

**DİSKRİMİNANT ANALİZİNDE AYIRMA 'NIN GÜCÜ İLE
FAKTÖR ANALİZİNDE KULLANILAN KOVARYANS
MATRİSİ ARASINDAKİ İLİŞKİ
ÜZERİNE BİR ÇALIŞMA**

Ahmet KAYA(*)

ÖZET

Bu çalışma çok değişkenli istatistik analizde birimlerin sınıflandırılması ve boyut indirgeme amacına yönelik olarak ayırma analizinde(discriminant analysis), bir veri seti kullanılarak, gruplar içi homojenliği en küçük, gruplararası heterojenliği en büyük yapacak ayırma işlemi ve etkinliğini, bu etkinliğin faktör analizinde(factor analysis), faktörler elde etmek amacıyla kullanılan kovaryans matrisi ile olan ilişkisini incelemek amacıyla yapılmıştır.

1. GİRİŞ

Uygulamalı bilimlerde, bilimsel değerlendirmelere temel oluşturan veriler deney birimlerinin özelliklerini gözlemekle oluşur. Deney birimleri; insan, hayvan, işletme gibi canlı ve cansız varlıklar olabilir. Ölçümü yapılan özellikler ise bilimsel incelemenin konusunu oluşturan değişkenlerdir. Çok değişkenli teori, regresyon, korelasyon ve varyans analizi yöntemlerinin genellenmesine olanak sağlamaktadır. Ancak özellikle hipotez testlerinin önem taşıdığı durumlarda standart uygulamalar pek yararlı olamamaktadır. Daha ayrıntılı çözümlere gidildiğinde testlerin gücü azalmakta ve güven aralıkları genişlemektedir. Çok değişkenli analizde bazı yöntemler belli bir olasılık teorisine dayanmamaktadır. Böylece pratik ve kullanışlı yöntemler temelde istatistiğin genel çerçevesinden uzaklaşmaktadır. Ancak, sebep-sonuç ilişkisinde bir olayı anlatmak için birden fazla değişkene gereksinim duyulabilir. Bazı uygulamalarda ise, kurulan modeller, gereksiz bazı değişkenleri de içerebilir. Bu durum, gereğinden çok işlemin yapılmasına ve zaman kaybına neden olurken, daha yalın modellerle ifade edilmesi mümkün olan problemlerin karmaşık bir yapı kazanmasına yol açmaktadır (Okur, 1982).

Bundan hareketle ayırma analizi ve faktör analizi kavramlarını tanımlamaya çalışalım. Birimleri veya bireyleri enaz hata ile ait oldukları kitlelere ayırmak için yapılan işlemlere "ayırma analizi" denir.

Ayırma analizi için p değişkenden bulunacak bağıntı,

(*) Öğr. Gör. Ege Üniversitesi Tire Meslek Yüksekokulu, Tire-İZMİR

$$Y_i = V_1 X_{i1} + V_2 X_{i2} + \dots + V_p X_{ip} \quad (1)$$

biçimindedir.

Burada X_1, X_2, \dots, X_p değişkenleri, V_1, V_2, \dots, V_p ise değişkenlere ilişkin katsayıları göstermektedir. Bu fonksiyona "ayırma fonksiyonu" denir.

Böyle bir fonksiyon bulunurken gruplararası değişimin grupiçi değişime oranının enbüyük olması gerekir.

Yani;

$$F = \text{Max}(\text{Gruplararası Değişim} / \text{Grupiçi Değişim})$$

Fisher tarafından tanımlanan iki varyans oranı ;

$$L = \frac{v'Av}{v'Wv} \Big|_{\max} \text{ olmalıdır.} \quad (2)$$

Buradan elde edilecek L_i özdeğerlerine karşılık gelen özvektörler belirtilen koşulları sağlayan ayırıcı fonksiyonlardır. Bunun sonucu olarak (2) nolu eşitliğin v 'ye göre türevi alınıp gerekli düzenlemeler yapılırsa, L_i özdeğerlerinin,

$|W^{-1}A - LI| = 0$ determinantının çözümünden elde edildiği görülebilir.

$r = \min(k - 1, p)$ olmak üzere elde edilen L_1, L_2, \dots, L_r özdeğerlerine karşılık gelen r tane özvektör, ayırıcı fonksiyon işlevi görür.

