T.C. Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt: 2, Sayı: 2, Yıl: 1987

GRUPLARIN KARŞILAŞTIRILMASINDA BİRER İSTATİSTİKSEL YÖNTEM OLARAK REGRESYON ANALİZİ VE AYRIŞIM ANALİZİ

Nalân CİNEMRE(*)

Ö Z E T

Uygun bir analitik teknik seçimiyle ilgili çalışmalarda çoğu zaman bağımlı değişkenin yalnızca «0» veya «1» değerini aldığı durumlarla karşılaşılabilmektedir. Örneğin, bir eğitim uzmanı öğrencileri not ortalamalarını gözönünde bulundurarak «başarılı» ve «başarısız» şeklinde, bir tıp uzmanı kendisine başvuranları «kanserli» ve «kanserli olmayanlar» şeklinde ayırmayı isteyebilir. Bir ekonomist araştırmasını yığını «ev sahibi olanlar» ve «ev sahibi olmayanlar» şeklinde iki gruba ayırarak sürdürmeyi isteyebilir. Regresyon analizi ve ayrışım analizi bu gibi durumlarda kullanılabilecek uygun birer istatistiksel tekniklerdir.

Bu çalışmada bağımlı değişkenin yalnızca «0» veya «1» değerini aldığı durumda regresyon analizi ile ayrışım analizi üzerinde durulmaktadır. Grup sayısının 3 ya da daha fazla sayıda olması durumunda da kullanılabilen ayrışım analizi, grup sayısının iki olması durumunda «iki gruplu ayrışım analizi» olarak anılmakta ve elde edilen fonksiyona «ayrışım fonksiyonu» denilmektedir. Grup sayısının iki olması durumunda elde edilen regresyon fonksiyonuna ise «doğrusal olasılık fonksiyonu» denilmektedir.

Aşağıdaki bölümlerde ayrışım fonksiyonu ile doğrusal olasılık fonksiyonunun elde edilişlerine elementer bir bakış açısıyla yaklaşılmakta ve iki fonksiyon arasındaki benzerliklerin ortaya çıkarılmasına çalışılmaktadır. Her iki tekniğin uygulanabilirliğini göstermek amacıyla çalışmanın sonunda minyatür bir veri üzerinde matematiksel işlemlere yer verilmektedir.

(*) Öğr. Gör. Dr., Çukurova Üniversitesi, İ.İ.B.F., İktisat Bölümü.

GİRİŞ

Regresyon analizi ile ayrışım analizinin uygulama ve fonksiyonlarında bulunan katsayıların yorumu birbirlerine benzemektedir. Her iki analizin temel amacı, iki ya da daha fazla sayıdaki bağımsız değişken ile bir bağımlı değişkenin tahmin edilmesi ya da tanımlanması şeklinde açıklanabilir. Ancak; iki analiz arasında, incelemeye alınan bağımlı değişkenin niteliği bakımından önemli farklar bulunmaktadır. Ayrışım analizi, bağımlı değişkenin sınıflama değişkeni olması durumunda uygun bir yöntemdir. Regresyon analizi ise, metrik nitelikli bağımlı değişkenler için uygundur (Cooley ve Lohnes, 1962, s. 85). Bununla birlikte, başlangıç noktaları birbirinden oldukça farklı olan iki analiz ile elde edilen sonuçlar büyük benzerlikler göstermektedir.

Nitekim, Fisher (1936, s. 119), bazı koşulların sağlanması durumunda regresyon analizi ile belirlenen regresyon katsayılarının, ayrışım analizi ile bulunan ayrışım katsayılarıyla orantılı olduğunu göstermiştir.

1. Grupların Karşılaştırılmasında İki İstatistiksel Yöntem

Çalışmanın bütününde, araştırma konusu olan yığının π_1 ve π_2 ile gösterilen iki gruptan meydana geldiği varsayılmaktadır. Gruplardaki birim sayısı sırası ile n_1 ve n_2'dir. Yığını oluşturan $n(n_1+n_2)$ birimin herbiri için, p sayıdaki değişken itibariyle birer ölçüm yapıldığı kabul edilmektedir. Ayrıca, (i)'inci birime ilişkin ölçümler (X_{i1}, X_{i2}, ... X_{ip}) olup, p elemanlı bir X vektörü (X) olarak gösterilmektedir.

~

1.1. Regresyon Analizi

Bilindiği gibi klasik regresyon analizi yönteminde, iki ya da daha çok sayıdaki değişken arasında bir bağıntının olup olmadığı ve bu bağıntının bir denklemle nasıl ifade edileceği araştırılabilir. Diğer taraftan, herhangi bir bağıntıda niceliksel faktörlerin etkisini göstermek için yalnızca «0» ve «1» değerlerini alabilen «iki-değerli» bağımlı değişken kullanılabilir. Çalışmanın bu kısmında «iki-değerli» bağımlı değişkenin kullanıldığı regresyon analizi ele alınmaktadır.

