleri, Markov ağlar, yön verilmemiş grafik modelleri kullanırlar (Buntine, 1996).

*Araş. Gör., Gazi Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü, e-mail: hulya@gazi.edu.tr
(Haberleşme adresi)

**Prof. Dr., Gazi Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü, e-mail: serbas@gazi.edu.tr

İki düğüm arasındaki kenarlar, oklarla gösterilmiş ise (bir yöne sahip ise), buna yön verilmiş grafik denir. a 'dan b 'ye bir ok çizilmiş ise, a , b 'nin nedeni demektir.

Bu durumda $a \rightarrow b$ şeklinde gösterilir. Bir G grafiğinde tüm kenarlar çizgi şeklinde ise, G 'nin yapısına yön verilmemiş grafik denir. Bu durumda $\{a,b\} \in E$ ve $\{b,a\} \in E$ dir (Şekil 1), (Oral Erbaş, Bayrak, 1999).



Şekil 1. (a) Yön Verilmemiş Grafik
(b) Yön Verilmiş Grafik

Şekil 1'de (a) ve (b) grafikleri, aynı düğümler arası kenarlara sahiptir. (b) grafiği yön verilmiş kenarlara sahip iken, (a) grafiği yön verilmemiş kenarlara sahiptir. (a) grafiği $V=\{A,B,C,D\}$ düğümler kümesi ve

$E=\{(A,B),(B,A),(A,C),(C,A),(A,D),(D,A)\}$ kenarlar kümesine sahiptir. (b) grafiği ise düğümler kümesi aynı ancak $E=\{(A,C),(A,D),(B,A)\}$ kenarlar kümesine sahiptir.

Grafik, yön verilmiş ve döngüsel olmama durumlarının her ikisini de sağlar ise, grafiğe yön verilmiş döngüsel olmayan grafik denir (Liarokapis, 1999).

2. GRAFİK TEORİSİNDE TEMEL KAVRAMLAR

Düğümlerin sıralanmasından oluşan yapıya yol denir. Örneğin, A ve B düğümleri arası herhangi bir yol, $A \rightarrow C \leftarrow D \rightarrow E \rightarrow F \leftarrow B$ ile verilebilir.

$A \rightarrow \dots \rightarrow B$ biçiminde bir yol A 'dan B 'ye yön verilmiş yol olarak adlandırılır.

Aynı düğüm ile başlayan ve biten yola döngü denir. Örneğin, $A \rightarrow \dots \rightarrow A$ biçiminde ise, bir döngü söz konusudur.

Yön verilmiş bir grafikte ebeveyn/çocuk ilişkilerinden söz edilir. A 'dan B 'ye bir kenar varsa A , B 'nin ebeveyni ve B , A 'nın çocuğudur denir ($A \rightarrow B$). Dede/torun ilişkisi, ebeveyn/çocuk ilişkisinin uzantısıdır. A , B 'nin ebeveyni ve B , C 'nin ebeveyni ise ($A \rightarrow B \rightarrow C$), A, C 'nin dedesi ve C , A 'nın torunudur denir. Aile, düğüm ve düğümün ebeveynlerinden oluşan bir kümedir (Richardson, 1997).

3. KOŞULLU BAĞIMSIZLIKLAR İLE İLGİLİ TEMEL ÖZELLİKLER

$G=(V,E)$ grafiğinde, A,B ve C , V 'nin alt kümelerini gösterebilir. Bir G grafiğinde, C düğümü verildiğinde A düğümünün, B düğümünden koşullu bağımsızlığı $A \perp B \mid C$ ile gösterilebilir. Böylece, koşullu bağımsızlık özellikleri aşağıdaki gibidir:

- $A \perp B$ ise $B \perp A$ 'dır.
- $A \perp B \mid C$ ve U, A 'nın alt kümesi ise $U \perp B \mid C$ 'dir.
- $A \perp B \mid C$ ve U, B 'nin alt kümesi ise $A \perp B \mid (C \cup U)$ 'dir.
- $A \perp B \mid C$ ve $A \perp D \mid (B \cup C)$ ise $A \perp (B \cup D) \mid C$ 'dir (Lauritzen, 1996; Lauritzen ve diğerleri, 1990).

4. BAYES AĞLARDA TOPOLOJİK SIRALAMA

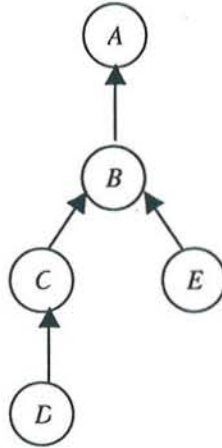
Yön verilmiş döngüsel olmayan grafikler için; bir X düğümünün ebeveynleri ($eb(X)$) topolojik sıralamada önce gelir. Bundan dolayı, yön verilmiş döngüsel olmayan grafikler, düğümlerin doğrusal bir sıralanmasına sahiptir. Düğümlerin böyle sıralanması topolojik sıralama olarak adlandırılır (Cowell, 1999). Topolojik sıralamayı bulmak için basit bir algoritma şöyle olabilir. İlk olarak, grafik ve boş bir liste alınır. Ebeveyne sahip olmayan herhangi bir düğüm grafikten çıkarılır ve tüm düğümler listenin sonuna eklenerek sıralaması elde edilmiş olur. Grafiğin döngüsel olduğu düşünüldüğünde, hiçbir düğüm, ebeveyne sahip olmadığından grafik, birçok durumda sağlanır. Bu algoritma, grafiğin döngüsel olduğunu kontrol etmenin bir yoludur. Başka bir durum ise, çocuksuz düğüm grafikten çıkarılıp tüm düğümler listenin başına eklenerek elde edilir.

5. OLASILIK HESAPLAMALARI İÇİN NİÇİN BAYES AĞLARA İHTİYAÇ DUYULUR?

Bayes ağlar, farklı değişkenler arası bağımlılıkları açıklar. Bayes ağlar'da doğrudan bağımlılıklardan söz edilir. İstenilen tüm olasılıkları hesaplamak için, koşullu bağımsız olan düğümlerden yararlanılır. Böylece, olasılıkları hesaplamak için daha az sayıda işleme gerek duyulur. Düğümler, koşullu olarak bağımsız olduğu zaman, ortak olasılık dağılımını hesaplamak basitleşir. Örneğin; beş değişkenden (düğüm) oluşan A,B,C,D,E ağına sahip olunduğu düşünülür. Olasılık teorisinde bilinen zincir kuralı ile $P(A,B,C,D,E)$ ortak olasılık dağılımı hesaplanabilir. Böylece bu beş düğüme ilişkin ortak olasılık dağılımı, aşağıdaki şekilde yazılır.

$$P(A,B,C,D,E)=P(A \mid B,C,D,E).P(B \mid C,D,E).P(C \mid D,E).P(D \mid E).P(E)$$

Ancak; bir Bayes ağda bağımlılıklar daha kolay görülür. Bunun için aşağıda bir ağ verilmiştir (şekil 2). Bu modele ilişkin topolojik sıralama (A,B,C,E,D) 'dir.



