

Diskriminant Analiz ve Eğitimde Bir Uygulama

Murat ALTUN*

ÖZET

Bu çalışmada çok değişkenli istatistiksel analiz türlerinden biri olan diskriminant analiz ve dayandığı matematik temeller kısaca tanıtılmış ve basic programlama dilinde bir program geliştirilmiştir.

Sonra bu program kullanılarak lise birinci sınıf öğrencilerinin ikinci sınıfta akademik kol seçimleri analiz edilmiştir. Programın dökümünden kısaltılmış örnekler ekler bölümünde tablolar halinde sunulmuştur.

SUMMARY

Discriminant Analysis and an Application in Education

In this study, at first the discriminant analysis and its mathematical ground works have been explained. Then, a basic computer program has been prepared. After that this program has been applied for the discrimination of the lycee students in academic branches.

TEORİK BİLGİ

Diskriminant Analizinin Konusu ve Amaçları

Diskriminant analizinin konusu hangi populasyona ait oldukları bilinmeyen bireylerin ayırımı ve sınıflandırılmasıdır.

Populasyonu oluşturan bireyler belirli özellikleri bakımından değişik gruplar oluştururlar. Bu grupların farklı olup olmadığı çok değişkenli varyans analizinin konusudur. Farklı olduklarına karar verildiği takdirde bu farklılığın kö-

* U.Ü. Eğitim Yüksekokulu Öğretim Görevlisi.

kenini arařtırmak gündeme gelir ki bu diskriminant analizinin konusudur. Farklılıđın k keni analize esas olan p tane deđiřken yerine bunların dođrusal ya da eđri-sel bir kombinezonu olan yeni deđiřkenleri (diskriminant fonksiyonlarını) tanım-lamak ve b ylece gruplar arasındaki mesafeyi maksimum yapmak suretiyle arařtırılır.

Çođunlukla grupları ayırmak iin diskriminant fonksiyonlarının bir, iki veya  tanesi yeterlidir. B ylece bireyler daha az boyutlu bir sistemde incelene-bilmiř olmaktadır. Bunun yanında gruplardan birine ait olduđu bilinen ancak hangisine ait olduđu bilinmeyen bir bireyin ait olduđu grubun belirlenmesi de dis-kriminant fonksiyonlarına bađlı olarak geliřtirilen kurallarla m mkündür.

Diskriminant Fonksiyonlarının Elde Ediliři

Daha  nce belirtildiđi gibi diskriminant fonksiyonları orijinal deđiřkenlerin (X lerin) birer lineer kombinezonudur. Ancak X_1 lerin her lineer kombinezonu diskriminant fonksiyonu deđildir. Diskriminant fonksiyonu olabilmesi iin grupla-rı birbirinden en iyi řekilde ayırması gerekir. Gruplararası farklılıđı ortaya koya-bilecek fonksiyonlar "Gruplararası arpımlar ve Kareler Toplamı"nın, "Gruplar İi arpımlar ve Kareler Toplamı"na oranına (F') bađlı olarak elde edilirler.

Birden ok deđiřkenin bulunduđu durumlarda "Gruplararası arpımlar ve Kareler Toplamı Matrisi (B)", "Gruplar İi Kareler Toplamı Matrisi (W)" ve C_i katsayılar ( zvekt r katsayıları) olmak  zere

$$F = \frac{C' B C}{C' W C} = \lambda$$

řeklinde elde edilir.

$$\frac{\partial \lambda}{\partial C} = 0$$

Buradan $BC - WC \cdot \lambda = 0$ elde edilir. Bu eřitliđin her iki yanını W ile arpılarak oluřturulan $W^{-1} B - \lambda I = 0$ denkleminde uygun C_i katsayıları bulun-mak suretiyle diskriminant fonksiyonları elde edilir.

Karar Amalı Diskriminant Analiz

Bir bireyin herhangi bir gruba yakınlıđını (ya da benzerliđini) belirleyen karar kuralları iki grupta toplanabilir.