İlişkili değişkenlerden oluşan bir sistemde, değişkenler arasındaki ilişkileri yakından incelemeyi amaçlayan yönteme "faktör analizi" denir.

Faktör analizi için p değişkenden oluşan bağıntı ,

$$Z_i(1) = V_{i1} X_{i1} + \dots + V_{ip} X_{ip} + b_i u_i \quad (3)$$

biçimindedir.

Burada X_1, X_2, \dots, X_p deęişkenleri, V_1, V_2, \dots, V_p ise deęişkenlere ilişkin katsayıları göstermektedir.

(2) no' lu modele ilişkin varsayımlar aőaęıdaki gibidir :

1. $E(\underline{X}) = \underline{0}$
2. $Var(\underline{X}) = \underline{0}$
3. $Kov(u_i, u_j) = 0$
4. $Kov(\underline{X}, \underline{u}) = \underline{0}$
5. $E(\underline{u}) = \underline{0}$
6. $u_i \approx N(0, \sigma^2_{u_i})$

2. UYGULAMA

Bir fabrikada alıőan iőiler birinci iői a, son iői l biiminde kodlanmak suretiyle yıl ii devamsızlık durumları, davranıő biimi ve aynı iőyerinde alıőma sureleri ile aıklanmak istenmektedir (*Green, 1972*).

Tablo-1, İői Bilgileri Tablosu

İőiler	Devamsız Gun	Davranıő Biimi	alıőma Suresi
a	1	1	1
b	0	2	1
c	1	2	2
d	4	3	2
e	3	5	4
f	2	5	6
g	5	6	9
h	6	7	4
i	9	10	8
j	13	11	7
k	15	11	9
l	16	12	10

Yukarıdaki verilerle iőilerin devamsızlık durumları, davranıő biimi ve alıőma suresi ile ne kadar aıklanabilmektedir sorusuna cevap bulunması gerekmektedir. Bu amala, davranıő biimi ve alıőma sureleri baęımsız deęişkenler kabul edilerek iőilerin devamsızlık nedenleri tespit edilmeye alıőılacaktır.

Diskriminant ve Faktör Analizi

Tablo-2 Regresyon Katsayılar Tablosu (RKT)

<u>Değişkenler</u>	<u>Katsayılar</u>	<u>St. Sapma</u>	<u>t-oranı</u>	<u>p</u>
Sabit	-2.183	1.097	-1.99	0.078
Dav. Biçimi	1.5377	0.3130	4.91	0.000
Çal. Süresi	-0.2243	0.3694	-0.61	0.559
R-kare = % 90.6				

Tablo-3, Varyans Analizi Tablosu (VAT)

<u>Varyans Analizi Sonuçları</u>				
<u>Kaynak</u>	<u>SD*</u>	<u>KT**</u>	<u>KO***</u>	<u>F</u>
Regresyon	2	320.93	160.46	43.43
Hata	9	33.32	3.70	
Toplam	11	354.25		

* Serbestlik Derecesi,

**Kareler Toplamı,

***Kareler Ortalaması

Bağımsız değişkenlere ilişkin Tablo-2' de verilen katsayıların önem seviyelerine bakıldığında davranış biçiminin önemli, çalışma süresi değişkeninin ise önemsiz çıktığı görülmektedir. Bu sonuçlardan hareketle işçileri önemli olan davranış biçimi değişkenine göre sıralamak gerekmektedir. Verilere ilişkin, grupları içi sapmalar ve gruplararası sapmalar matrisleri, davranış biçimi değişkeni esas alınarak aşağıdaki gibi elde edilmiştir.

Tablo-4 Gruplandırma Tablosu

<u>İşçi</u>	<u>Devamsız Gün</u>	<u>Davranış Biçimi</u>	<u>Çalışma Süresi</u>
a	1	1	1
b	0	2	1
c	1	2	2
d	4	3	2
e	3	5	4
f	2	5	6
g	5	6	9
h	6	7	4
i	9	10	8
j	13	11	7
k	15	11	9
l	16	12	10