Şekil 2. Beş Dügüme İlişkin Yön Verilmiş Döngüsel Olmayan Grafik

Şekil 2’den bu beş düğüme ilişkin, $P(A,B,C,D,E)$ ortak olasılık dağılımını hesaplamak basitleşir. Bu hesaplama,

$$P(A,B,C,D,E) = P(A|B) \cdot P(B|C,E) \cdot P(C|D) \cdot P(D) \cdot P(E)$$

ile elde edilir.

6. BAYES AĞLARDA KOŞULLU BAĞIMSIZLIKLAR

6.1. Bayes Ağlar da Koşullu Bağımsızlığın Yön Verilmiş Markov Özelliği

Bayes ağlarda, keyfi olarak sıralanan (X_1, X_2, \dots, X_n) gibi n tane kesikli değişkenler üzerinde P dağılımına sahip olunduğu varsayalım. Değişkenlerin kümesi $U = \{X_1, X_2, \dots, X_n\}$ ve bu değişkenlerin ortak olasılık dağılımı $P(U)$ olsun. Bir Bayes ağ, U üzerinde temsil edilir ve koşullu olasılıklar U için sağlanır. O zaman; $P(U)$, ağda belirlenen koşullu olasılıklardan hesaplanır. Düğümün ebeveynleri üzerinde koşullandırma yapılarak, her düğüm için koşullu olasılıklar $P(X_j|eb_j)$ belirlenir. Tüm değişkenler kümesi üzerinde ortak olasılık,

$$P(U) = \prod_j P(X_j | eb_j) \quad (1)$$

ifadelerin çarpımıyla verilir. Eş. 1, Bayes ağda belirlenen tüm koşullu olasılıkların çarpımıdır (Jensen, 1996).

Yön verilmiş döngüsel olmayan grafiğe, geri dönüşlü faktörleştirme de denir. Bu, yön verilmiş döngüsel olmayan grafik üzerinde oluşan dağılım olarak da bilinir (Cowell, 1999).

Yön verilmiş Markov özelliği, koşullu bağımsızlık özelliğidir. Bir değişkenin ebeveynleri verildiğinde $(eb(X))$ torunları olmayan değişkenlerden $(nd(X))$ koşullu bağımsız olma durumudur (Cowell, 1999).

$$X \perp_{nd} (X) \setminus eb(X) \quad (2)$$

eb_j , X_j 'nin atalarının minimal bir kümesi ise X_j , diğer tüm atalarından bağımsızdır. eb_j , X_j 'nin Markov ailesidir. Başka bir deyişle; eb_j ,

$$P(x_j / eb_j) = P(x_j / x_1, x_2, \dots, x_{j-1}) \quad (3)$$

eşitliğini sağlayan $\{x_1, x_2, \dots, x_{j-1}\}$ 'in herhangi alt kümesidir. Bu yapı, koşullu bağımsızlık ilişkilerini taşıması nedeni ile Bayes ağı tanımlar.

Eş.1'e göre; eb_j ebeveyn kümesi verildiğinde, her X_j değişkeni, diğer geriye kalan tüm değişkenlerden koşullu bağımsızdır denir ($\{X_1, X_2, \dots, X_{j-1}\} \setminus eb_j$). Bağımsızlığın kümesi aşağıdaki şekilde yazılabilir.

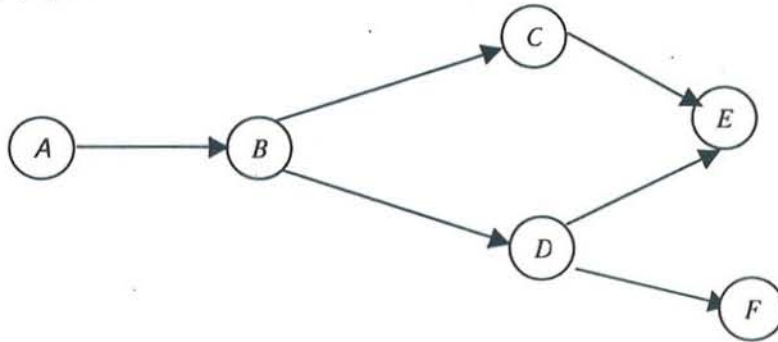
$$X_j \perp \{X_1, X_2, \dots, X_{j-1}\} \setminus eb_j \quad i=1, 2, \dots, n \quad (4)$$

Bu tür bağımsızlıklara Markov bağımsızlığı denir (Pearl, 2000).

Bir olasılık fonksiyonu P , yön verilmiş döngüsel olmayan grafik G 'ye bağlı Eş. 3' deki ayrıştırmayı sağlarsa G , P 'ye karşılık gelir. G ve P tutarlıdır ya da P , G 'ye Markov bağlantılıdır denir ve grafiğin olasılığa karşılık gelmesi ya da koşullu olasılıkların tanımlanması olarak düşünülebilir (Pearl, 2000).

Kesikli değişkenlerin $U = \{X_1, X_2, \dots, X_n\}$ kümesi üzerinde P ortak olasılık dağılımı ve yön verilmiş döngüsel olmayan grafik G verilsin. G 'de düğümler ve X 'de değişkenler arası bire-bir uygunluk varsa G , P 'yi temsil eder. Yani; P , Eş.1'deki ayrışım ile gösterilir (Pearl, 1993).

Örneğin; şekil 3'de, verilen yön verilmiş döngüsel olmayan grafik için Eş.1'e göre ortak olasılık aşağıda verilmiştir. Bu modele ilişkin topolojik sıralama (A, B, C, D, E, F) 'dir.



Şekil 3. Altı Düğümden Oluşan Bayes Ağ Modeli

Elde edilen Markov bağımsızlıkları aşağıdadır.

1. $E \perp \{A, B\} \setminus \{C, D\}$
2. $F \perp \{A, B, C, E\} \setminus \{D\}$
3. $\{C, D\} \perp \{A\} \setminus \{B\}$

6.2. Birleşme Ağacı İle Koşullu Bağımsızlıkların İncelenmesi

İlk olarak birleşme ağacının kurulması ile ilgili gereken tanımlamalar verilsin (Geiger et. al.,1990;Stephenson, 2000).

Bir alt grafik tam ise, bu alt grafiğe takım denir. Bir grafik verildiğinde, takımların tanımlanmasına gereksinim duyulur ve grafik modellerde önemli bir kavramdır. Verilen herhangi bir yol için giriş, kenardır. Yön verilmemiş grafiklerde, ard arda gelen düğüm çiftleri komşudur ve eğer döngüdeki ardışık düğüm çiftlerinden başka komşu yoksa o zaman döngü kirişsizdir denir (Stephenson, 2000).