- Geometrik Esaslı Karar Kuralları,
- Probabilistik Esaslı Karar Kuralları.

Ařađıda uygulamada geometrik esaslı karar kurallarından "Diskriminant Fonksiyonları İle Sınıflandırma" y ntemi kullanılmıřtır.

Bu y ntemle sınıflamada bir X_i vekt rünün grup ortalama vekt rlerinden uzaklıđı (mahalonobis uzaklıđı) diskriminant fonksiyonları  zerinden hesaplanır.

$$f_i(X) = C_1X_1 + C_2X_2 + \dots + C_pX_p$$

gibi r tane ayırıcı faktör (Diskriminant Fonksiyonu) ün elde edildiğini varsayalım. Bir X_i bireyinin bu diskriminant fonksiyonu (ayırıcı faktör eksen) üzerindeki izdüşümü (diskriminant skoru) bulunur. Bunlar $f_1(X_i), f_2(X_i), \dots, f_r(X_i)$ olsunlar.

Grup ortalama vektörlerinin diskriminant skorları da benzer şekilde hesaplanır. Bunlar m tane grup için

$$f_1(\bar{X}_1), f_2(\bar{X}_1), \dots, f_r(\bar{X}_1)$$

$$f_1(\bar{X}_2), f_2(\bar{X}_2), \dots, f_r(\bar{X}_2)$$

$$\dots$$

$$f_1(\bar{X}_m), f_2(\bar{X}_m), \dots, f_r(\bar{X}_m)$$

olsunlar.

Şimdi X_i vektörünün k'ncü grubun ortalama vektöründen uzaklığının karesi her bir eksen üzerinden analitik olarak

$$D^2_{ik} = (f_1(X_i) - f_1(\bar{X}_k))^2 + (f_2(X_i) - f_2(\bar{X}_k))^2 + \dots + (f_r(X_i) - f_r(\bar{X}_k))^2$$

biçiminde tanımlıdır. Bir vektörün her bir eksene uzaklığı hesaplandıktan sonra $\text{Min}(D^2_{i1}, D^2_{i2}, \dots, D^2_{im})$ seçilir ve birey bu gruba atanır.

UYGULAMA

Diskriminant analizin çeşitli konularda birçok uygulaması mevcuttur. Özellikle tarım, tıp, psikoloji, sosyoloji ve biyolojide ayırım ve sınıflandırma işlemlerinde yaygın şekilde kullanılmaktadır. Burada uygulama alanı eğitim seçilmiş ve lise öğrencilerinin akademik kollara (fen kolu, edebiyat kolu gibi) ayrılması üzerinde çalışılmıştır.

Problemin Ortaya Çıkışı

Hali hazırda üniversiteye öğrenci hazırlamayı temel amaç edinen genel liselerde (meslek liseleri hariç) öğrencilerin tümü birinci sınıfta aynı dersleri okumakta; ikinci sınıfta iki kola (gruba) ayrılarak farklı dersler okumaktadırlar. Fen kolu öğrencileri de lise üçüncü sınıfta kendi aralarında tekrar iki gruba (matematik ve tabii bilimler kolları) ayrılmaktadırlar.

Kollara ayrılmada temel amaç, öğrencilerin yeteneklerine uygun ortamlarda eğitim almaları ve buna bağlı olarak yeteneklerine uygun iş edinmelerini temin etmektir.

Lise eğitiminin nasıl olması gerektiği bu tartışmanın dışındadır. Mevcut eğitim sisteminde öğrenciler iki seçenek (fen ve edebiyat) karşısında bırakılmaktadır.

Burada öğrencilerin buna göre sınıflandırılmasının ne denli isabetli olup olmadığı incelenmektedir.