Tablo-5 Gruplar İçin Sapmalar Tablosu

İşçi	Davranış Biçimi		Çalışma Süresi	
	Gruplariçi Sapmalar İşlemi		Gruplararası Sapmalar İşlemi	
a	-1	-0.5	-4.25	-3.42
b	0	-0.5	-4.25	-3.42
c	0	0.5	-4.25	-3.42
d	1	0.5	-4.25	-3.42
e	-0.75	-0.75	-0.5	-0.17
f	-0.75	1.25	-0.5	-0.17
g	0.25	0.25	-0.5	-0.17
h	1.25	-0.75	-0.5	-0.17
i	-1	-0.5	4.75	3.58
j	0	-1.5	4.75	3.58
k	0	0.5	4.75	3.58
l	1	2.5	4.75	3.58

Davranış biçimleri değerlerine ilişkin değişimler gözönünde bulundurularak işçiler üç grupta toplanabilir. Buna göre a, b, c ve d işçileri 1. grupta, e, f, g ve h işçileri 2. grupta, i, j, k ve l işçileri ise 3. grupta toplanmıştır. Bu gruplama işleminde her grubun kendi içinde minimum varyansa sahip olacak şekilde gruplara ayrıldıkları görülebilir. Yukarıdaki ayırma işlemine bağlı olarak elde edilen, gruplariçi sapmalar ve gruplararası sapmalar matrisleri aşağıdaki gibi elde edilir.

$$W = \begin{bmatrix} 6.75 & 1.75 \\ 1.75 & 8.75 \end{bmatrix} \quad A = \begin{bmatrix} 163.50 & 126.50 \\ 126.50 & 98.17 \end{bmatrix}$$

Gruplariçi ve gruplararası sapmalar matrislerinden hareketle L_i özdeğerine ilişkin özvektör elde edildiğinde, $L_1=29.44$ ilişkin özvektör değerleri 0.905 ve 0.425 olarak, $L_2=0.0295$ özdeğerine ilişkin özvektör elemanları ise, -0.612 ve 0.791 biçiminde elde edilmiştir. Görüldüğü gibi ikinci özdeğer birinci özdeğere göre çok küçük çıkmakta dolayısıyla, birinci özdeğere ilişkin elde edilen ayırma fonksiyonu yeterli olmaktadır.

Khi-kare önem kontrolü ile ayırmanın gerekli görüldüğü bu yöntemde birinci özdeğere karşı karşı gelen özvektör değerleri 29.035 ve 0.247 olarak elde edilmiştir. Bu şekilde elde edilen ayırma gücü % 99.9 gibi bir orana yaklaşmaktadır.

Birinci gözlem X_1 'e ilişkin ayırma skoru,

$$w_{1(1)} = 0.905(-5.25) + 0.425(-3.92) = -6.42$$

elde edilmiştir.

Grup ortalamaları için ayırma skorları ise,

$$w_1(\text{Grup 1}) = 0.905(-4.25) + 0.425(-3.42) = -5.30$$

$$w_2(\text{Grup 2}) = 0.905(-0.50) + 0.425(-0.17) = -0.52$$

$$w_3(\text{Grup 3}) = 0.905(4.75) + 0.425(3.58) = 5.82$$

Yukarıdan görülebileceği gibi ayırma skorlar toplamı sıfırdır.

Şimdi aynı verilerle faktör analizi için kovaryans matrisi elde edilirse,

$$C = \begin{bmatrix} 14.19 & 10.69 \\ 10.69 & 8.91 \end{bmatrix}$$

Yukarıdaki kovaryans matristen elde edilen özdeğerler, $L_1 = 22.56$, $L_2 = 0.54$ biçimindedir.

Özdeğerlere karşı gelen özvektörler,

$$v_1 = \begin{bmatrix} 0.787 \\ 0.617 \end{bmatrix} \quad v_2 = \begin{bmatrix} 0.617 \\ -0.787 \end{bmatrix}$$

olarak elde edilmiştir. Özdeğerlerden oluşan matris ,

Ahmet Kaya

$$T = \begin{bmatrix} 0.787 & 0.617 \\ 0.617 & -0.787 \end{bmatrix}$$

şeklinde elde edilmiştir. Kovaryans matrisinden bulunmuş bileşen yükleri matrisi ise aşağıdaki gibi elde edilmiştir.