Birleşme ağacını oluşturmak için aşağıda adımlar sıralanmıştır.

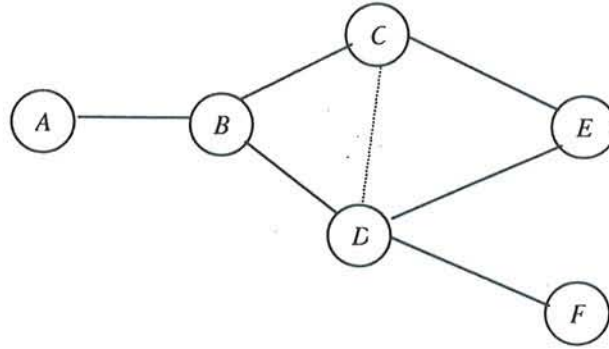
Adım 1: Moral grafik, her düğümün ailelerinin evlendirilmesi ile oluşan yön verilmemiş bir grafikdir (Lauritzen and Spiegelhalter, 1988). Her aile arasına yön verilmemiş kenar eklenir. Bundan sonra, orijinal kenarların hepsinin yönleri kaldırılır.

Adım2: Üçgenleştirilmiş grafik, yön verilmemiş bir grafikdir. Burada her döngü, en az dört düğüm ve en az bir kirişe sahip olmalıdır (Jensen, 1996). Üç düğüme sahip olan döngünün herhangi bir kirişe sahip olması mümkün değildir.

Verilen Bayes ağ, moral grafik haline getirildikten sonra, takımlar, bazen kolaylıkla görülmemektedir. Bundan dolayı, moral grafiğe, yapay bağlantılar eklenerek, grafik üçgenleştirilmiş grafik haline getirilir. Takımlar da, bu grafikten görülebilir. Takımlar sayesinde, değişkenler arasındaki koşullu bağımsızlık ilişkileri kolaylıkla görülmektedir. Elde edilen moral grafik, üçgenleştirilmiş grafikdir. Bu grafikte uzunluğu 3'den büyük her döngü kirişe sahip olmalıdır. Üçgenleştirilmiş grafiği elde edebilmek için büyük ağlarda, yapay bağlantıların nasıl ekleneceği literatürde vardır. Yapay bağlantıların minimal sayısının bulunması problemi NP-hard olarak adlandırılır (Jensen et.al., 1990). Bu çalışma da, amacımız, fikir takımlarının minimal kümesini alarak yapay bağlantıları eklemektir.

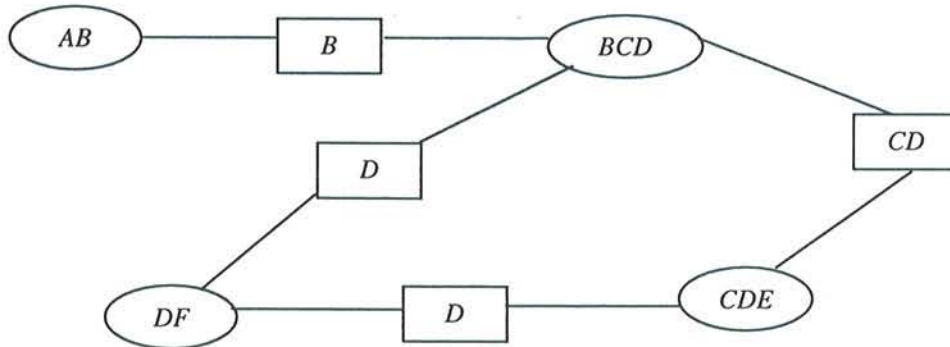
Adım 3: Birleşme grafiği, her A değişkeni, $eb(A) \neq \emptyset$ ile $eb(A) \cup \{A\}$ takımı ile kurulur. Her iki takım arasına, boş olmayan arakesit eklenir. Elde edilen grafiğe, birleşme grafiği denir. Elde edilen birleşme grafiği, bir döngüye sahip ise kenarların her biri döngüleri kırarak ortadan kaldırılır. Döngü üzerindeki tüm arakesitler aynı değişkeni kapsar. Böyle bir durum söz konusu ise, döngü ortadan kaldırılarak birleşme ağacı elde edilir.

Moral grafik, üçgenleştirilmiş bir grafik ise moral grafik için birleşme grafiği, birleşme ağacına sahiptir (Jensen, 1996). Şekil 3 için moral ve üçgen grafik aşağıda verilmiştir.



Şekil 4: Moral ve Üçgen Grafik

Birleşme ağaçlarını kurabilmek için, ilk önce takımlar elde edilir. B düğümünün ebeveyni $eb(B)=\{A\}$ ile $\{A,B\}$ takımı ve $\{C,D\}$ düğümünün ebeveyni $eb\{C,D\}=\{B\}$ ile $\{B,C,D\}$ takımları ele alınır. Bu iki takım B arakesiti ile bağlanır. Daha sonra, $eb(E)=\{C,D\}$ ile (C,D,E) takımı oluşturulur. (C,D,E) ve (B,C,D) takımları da $\{C,D\}$ arakesiti ile bağlanır. $eb(F)=\{D\}$ ile (F,D) takımı elde edilir. $\{F,D\}$ takımı, (C,D,E) ve (B,C,D) takımları ile $\{D\}$ arakesiti ile bağlanır. Ortak düğüme sahip olan takımlar arakesitleri ile bağlanarak birleşme grafiği elde edilir (şekil 5). Arakesitler, kare kutular ile gösterilmiştir.



Şekil 5: Birleşme ağacı

Elde edilen birleşme ağacına göre koşullu bağımsızlıklar şöyledir.

1. $\{B\} \perp \{E\} \setminus \{C,D\} \Rightarrow \{E\} \perp \{A,B\} \setminus \{C,D\}$
2. $\{B,C\} \perp \{F\} \setminus \{D\} \Rightarrow \{B,C,A,E\} \perp \{F\} \setminus \{D\}$
 $\{C,E\} \perp \{F\} \setminus \{D\} \Rightarrow \{A,B,C,E\} \perp \{F\} \setminus \{D\}$
 $\{A,B,C,E\} \perp \{F\} \setminus \{D\}$
3. $\{A\} \perp \{C,D\} \setminus \{B\}$

6.3.Yönsel-ayrılma Kriteri

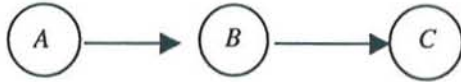
6.3.1. Bayes Ağlarda Olayların Tanımlanması

İki çeşit olay vardır: kuvvetli olay, zayıf olay (Fenton, 1997). Kuvvetli olay: X rasgele değişkeni için kuvvetli olay, X ile ilgili mümkün durumların kesin bir değer alması olayıdır. Örneğin; X , futbol takımı için belli bir oyunun sonuçlarını (kazanma, kaybetme, berabere kalma) gösterebilir. Bu olayın kuvvetli bir olay olması, bilgi alanı maçın kesinlikle kazanılması olayıdır. Bu durumda kazanma değeri olarak bilinen X rasgele değişkeni kuvvetli olaydır. Zayıf olay: X ile ilgili durumların önsel olasılık değerlerinin güncelleştirilmesine imkan sağlayan herhangi bir olaydır. Örneğin, bir futbol takımının ilk yarı devrede skoru 3-0 olduğu bilindiğinde, takımın maçı kazanma olasılığının yüksek olduğu söylenebilir. Bunun yanında kaybetme ve berabere kalma olasılıklarının her ikisi de oldukça düşüktür. Kaybetme ve berabere kalma değişkenleri zayıf olaydır.