Öğrencilerle İlgili Bilgiler

Öğrencilere lise birinci sınıfta

X1 : Türk Dili ve Edebiyatı	X7 : Fizik
X2 : Tarih	X8 : Kimya
X3 : T.C. İnkılap Tarihi	X9 : Yabancı Dil
X4 : Coğrafya	X10 : Beden Eğitimi
X5 : Matematik	X11 : Din Kültürü ve Ahlâk Bilgisi
X6 : Biyoloji	

ve bunlardan başka liseler arasında farklılık gösteren ikişer seçmeli ders okutulmaktadır.

Analiz Bursa Yıldırım, Erkek ve Çınar Liselerinin 1986-1987 öğretim yılında lise birinci sınıf okumuş toplam 120 öğrencisi üzerinde yapılmıştır. Analiz öncesinde öğrenciler lise ikinci sınıfta ayrıldıkları kollar esas alınarak iki gruba ayrılmış ve yukarıda belirtilen onbir dersten aldıkları notlar tesbit edilmiştir (Tablo: 1).

Öğrencilerin liselere dağılımı şöyledir:

	Fen Kolu	Edebiyat Kolu	Toplam
Yıldırım Lisesi	21	21	42
Erkek Lisesi	21	21	42
Çınar Lisesi	21	17	36
TOPLAM	63	57	120

Tablodan da anlaşılacağı gibi 63 öğrenci fen grubunu, 57 öğrenci edebiyat grubunu oluşturmuştur.

Diskriminant Fonksiyonları ve Diskriminant Skorları

Öğrenci notları X_1, X_2, \dots, X_{11} sırasıyla veri matrisinin kısaltılmış bir biçimi (Tablo: I) ile verilmiştir. Bu matriste yer alan 5'in altındaki notlar öğrencilerin öğretmenler kurulu kararıyla veya borçlu olarak sınıf geçtikleri derslere aittir.

Genel ortalama vektörü ve değişkenlerin standart sapmaları (Tablo: II) incelendiğinde 5.22 ortalama ile matematik en başarısız ders olmuş, bunu 5.35 ile fizik izlemiştir. En başarılı ders 7.95 ortalama ile Beden Eğitimi olmuş, bunu 7.45 ile Din Kültürü ve Ahlak Bilgisi izlemiştir.

En büyük iki standart sapma matematik ve kimya derslerine ait olup, 1.93 ve 1.75'tir. Bu durum bu iki dersin notlarının diğer derslere göre daha çok değişkenlik gösterdiğini ortaya koymaktadır. En az sapma gösteren Beden Eğitimi dersi olmuştur (1.07).

Basic programlama dilinde geliştirilen bilgisayar programı aracılığı ile Genel Çarpımlar ve Kareler Toplamı Matrisi (T), Gruplar İçi Çarpımlar ve Kareler Toplamı Matrisi (W), Gruplararası Çarpımlar ve Kareler Toplamı Matrisi (B) hesaplanmış ve buradan özdeğerlere esas olan $W^{-1} B$ matrisi elde edilerek birinci ve ikinci özdeğerler bulunmuştur.

$$\lambda_1 = 0.679617$$

$$\lambda_2 = 0.000018 \text{ çıkmıştır.}$$

Özdeğerlerin önemli olup olmadığını anlamak için i'inci kök için tanımlanan χ^2 değişkenini λ_1 için hesaplayalım.

$$\chi^2 = (n - \frac{p + m}{2}) \ln(1 + \lambda_i)$$

$$\chi^2 = (120 - \frac{11 + 2}{2}) \ln(1 + 0.6796) = 65,598$$

Bu değişken $p + m - 2$ serbestlik dereceli χ^2 dağılışı gösterir. $sd = 11$ olup $\alpha = 0.05$ önem seviyesinde $\chi^2_{11} = 15.68$ 'dir. $X^2_{11} < X^2$ olduğundan birinci kökün temsil ettiği varyasyon önemlidir. Buna göre birinci diskriminant fonksiyonu $f_1(X) = -0.0006X_1 - 0.2479X_2 - 0.0109X_3 - 0.0969X_4 + 0.2466X_5 + 0.251X_6 + 0.1674X_7 + 0.3098X_8 - 0.0444X_9 - 0.2464X_{10} + 0.0397X_{11}$.