İlk önce Y bağımlı değişkenini, X bağımsız değişkeninin doğrusal birleşimi olarak yazalım.

$$Y_i = \alpha + k_1X_{i1} + k_2X_{i2} + \ldots + k_pX_{ip} + \varepsilon_i \qquad (1)$$

Yukarıdaki eşitlikte; $i=1, 2, \ldots, n_1, (n_1+1), \ldots n_2, (n_2+1), \ldots,$ n'dir. Denklemde yer alan bağımsız değişken katsayılarının tahmin edilmesinden önce şu tanımları yapalım:

$Y_i = 1$ $\quad Y_i \; \theta \; \pi_1$ için

$Y_i = 0$ $\quad Y_i \; \theta \; \pi_2$ için

Buradan, Y değişkeninin tüm değerlerinin dikkate alınmasıyla hesap edilen ortalaması ve varyansı aşağıdaki gibi olacaktır.

$$\overline{Y} = \frac{n_1}{n_1+n_2} = \frac{n_1}{n} \qquad (2)$$

$$\text{Var } (Y) = \sum_{i=1}^{n_j} \sum_{j=1}^{2} Y_{ji} - (n_1+n_2)\,\overline{Y}^2$$

$$= \frac{n_1 n_2}{n_1+n_2} = r \qquad (3)$$

Ayrıca, X_{tji} ; (J)'inci gruptaki (i)'inci gözlemin (t)'inci bağımsız değişken için ölçülen değeri olsun. Burada $i=1, 2, \ldots, n_j$; $t=1, 2, \ldots p$, $j=1, 2$ dir. Buna göre, (t)'inci değişkenin (j)'inci gruptaki ortalaması :

$$\overline{X}_{tj.} = \frac{\sum_{i=1}^{n_j} X_{tji}}{n_j} \qquad (4)$$

olur. (t)'inci değişkenin yığındaki ortalaması ise aşağıdaki gibi tanımlanabilecektir.

$$\overline{X}_{t..} = \frac{\sum_{i=1}^{n_j} \sum_{j=1}^{2} X_{tji}}{n} = \frac{n_1\overline{X}_{t1.} + n_2\overline{X}_{t2.}}{n} \qquad (5)$$

Yapılan bu tanımlara göre, t değişkeninin grup ortalamaları arasındaki fark şu şekilde olacaktır:

$$d_t = \overline{X}_{t1.} - \overline{X}_{t2.}$$

Diğer taraftan, p sayıdaki değişkenin gözönünde bulundurulmasıyla hesap edilecek grup ortalamaları arasındaki farklar aşağıdaki gibi bir sütun vektörü ile ifade edilebilir.

$$\underset{\sim}{d} = (d_1 \; d_2 \; \ldots\ldots\ldots \; d_p)' \qquad (6)$$

X_t bağımsız değişkeni ile Y'nin kovaryansı ise,

$$S_{ty} = \sum_{i=1}^{n_j} \sum_{j=1}^{2} X_{tji} \, Y_{ji} - n\bar{Y}\bar{X}_{t..}$$

$$= n_1\bar{X}_{t1.} - \frac{n_1}{n_1+n_2} \left(n_1\bar{X}_{t1.} + n_2\bar{X}_{t2.} \right)$$

$$= \frac{n_1 n_2}{n_1+n_2} \cdot d_t = r d_t \qquad (7)$$

şeklinde hesap edilebilir.

Bu tanımların da kullanılmasıyla (1) no'lu eşitlikte bulunan regresyon katsayılarının tahmin edilmesinde, gerçek Y değerleri ile tahmin edilen $\hat{Y}$ değerleri arasındaki farkların (hataların) kareleri toplamını en küçükleyecek şekilde en iyi doğruyu tahmin etme yöntemi olarak tanımlanabilen (Akdeniz, 1976, s. 419) en küçük kareler (EKK) yöntemi kullanılır.
Yöntem;

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \qquad (8)$$

şeklinde tanımlanabilen kareler toplamının en küçüklenmesi esasına dayanmaktadır.