6.3.2. Bayes Ağlarda Bağlantı Durumlarının Tanımlanması

Burada üç tane bağlantı durumundan söz edilir: serisel bağlantı, iraksayan bağlantı, yakınsayan bağlantı (Jensen, 1996; Fenton, 1997).

Serisel bağlantı: Şekil 6'daki Bayes ağı göz önüne alınsın. A, B ve C değişkenleri doğru ve yanlış mümkün durumlarıyla kesikli bir değişkendir. A olayının gerçekleştiği varsayılınsın. Bu bilgi, sırasıyla B ve C olaylarının gerçekleşebileceğini artırır. A hakkındaki olay, B üzerinden C 'ye iletilebilmelidir.

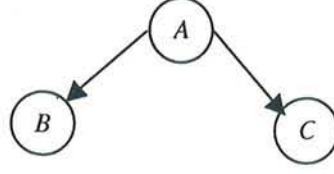


Şekil 6. Serisel Bağlantı

B olayının bilindiği varsayılınsın (B olayı kuvvetli bir olaydır). Bu durumda, A olayı, C ile ilişkisizdir. A olayı ile C olayı bağımsız olarak oluşur. A 'da ki olay C 'ye iletilemeyecektir. Serisel bağlantı da, B bilinmiyorsa ise olay, A 'dan C 'ye iletilebilecektir. Ancak; B verildiğinde, A ve C yönsel-ayrılmıştır.

Sonuç olarak, bağlantıda bir değişkenin durumu bilinmiyorsa, olay serisel bağlantı vasıtasıyla iletilebilmelidir.

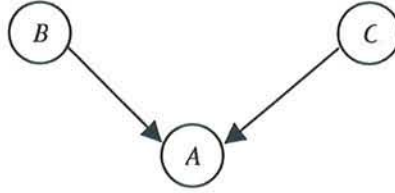
Iraksayan bağlantı: Şekil 7'deki Bayes ağı düşünölsün. A, B , ve C değişkenleri doğru ve yanlış mümkün durumlarıyla kesikli bir değişkendir. A hakkındaki herhangi bir olay, B ve C 'nin her ikisine de iletilmelidir. Başka bir deyişle; A olayı, B ve C olaylarının her ikisinin de kesinliğini etkileyecektir. A olayındaki fikir kuvvetli ise, sırasıyla B ve C olaylarındaki fikir artar.



Şekil 7. İraksayan Bağlantı

B hakkındaki bilginin C 'ye iletilip iletilemeyeceği ile ilgilenilsin. A ile ilgili olayın kuvvetli bir olay olduğu düşünölsün. Bu durumda, B hakkındaki olay, C hakkındaki fikir herhangi bir durumda deęiřtirmeyecektir. Yani, C 'nin deęeri sadece A 'nın kesinlięi ile etkilenecektir. A bilindięi zaman B ve C baęımsız olarak oluřur. B ve C 'nin baęımsızlıęı, A 'nın kesinlięi ile kořullandırılmıřtır. Bundan dolayı, A verildięinde, B ve C kořullu olarak baęımsızdır. İraksayan baęlantı ile; A olayı bilinmiyorsa, A üzerindeki olay B 'den C 'ye iletilecektir. A verildięinde B ve C yönsel ayrılmıřtır.

Yakınsayan baęlantı: Şekil 8'deki Bayes aęı düşünölsün. A , B ve C deęiřkenleri doęru ve yanlış mümkün deęerleri ile kesikli bir deęiřkendir.



Şekil 8. Yakınsayan Bağlantı

B ya da C hakkındaki herhangi olay A 'ya iletilebilecektir. A olayının, B ve C arasındaki olaylara iletilip iletilemeyeceęi ile ilgilenilir. A hakkında bilgi yoksa, A 'nın aileleri B ve C baęımsızdır. Bununla beraber, A hakkında herhangi bir řey biliniyorsa, (zayıf olay) A 'nın aileleri baęımlı olarak oluřur. B ve C , A üzerinde kořullu olarak baęımlıdır. A yakınsayan düęümü, zayıf ya da kuvvetli olayları içerirse, olay sadece B ve C 'nin aileleri arasında iletilecektir.

Sonuç olarak, serisel ve iraksayan baęlantılar durumunda, kořullu baęımsızlıęın saęlanması için, kuvvetli bir olay gerekir. Ancak, yakınsayan bir baęlantıda, kořullu baęımsızlıęın saęlanması için, zayıf ya da kuvvetli bir olay gerekmektedir (Fenton, 1997).

Yukarıda verilen üç durumda, olayın bir deęiřken vasıtasıyla iletilebileceęini, tüm durumlarda göröldü. Bu kriter, aę içerisinde bilinen bir olay verildięinde, herhangi deęiřken çiftlerinin baęımlı olup olmayacaęına karar vermek için imkan saęlar. Kořullu baęımsızlıklar, yönsel-ayırma kriteri kullanılarak görölebilir. Kural ařaęıdaki biçimde ifade edilmiřtir.

A ve B değişkenleri arası tüm yollar için X orta değişkeni *var ise*,

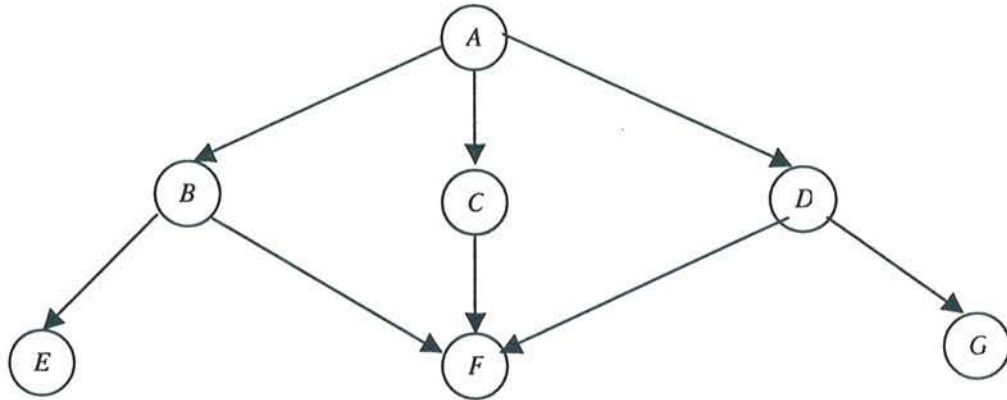
- bağlantı serisel ya da iraksayan ve X 'nin durumu bilinmesi durumunda
- bağlantı yakınsayan ve ne X ne de X 'nin herhangi torunları olayda bulunmaması durumunda A ve B değişkenleri yönsel-ayrılmıştır denir.