Birinci köke karşılık elde edilen diskriminant fonksiyonu grupları ayırma da tek başına yeterli olduğundan ikinci diskriminant fonksiyonuna ihtiyaç yoktur.

Tablo III ile öğrencilerin f_1 ve f_2 diskriminant fonksiyonları üzerinden aldıkları değerler (diskriminant değerleri) verilmiştir.

Birinci diskriminant fonksiyonu üzerinden elde edilen skorlar bakımından ilk üç sırayı 46, 3, 4 nolu, son üç sırayı 78, 76, 77 nolu öğrenciler almışlardır.

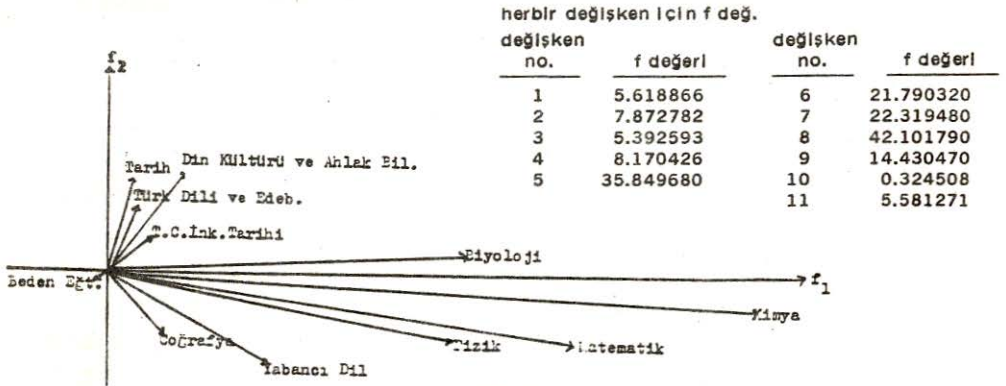
Bu faktör bakımından pozitif skora sahip bireyler faktörün temsil ettiği yetenekler bakımından iyi durumdadır.

Diskriminant Fonksiyonunun Tanınması ve Adlandırılması

Birinci diskriminant fonksiyonunun oluşumuna en çok katkıda bulunan değişkenler sırayla kimya, matematik, biyoloji, fizik derslerinin notlarıdır. Bu bakımdan bu faktöre "Matematik ve Fen Yeteneği" adı verilebilir. λ_1 özdeğerinin ayırım gücü 0.999 olduğu için bu faktör grupları ayırmaya yeterlidir.

Fen ve Edebiyat gruplarının koordinat sistemindeki yerlerini gösteren grafikte derslerle ilgili vektörlerin başlangıç noktası, grupların merkez koordinatlarının ortalamaları seçilmiştir. Her bir değişkenin f_1 ve f_2 vektörleri ile korelasyonlarının belirlediği nokta başlangıca birleştirilmiş ve vektörler ilgili buldukları F değerleri ile çarpılarak uzunlukları elde edilmiştir. Aşağıdaki grafikte matematik, biyoloji, fizik, kimyanın birinci diskriminant fonksiyonuna yönelmiş

olması bu gruplarda anılan derslerin ortalamalarının, uzunlukları ise ayırım güçlerinin yüksek olduğunu ortaya koymaktadır.



Şekil: 1
Değişkenlerin Diskriminant Fonksiyonları İle İlişkileri

Bireylerin Sınıflandırılması

Bireylerin hangi gruba daha yakın olduklarını belirlemek için "diskriminant fonksiyonları ile sınıflandırma" yönteminden yararlanılmıştır. Yapılan sınıflandırma sonucunda fen kolundaki 63 öğrenciden 16'sı (% 25) edebiyat koluna, edebiyat kolundaki 57 öğrencinin 10'u (% 18) fen koluna yakın çıkmışlardır (Tablo: IV).