$$F = \begin{bmatrix} 0.787 & 0.617 \\ 0.617 & -0.787 \end{bmatrix} \begin{bmatrix} 4.75 & 0 \\ 0 & 0.73 \end{bmatrix} = \begin{bmatrix} 3.74 & 0.45 \\ 2.93 & 0.57 \end{bmatrix}$$

$$L_1 = (3.74)^2 + (2.93)^2 = 22.56, \quad L_2 = (0.45)^2 + (0.57)^2 = 0.54.$$

Görüldüğü gibi birinci bileşenin varyansı, 22.56, ikinci bileşenin varyansı ise 0.54 olarak elde edilmiştir. Birinci bileşenin değişimi açıklama yüzdesi, $22.56/(22.56+0.54) = 0.98$ 'tir. Bu durumda ikinci bileşenin değişimi açıklama % 2 düzeyinde kalmaktadır. Dolayısıyla ikinci bağımsız değişkenin kullanımı önemli bulunmamıştır, bu sonuç varyans analizi sonuçlarında da elde edilmişti.

3. SONUÇ

Analiz sonuçları dikkatle incelenecek olursa, tek bağımlı iki bağımsız değişkene ilişkin olarak yapılmış çalışmada basitleştirme ve boyut indirgeme, birimleri sınıflama, bağımlılık yapısının incelenmesine yönelik elde edilen sonuçlarda tam bir tutarlılık görülmektedir. Bundan hareketle ayırma analizi yapmak amacıyla elde edilen gruplarıçi sapmalar ve gruplararası sapmalar matrisi kullanılarak elde edilen özdeğerler ve bunlara karşı gelen özvektörler kullanılarak ayırma fonksiyonları elde edilmiştir. Elde edilen bu fonksiyonlar ile yapılan ayırma işleminin ne kadar etkin olduğu ortaya konmuştur. Ayırma analizi ayrıca normallik ve kovaryans homojenliği altında olasılık ve istatistik yöntemlerin uygulanmasına imkan vermektedir. Ayrıca bu sonuçlar, gruplarıçi homojenliğin ve gruplararası heterojenliğin ortaya konması bakımından da bir ölçüt olabilmektedir. Faktörlerin elde edilmesinde kullanılan kovaryans matrisi ise gözlemlerarası bağımlılık yapısını ortaya koyma açısından bir ölçüt olarak kullanılmaktadır. Faktör analizi uygulamasına başlamadan önce değişkenler arasındaki korelasyonları incelemek yararlıdır. Eğer bu korelasyonlar çok düşük ise, analizi yapmanın bir yararı olmaz. Ayrıca korelasyonlar bakımından değişkenlerin kümeleme eğilimleri hangi tür faktörlerin ortaya çıkabileceği

konusunda fikir verebilir. Ancak unutulmaması gerekli nokta, faktörlerin gözlenemez olmaları ve araştırmacı tarafından modele sokulmuş bulunmalarıdır.

ABSTRACT

In the present study the problem of the classification of units and discrimination by reducing the dimension with regard to the minimization of within group variation and the maximisation of among group heterogeneity was considered. The efficiency of this process was later related with the covariance structure of the factor analysis. The result of this relation was studied on a numerical example considering data obtained on the workers in a factory.

TEŞEKKÜR

Bu çalışmanın sonuçlandırılmasında, destek, öneri ve yardımları ile katkı yapan, Ege Üniversitesi Bilgisayar Mühendisliği Bölümü, Bilgisayar Bilimleri Ana Bilim Dalı Başkanı Prof. Dr. Fikret İKİZ'e teşekkürü borç bilir, saygılar sunarım.

KAYNAKÇA

- Green, E. P, (1972); *Mathematical Tools For Applied Multivariate Analysis*, Wharton School University of Pennsylvania.
- Morrison, D.F, (1972); *Multivariate Statistical Methods*, McGraw-Hill International Company.
- Okur, C., (1982); *Çok Değişkenli İstatistik Analizi Ders Notları*, Ege Üniversitesi Bilgisayar Bilimleri Mühendisliği Bölümü, Bornova/İzmir.
- Rummel, R J, (1979); *Applied Factor Analysis*, Northwestern University press.
- Saraç, İ., (1980); Faktör, *Diskriminant ve Sınıflandırma Analizlerinin Birlikte Kullanımlarına İlişkin Bir Çalışma*, Ege Üniv. Elektronik ve Hesap Bil. Enst. Basılmamış Yüksek Lisans Tezi, Bornova/İzmir.
- Tatlıdil, H. (1997); *Uygulamalı Çok Değişkenli İstatistiksel Analiz*, Akademi Matbaası, Kızılay/Ankara.