A ve B değişkenleri yönsel-ayrılmış değil ise, A ve B değişkenlerine yönsel-bağlantılıdır denir (Madsen and Jensen, 1999; Neal, 2000).

Şekil 3 için yönsel-ayrılma kriterine göre elde edilen koşullu bağımsızlıklar şöyledir:

- * $\{E\} \perp \{A\} \setminus \{C\}$
- $\{E\} \perp \{B\} \setminus \{C\}$
- $\{E\} \perp \{A\} \setminus \{D\} \Rightarrow \{E\} \perp \{A,B\} \setminus \{C,D\}$
- $\{E\} \perp \{B\} \setminus \{D\}$
- * $\{E\} \perp \{F\} \setminus \{D\}$
- $\{A\} \perp \{F\} \setminus \{D\} \Rightarrow \{F\} \perp \{A,B,E\} \setminus \{D\} \Rightarrow \{F\} \perp \{A,B,C,E\} \setminus \{D\}$
- $\{B\} \perp \{F\} \setminus \{D\}$
- * $\{A\} \perp \{C\} \setminus \{B\}$
- $\{A\} \perp \{D\} \setminus \{B\} \Rightarrow \{A\} \perp \{C,D\} \setminus \{B\}$

Değişik Bayes ağ modelleri üzerinde üç farklı şekilde koşullu bağımsızlıklar elde etmeye çalışalım.



Şekil 9. Yön Verilmiş Döngüsel Olmayan Grafik

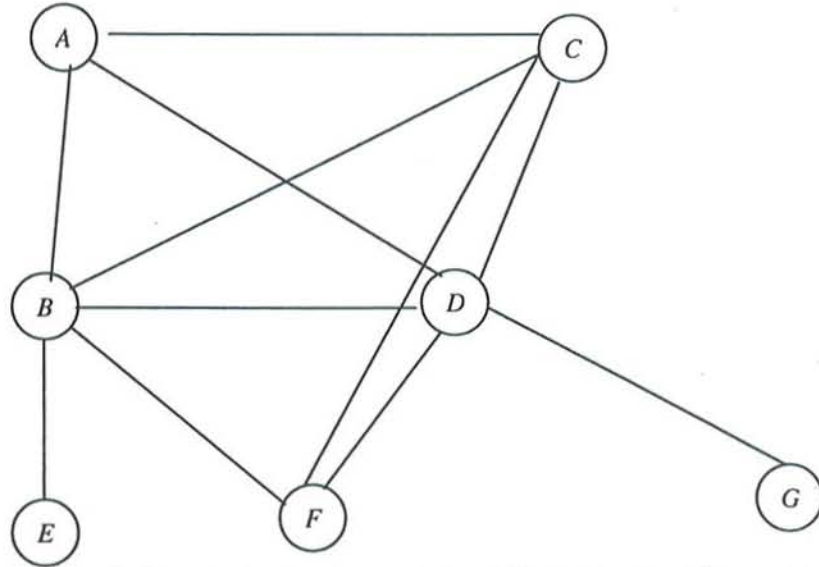
Bu modele ilişkin, topolojik sıralama (A,B,C,D,E,F,G) 'dır.

Ortak olasılık dağılımı,

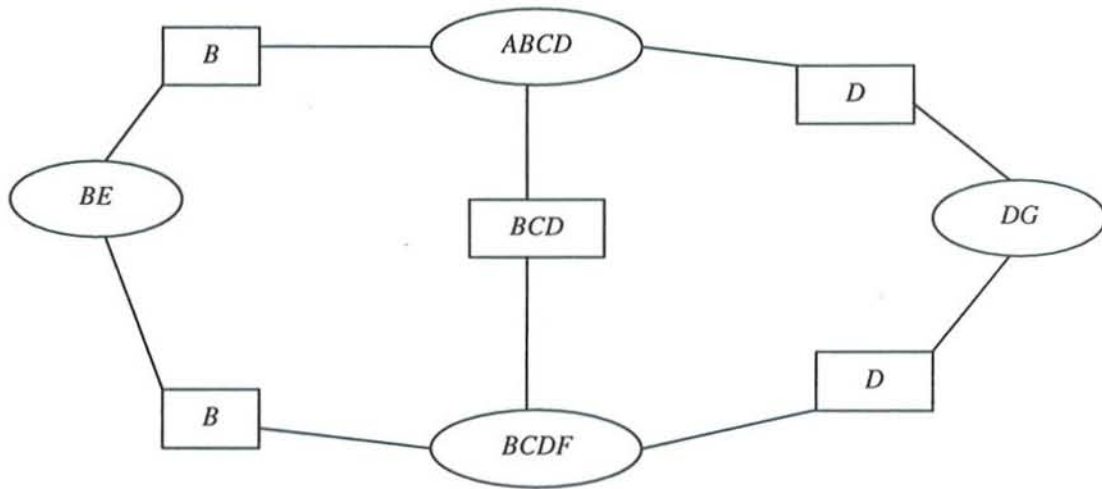
$$P(A,B,C,D,E,F,G) = P(A) \cdot P(B|A) \cdot P(C|A) \cdot P(D|A) \cdot P(E|B) \cdot P(G|D) \cdot P(F|B,C,D)$$

olarak ifade edilir.