TARTIŞMA VE SONUÇ

Diskriminant analizi ile ulaşılan yargıların güvenilirliği örneğe alınan birey sayısının popülasyonu temsil edebilecek kadar çok olması, analiz öncesindeki gruplamanın isabetliliği, ölçümlerin sağlıklılığı ve seçilen değişkenlerin ayırıcı olabilmeleriyle yakından ilgilidir.

Eğitimle ilgili bu uygulamadan çıkan sonuçlar şöyle sıralanabilir:

1- Öğrencilerin lise ikinci sınıfta fen ve edebiyat kollarına ayrılmaları tamamen kimya, matematik, fizik, biyoloji derslerinden etkilenmektedir. Edebiyat kolunun oluşumuna Türk Dili ve Edebiyatı, Tarih, Coğrafya gibi derslerin kayda değer bir etkisinin olmadığı görülmüştür.

Bir başka söyleyişle öğrenciler fen kolunu seçemiyor iseler, zorunlu olarak edebiyat kolunu seçmektedirler.

2- Mevcut kol seçimi % 25, % 18 gibi önemli miktarlarda yanılmalara (yanlış sınıflamalara) yol açmaktadır. Bu durum kol seçiminin tamamen derslerle ilgili olması gerekirken, dersler dışında başka faktörlerden de etkilendiği şüphesini doğurmaktadır.

Bu yüzden kol seçimlerinde, daha geniş örnekler üzerinde yapılmış çalışmalarından elde edilen diskrimine edici fonksiyonların kullanılması daha az hatalı sınıflamaları ortaya koyabilir.

Kol seçimleri üzerinde dersler dışındaki diğer faktörler de (aile tutumu, üniversite sınavları gibi) ayrıca incelenmelidir.

3- Lise ikinci, üçüncü ve daha sonraki yıllarda fen ve edebiyat alanlarında başarılı olmuş kimselerden oluşturulacak grupların notları üzerinden geliştirilebilecek diskriminant fonksiyonları ile sınıflama (kol seçme) daha isabetli olabilir. Bu durumda her iki kola uzak bulunan öğrenciler bu iki kolun dışında bir eğitim programına veya iş koluna yöneltilir.

KAYNAKÇA

1. EMİN, M.S.: *Çok Boyutlu Verilerin Bazı İstatistiksel Analiz Yöntemleri ve Uygulamaları*, "Yayımlanmamış Doktora Tezi", Gazi Üniversitesi Fen Bil. Ens., 1984.
2. ÖZTÜRK, Aydın, KARATAŞ, Şaban: *Diskriminant Analizi ve Bununla İlgili Bir Uygulama*, Atatürk Üniversitesi Ziraat Fak. Ziraat Dergisi, Cilt 6, Sayı 2, Sevinç Matbaası, Ankara, 1975, s. 251-263.
3. ÖZTÜRK, A. ve Diğerleri: *Çoklu Diskriminant Analizi ve Bununla İlgili Bir Uygulama*, Ege Üniversitesi Elektronik Hesap Bilimleri Enstitüsü Dergisi, Cilt 1, Sayı 1, Ege Üniv. Matbaası, İzmir, 1978, s. 31-41.
4. KORUM, U.: *Matematiksel İstatistiğe Giriş*, Ankara Üniv. Siyasal Bilgiler Fak. Yayını, Ankara, 1971.
5. MORRISON, F.D.: *Multivariate Statistical Methods*, Second Edition, Mc Graw Hill Kogakuska LT. D., New-York, 1976.