Genel Markov Koşulu	Birleşme ağacına göre	Yönel-ayırma kriteri
$F \perp \{A,E,G\} \setminus \{B,C,D\}$	$\{A\} \perp \{F\} \setminus \{B,C,D\} \Rightarrow F \perp \{A,E,G\} \setminus \{B,C,D\}$	$\{A\} \perp \{F\} \setminus \{B\}$ $\{A\} \perp \{F\} \setminus \{D\} \Rightarrow \{A\} \perp \{F\} \setminus \{B,C,D\}$ $\{A\} \perp \{F\} \setminus \{C\}$
$E \perp \{A,C,D,F\} \setminus \{B\}$	$\{E\} \perp \{A,C,D\} \setminus \{B\}$ $\{E\} \perp \{C,D,F\} \setminus \{B\} \Rightarrow E \perp \{A,C,D,F\} \setminus \{B\}$	$\{E\} \perp \{F\} \setminus \{B\} \Rightarrow \{E\} \perp \{A,C,D,F\} \setminus \{B\}$
$G \perp \{A,B,C,E,F\} \setminus \{D\}$	$\{G\} \perp \{B,C,F\} \setminus \{D\}$ $\{G\} \perp \{A,B,C\} \setminus \{D\} \Rightarrow G \perp \{A,B,C,E,F\} \setminus \{D\}$	$\{F\} \perp \{G\} \setminus \{D\} \Rightarrow$ $\{G\} \perp \{A,B,C,E,F\} \setminus \{D\}$

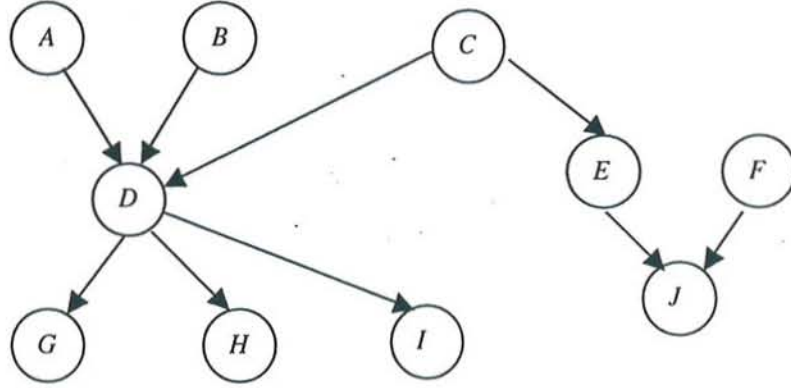


Şekil 10. Yön Verilmiş Döngesel Olmayan Grafik İçin Moral ve Üçgen Grafiği



Şekil 11. Yön Verilmiş Döngesel Olmayan Grafik İçin Birleşme Ağacı

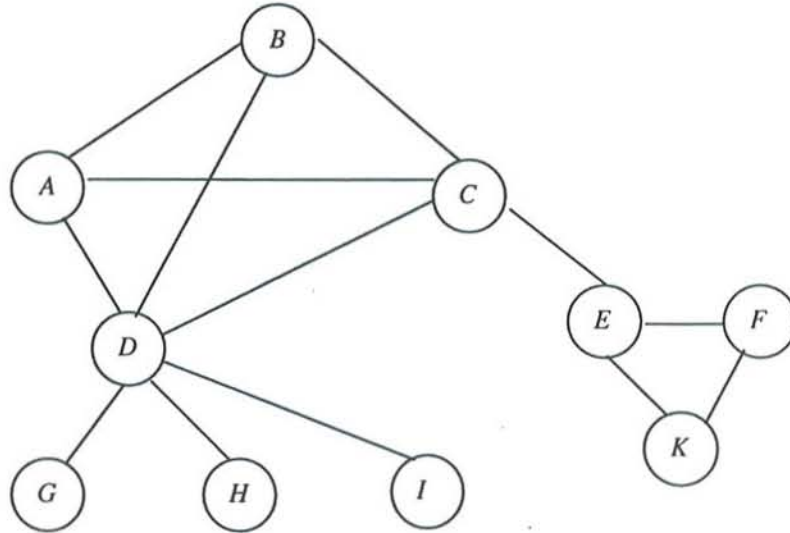
Başka bir Bayes ağ modeli şekil 12'de gösterildiği gibi olsun. Bu modele ilişkin topolojik sıralama (A,B,C,D,E,F,J,G,H,I)'dır.



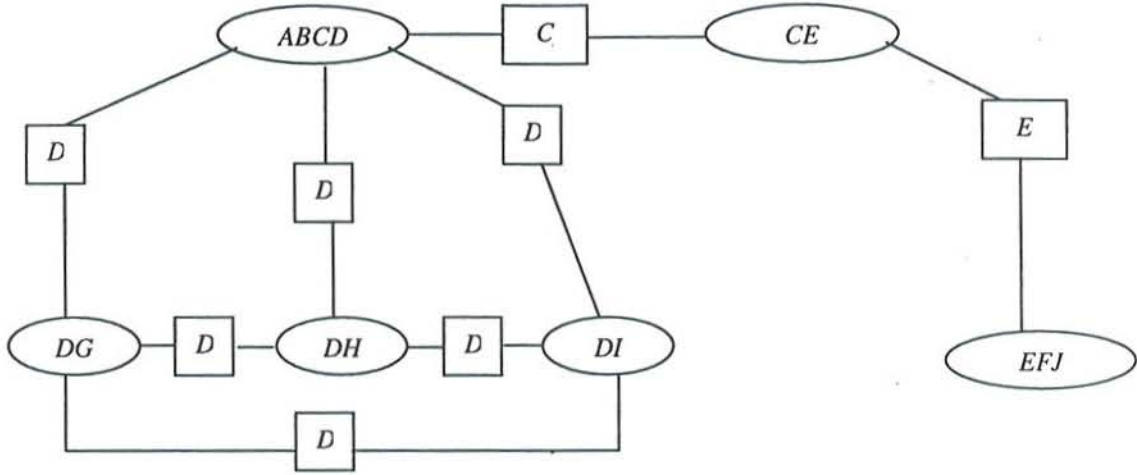
Şekil 12. On Düğümden Oluşan Bayes Ağ Modeli

Şekil 12'ye göre ortak olasılık dağılımı,
 $P(A,B,C,D,E,F,G,H,I,J)=P(A).P(B).P(C).P(D/A,B,C).P(E/C).P(G/D).P(H/D).P(I/D).P(JE,F)$ 'dır.

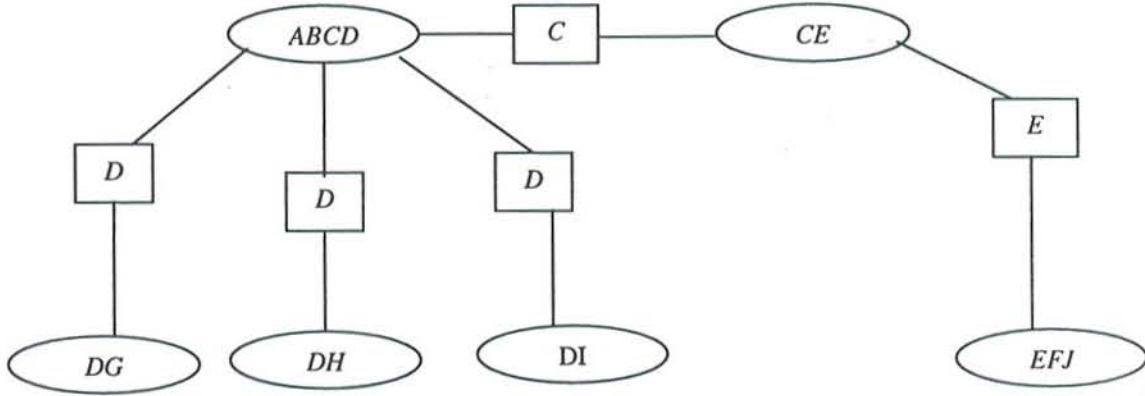
Genel Markov Koşulu	Birleşme ağacına göre	Yönel-ayrılma kriteri
$\{G,H,I\} \perp \{A,B,C\} \setminus \{D\}$	$\{A,B,C\} \perp \{G\} \setminus \{D\}$ $\{A,B,C\} \perp \{H\} \setminus \{D\} \Rightarrow \{G,H,I\} \perp \{A,B,C\} \setminus \{D\}$ $\{A,B,C\} \perp \{I\} \setminus \{D\}$	$\{A\} \perp \{H,G,I\} \setminus \{D\}$ $\{C\} \perp \{H,G,I\} \setminus \{D\} \Rightarrow \{G,H,I\} \perp \{A,B,C\} \setminus \{D\}$ $\{B\} \perp \{H,G,I\} \setminus \{D\}$
$\{J\} \perp \{A,B,C,D,G,H,I\} \setminus \{E,F\}$	$\{C\} \perp \{F,J\} \setminus \{E\} \Rightarrow$ $\{J\} \perp \{A,B,C,D,G,H,I\} \setminus \{E,F\}$	$\{C\} \perp \{J\} \setminus \{E\} \Rightarrow$ $\{J\} \perp \{A,B,C,D,G,H,I\} \setminus \{E,F\}$
$\{E\} \perp \{A,B,D,G,H,I\} \setminus \{C\}$	$\{A,B,D\} \perp \{E\} \setminus \{C\} \Rightarrow$ $\{E\} \perp \{A,B,D,G,H,I\} \setminus \{C\}$	$\{D\} \perp \{E\} \setminus \{C\} \Rightarrow$ $\{E\} \perp \{A,B,D,G,H,I\} \setminus \{C\}$



Şekil 13. Yön Verilmiş Döngüsel Olmayan Grafik İçin Moral ve Üçgen Grafiği



Şekil 14. Yön Verilmiş Döngüsel Olmayan Grafik İçin Birleşme Grafiği



Şekil 15. Yön Verilmiş Döngüsel Olmayan Grafik İçin Birleşme Ağacı

7. SONUÇ VE ÖNERİLER

Bayes ağlar, grafikte koşullu olasılıkları güncelleştirmek için etkili metotlar geliştirir. Olasılık modellerini ortaya koyar. Kesin olmayışlığı içeren bilgi alanlarını biçimlendirir. Ayrıca, değişkenler arasındaki bağımlılık ve bağımsızlık kavramlarını etkili bir şekilde açıklar.

Bayes ağ modellerinde, modelin belirlenmesi önemlidir. Çünkü seçilen modele göre koşullu bağımsızlıklar değişecektir. Araştırmacı, amacına uygun bir şekilde modelini belirlemeye çalışmalıdır. Bu konu literatürde, uzman sistemler adı altında oldukça geniş bir yer kaplar. En iyi modelin belirlenmesinde uzman sistemlerin görüşlerine başvurulmalıdır.

Bu çalışmada kesikli değişkenler için Bayes ağlarda önemli olan koşullu bağımsızlık kavramı üzerinde durulmuştur. Koşullu bağımsızlık, yön verilmiş döngüsel olmayan grafikler ve nedensellik göz önüne alınarak anlatılmıştır. Bayes

ağlar, nedensel düğümlerin kümesi verildiğinde, koşullu bağımsızlık ilişkilerinin çıkarımını sağlar. Böylece, yön verilmiş döngüsel olmayan grafiklerde, koşullu bağımsızlıklar, çeşitli değişkenler arası ilişkileri yorumlamak için kolaylık sağlayan grafiksel bir gösterimdir.

Bayes ağlarda, koşullu bağımsızlığın incelenmesi için üç farklı yol ve bu yollar arasındaki ilişkiler ortaya konulmaya çalışılmıştır. Üç farklı Bayes ağ modeline yer verilmiş ve her bir model için üç yol ile koşullu bağımsızlıklar incelenmiştir.

Bayes ağlarda koşullu bağımsızlıklar, sayısal hesaplamalara gitmeksizin sadece ağdaki bağımsızlıkların maksimum sayısını tanımlar (Geiger vb. diğerleri, 1990). Koşullu bağımsızlığın incelenmesi için kullanılan yollar, okların yok olması ya da bulunmamasını ifade eder. İlk kullanılan yolda, ilgilenilen tüm değişkenler için, koşullu bağımsızlıkların bir listesini verir. Bu yöntem, maliyeti yüksek olan bir yöntemdir. Yönsel-ayırılma kriteri, nedensel ilişkilerin gösterimi olarak en iyi şekilde ifade edilmektedir. Yönsel-ayırılma kriterinin en iyi çözümü verdiği gözlenmiştir. Birleşme ağacı ile elde edilen koşullu bağımsızlıklar ile yönsel-ayırılma kriteri ile elde edilen koşullu bağımsızlıkların birbirine daha yakın olduğu gözlenmiştir (Oliver ve Smith, 1988). Ancak, her üç modelde de koşullu bağımsızlık özellikleri kullanılarak, elde edilen koşullu bağımsızlıkların aynı olduğu gösterilmiştir.

KAYNAKLAR

- BUNTINE, W., (1996), *A guide to the literature on learning probabilistic networks from data*, IEEE transactions on Knowledge and data Engineering, 8(2), 195-210.
- COWELL, R.G., (1999), *Introduction to inference in bayesian networks*, In Learning in Graphical Models, 9-26.
- EDWARDS, D., (1995), *Introduction to Inference Bayesian Networks*, Springer-Verlag, New York.
- FENTON, N., (1997), *Basics of BBNs*, [http://www.csr.city.ac.uk/people/norman.fenton/bbns/Details]. Erişim tarihi: 02.06.2000
- GEİGER, D., VERMA, T. AND PEARL, J., (1990), *Identifying independence in Bayesian networks*, Networks, Vol. 20, 507-534.
- JENSEN, F.V., OLESEN, K.G. AND ANDERSEN, S.K. (1990), *An algebra Bayesian belief universes for knowledge-based systems*, Networks, Vol. 20, 637-659.
- JENSEN, F.V., (1996), *An Introduction to Bayesian Networks*, UCL Press Ltd., London.
- LAURITZEN, S.L., (1996), *Graphical Models*, Oxford University Press, Oxford.