E K L E R :

Tablo: 1
Diskriminant Analizi Veri Matrisi

1	8	6	7	6	8	6	8	7	10	8	8
2	7	9	9	8	5	7	6	8	6	8	9
3	9	9	10	9	9	10	10	10	10	9	10
4	8	9	10	8	8	9	8	10	8	8	10
5	6	6	6	6	8	6	3	5	5	7	7
6	6	7	6	5	5	5	3	6	6	8	6
7	5	5	7	5	7	6	5	5	6	8	6
8	7	8	8	8	8	7	7	9	6	10	8
9	6	8	9	6	6	7	6	9	7	9	8
10	7	9	10	8	6	8	7	9	6	7	8
-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-

64	6	8	7	7	5	7	5	8	5	7	7
65	7	6	6	6	3	6	1	6	6	8	8
66	6	6	7	6	3	5	2	5	7	8	7
67	6	6	6	5	5	6	1	5	5	8	7
68	6	6	8	5	2	5	5	5	5	9	7
69	7	8	7	6	6	5	5	9	7	8	8
70	7	6	7	5	2	6	1	7	5	8	7
71	6	9	9	7	5	6	5	5	6	8	9
72	6	8	7	5	5	5	3	5	5	7	6
73	6	7	8	8	3	6	3	5	6	8	9
74	6	7	9	8	3	6	7	5	5	8	8
-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-

Tablo: 2
Değişkenlerin Standart Sapmaları

1 inci deę. stan. sap:	1.212418	ortalaması	6.025
2 inci deę. stan. sap:	1.413816	ortalaması	6.966667
3 inci deę. stan. sap:	1.480948	ortalaması	7.341667
4 inci deę. stan. sap:	1.438858	ortalaması	6.383333
5 inci deę. stan. sap:	1.937956	ortalaması	5.225
6 inci deę. stan. sap:	1.430362	ortalaması	6.433334
7 inci deę. stan. sap:	1.678585	ortalaması	5.35
8 inci deę. stan. sap:	1.739293	ortalaması	6.091667
9 inci deę. stan. sap:	1.724738	ortalaması	6.508333
10 inci deę. stan. sap:	1.079961	ortalaması	7.958334
11 inci deę. stan. sap:	1.549466	ortalaması	7.45

Değişkenlerle Disk. Skorları Arasındaki Krelasyonlar

1 inci deęişkenle	0.335164	0.282704
2 inci deęişkenle	0.393169	0.278896
3 inci deęişkenle	0.328651	0.130198
4 inci deęişkenle	0.400050	-0.163893
5 inci deęişkenle	0.758876	-0.067396
6 inci deęişkenle	0.620682	0.118640
7 inci deęişkenle	0.626990	-0.236373
8 inci deęişkenle	0.806172	-0.052917
9 inci deęişkenle	0.518946	-0.034237
10 inci deęişkenle	-0.082304	-0.507995
11 inci deęişkenle	0.334088	0.389902

Tablo: 3

Birey No:	Skor 1	Skor 2
1	1.436861	0.167678
2	0.181433	0.349989
3	2.717693	-0.101737
4	2.317656	0.892135
5	0.420298	0.647905
6	-0.742215	0.263917
7	0.512692	-438843
8	1.124996	-1.741022
9	0.849811	-0.074573
10	1.352271	0.494094
11	0.369567	1.034023
12	0.031349	-0.952691
13	0.264339	0.606995
14	-0.110505	-1.142883
15	-0.577316	-1.334114
16	-0.674732	2.518101
17	2.078561	-0.909875
18	0.303241	0.492963
19	1.262930	-0.730998
20	1.132992	-0.663804

Tablo: 4

Birey No:	1. Gruba Uz.	2. Gruba Uz.	Ait Olması Gereken Gr.
1	4.36073	4.76883	1
2	2.96345	3.05332	1
3	8.55597	8.95431	1
4	6.16551	6.63127	1
5	5.13920	5.24958	1
6	5.16197	4.98432	2
7	3.77549	3.95417	1
8	4.41056	4.73155	1
9	2.82726	3.19840	1
10	4.04376	4.45680	1
-	-	-	-
64	0.06248	1.25999	1
65	1.93955	1.03843	2
66	2.18766	0.92989	2
67	1.73048	1.01750	2
68	1.99636	0.72892	2
69	0.42574	1.18223	1
70	1.93631	1.20676	2
71	1.97066	1.11854	2
72	2.09472	1.37116	2
73	2.09319	0.82937	2
74	1.60604	0.75458	2