- LAURITZEN, S.L. AND SPIEGELHALTER, D.J., (1998), *Local computations with probabilities on graphical structures an their application to expert systems*, J.R. statist. Soc. B, 50(2), 157-224.
- LAURITZEN, S.L., DAWID, A.P., LARSEN, B.N., AND LEİMER, H.G., (1990), *Indepedence properties of directed Markov fields*, Networks, 20,491-505.
- LİAROKAPİS, D., (1999), *An introduction to belief networks*, [<http://www.cs.umb.edu/~dimitris>]. Erişim Tarihi: 02.11.2000.
- MADSEN, A.L.,AND JENSEN, F.V.,(1999), *LAZY propagation:A junction tree inference algorithm based on lazy evaluation*, Artificial Intelligence, 113,203-245.
- NEAL, R.M., (2000), *On deducing conditional independence from d-separation in causal graphs with feedback*, Journal of Artificial Intelligence Research, 12, 87-91.
- OLİVER, R.M. AND SMITH, J.Q., (1988), *Influence diagrams, Belief Nets and Decision Analysis*, John Wiley&Sons, New York.
- ORAL ERBAŞ, S., VE BAYRAK, H., (1999), *Grafiksel Modeller*, Gazi Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü, Ankara.
- PEARL, J., (1993), *From Bayesian Networks to Causal Networks*, 49th Session of the International Statistical Institue, Florence, Italy.
- PEARL, J., 2000, *Causality: Models, Reasoning and Inference*, Cambridge University Press, England.
- RİCHARDSON,T., (1997), *Introductin and d-separation*, UW department of statistics, [<http://www.stat.washington.edu/tsr/s538/lec1>].Erişim Tarihi: 4.11.2000
- STEPHENSON, T.A., (2000), *An introduction to Bayesian network theory and usage*, IDIAP Research Report 00-03.

A Study For Examining The Conditional Independence In Bayesian Networks

ABSTRACT

A Bayesian network is a directed acyclic graph which has conditional independence properties. Bayesian network consists of variables and the sets of the directed edge between variables. Edges denote probability dependence between variables. This dependence consists of the set of conditional probabilities. Conditional probability of a variable is determined by giving parents of each variable. When a node has no parent, a variable has an unconditional (marginal)

probability. In this study, conditional independence in Bayesian Networks is examined by following three ways. First of these ways is directed Markovian properties. Second is conditional independence which can be obtained by using moral and triangulated graph. Junction tree is constituted with cliques which can be obtained by using moral and triangulated graph. Conditional independence related with Bayesian network is obtained from junction tree. Third is the definition of the conditional independence by using d-separation criterion. The characteristics of conditional independence are given in three different ways and relations between these ways are examined.

Key Words: *Directed Acyclic Graphs, Directed Graphs, Undirected Graphs, Conditional Independence, Markovian Independence, Moral Graph, Triangulated Graph, Junction Tree.*

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 İngilizce dilde özeti araştırmanın sonunda verilmelidir. Araştırmanın İngilizce adı üstten 2 satır boşluk bırakılarak ortalı, her sözcüğün ilk harfi büyük, 14 punto harf büyüklüğünde, Abstract büyük harflerle ortalı, 12 punto harf büyüklüğünde koyu olarak yazılmalıdır. İngilizce özet soldan 5,5 cm ve sağdan 4,5 cm boşluk bırakılarak 200 kelimeyi geçmeyecek şekilde 11 punto harf büyüklüğünde italik olarak araştırmanın İngilizce adından sonra 3 satır boşluk bırakılarak yazılmalıdır.
- Anahtar kelimeler (Key words) her iki özeti bir satır altına, anahtar kelimeler ve key words koyu italik olarak yazılmalıdır.

9. Matematik simge ve formüllerin yazımında aşağıdaki hususlara dikkat edilir:

- Simgelerin ayırt edilmesi önemlidir. Özellikle büyük ve küçük harfler, düz ve koyu harfler, Klasik Yunan ve Latin harfleri, alt ve üst indisler, sıfır (0) rakamı ve O harfi, Bir (1) rakamı ve l harfi ayırt edilebilmelidir. Çoklu indislerden sakınılmalıdır.
- Denklemler word, standart (default) ölçülerde 1 tab (1,27 cm) içerden ve numara vermek gerekliyse bölüm numarasını içermeksizin en sağına parantez içinde yazılmalıdır. Uzun formüller metin içinde yer almamalıdır.
- Kesirler, metin içinde (/) işareti ile gösterilmelidir.
- Karmaşık ifadeler içeren denklemler olabildiğince kısaltma simgeleri kullanılarak yazılmalıdır.
- İç içe çoklu ayraçlar aynı formülde yer aldığı anda, sıra düzeni örneğin $\{[(0)]\}$ biçiminde olmalıdır.

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

Eksik Blok Düzenlerinin Dual Yapıları

The Dual Structure of Incomplete Block Designs

Hülya BAYRAK.....1

Çok Değişkenli Normal Dağılıma Sahip Örneklerdeki Aykırı Gözlemlerin Belirlenmesi için Bayesgil Bir Yaklaşım

A Bayesian Method to Identification of Outlier Observations in Multivariate Normal Distribution

Ufuk EKİZ.....11

Damar Darlık Derecesi ve Risk Faktörlerinin Homojenite Analizi ile İncelenmesi

Investigation of the Effect of Risk Factors on Narrowness Degree of Blood Vessel by Using Homogeneity Analysis

Necdet SÜT
Mevlüt TÜRE.....21

Yapay Sinir Ağı Performansına Etki Eden Faktörlerin Analizinde Taguchi Yöntemi: Hisse Senedi Fiyat Tahmini Uygulaması

Analyzing Performance of Artificial Neural Networks by Taguchi Methods: Forecasting Stock Market Prices

Alperen ÖZALP
A. Sermet ANAGÜN.....29

İki ve Üç Yönlü Tabloların Gözelerinde 5'den Küçük Beklenen Frekans Olması Durumunda, I. Tip Hata Olasılığının (α) Durumu

The Case of Cells of Two and Three Dimensional Tables Having Less Than 5 Frequencies, The Case of First Type Error (α) Probabilities

Mehmet AKYOL
Fikret GÜRBÜZ.....47

Yaşam Çözümlemesinde Yarışan Riskler ve Bir Uygulama

Competing Risks in the Survival Analysis and an Application

Durdu SERTKAYA
M. Tekin SÖZER.....63

Sürekli Değişken İçeren Grafikselle Modeller

Graphical Models for Continuous Variables

Hülya BAYRAK
Fikri GÖKPINAR
Berrin ÖZKAYA.....79

Bayes Ağlarda Koşullu Bağımsızlıkların İncelenmesi Üzerine Bir Çalışma

A Study for Examining the Conditional Independence in Bayesian Networks

Hülya OLMUŞ
Semra Oral ERBAŞ.....89