SSE'yi iki ayrı değerin toplamı olarak düşünmek ve aritmetik işlemleri buna göre sürdürmek, katsayıların tahmininde basitlik sağlamaktadır. Bu amaçla (8) no'lu eşitlik aşağıdaki gibi yazılabilir :

$$SSE = \sum_{i=1}^{n_1} [1 - (+ k_1X_{i1} + K_2X_{i2} + \ldots + k_pX_{ip})]^2 +$$

$$\sum_{i=n_1+1}^{n} [0 - (+ k_1X_{i1} + k_2X_{i2} + \ldots + k_pX_{ip})]^2 \qquad (9)$$

(9) no'lu ifadeyi en küçükleyecek tahmini değerleri bulabilmek için, SSE'nin bu elemanlar itibariyle kısmı türevleri alınıp sıfıra eşitlenir. α 'ya göre alınan kısmi türev aşağıdaki gibi yazılabilir:

$$\frac{\partial SSE}{\partial \alpha} = -2 \sum_{i=1}^{n_1} [1-(+ k_1X_{i1} + k_2X_{i2} + \ldots + k_pX_{ip})] -2$$

$$\sum_{i=n_1+1}^{n} -(\alpha + k_1X_{i1} + k_2X_{i2} + \ldots + k_pX_{ip}) \qquad (10)$$

(10) no'lu ifadenin sıfıra eşitlenip α'ya göre çözümlenmesi ile α'nın EKK'ler tahmin edicisi aşağıdaki gibi bulunabilir.

$$\hat{\alpha} = \overline{Y} - \sum_{t=1}^{p} k_t\overline{X}_{t..}$$

Ayrıca,

$$x_{it} = X_{it} - \frac{\sum_{i=1}^{n} X_{it}}{n} \qquad t = 1, 2, \ldots, p$$

yani, $X_{it} - \overline{X}_{.t}$ şeklinde tanımlanan yeni bağımsız değişkenler kümesi ile SSE aşağıdaki gibi yazılabilir.

$$SSE = \sum_{i=1}^{n_1} [1-(k_1x_{i1} + k_2x_{i2} + \ldots + k_px_{ip})]^2 +$$

$$\sum_{i=n_1+1}^{n} [0-(k_1x_{i1} + \ldots + k_px_{ip})]^2$$

Buradan $\underset{\sim}{k}$'nın EKK'ler tahmin edicisi;

$$\hat{\underset{\sim}{k}} = \underset{\sim}{S}^{-1} \frac{n_1 n_2}{n} \underset{\sim}{d} = \underset{\sim}{S}^{-1} \underset{\sim}{r} \underset{\sim}{d} \qquad (11)$$

şeklinde hesap edilir. Burada $\underset{\sim}{S}$, S_{tk}'nın $(p \times p)$ boyutlu matrisidir ve S_{tk} aşağıdaki gibi tanımlanmaktadır.

$$S_{tk} = \sum_{i=1}^{n_j} \sum_{j=1}^{2} X_{tji} X_{kji} - n \overline{X}_{t..} \overline{X}_{k..} \qquad (12)$$

Elde edilen tahmini değerlerin 1 no'lu eşitlikte yerine konması ile;

$$\hat{Y} = \sum_{i=1}^{p} \hat{k} X_{tji} + \hat{\alpha} \qquad (13)$$

elde edilir.

1.2. Ayrışım Analizi

Lachenbruch, ayrışım analizini iki ya da daha fazla sayıdaki gruptan birine ait olduğu bilinmekle birlikte, hangisine ait olduğu bilinmiyen birimin, üzerinde $X_1, X_2, \ldots, X_p$ değişkenleri için yapılmış ölçümlere dayanarak orijininin belirlenmesi işlemi olarak tanımlamaktadır (Lachenbruch, 1975, s. 1.).

Üzerinde (p) sayıdaki değişken için ölçüm yapılmış birimlerle ayrışım analizi, önceden belirlenmiş gruplar arasında en iyi ayırımı sağlayacak (p) sayıdaki bağımsız değişkenin doğrusal birleştirmesinin elde edilmesi şeklinde açıklanabilir. Ayrışım çözümlemesi için bu birleştirmeler aşağıdaki gibi gösterilebilir.

$$Y = Z_1X_1 + Z_2X_2 + \ldots + Z_pX_p, \quad \text{veya kısaca} \qquad (14)$$

$$Y = \underset{\sim}{Z}' \underset{\sim}{X}$$

Yukarıdaki eşitlikte: Y ayrışım değerini, $\underset{\sim}{Z}$ ayrışım katsayılarını ve $\underset{\sim}{X}$

bağımsız (ayrıştırıcı) değişkenleri göstermektedir. (14) no'lu eşitlikte yeralan ayrışım katsayılarının tahmin edilmesinde, Y değişkeni itibariyle elde edilecek grup ortalamaları arasındaki farkın karesinin, Y'nin varyansına oranının en büyüklenmesi yöntemi kullanılır (Mardia, Kent and Biby, 1979, s. 319).

Koşulu istatistiksel olarak ifade etmeden önce şu tanımları yapalım.

X vektör değişkeni itibariyle, grupların eşit oldukları kabul edilen varyans-kovaryans matrisleri aşağıdaki gibi yazılabilmektedir.

$$\underset{\sim}{\Sigma} = E(\underset{\sim}{X} - \underset{\sim}{\mu_j})(\underset{\sim}{X} - \underset{\sim}{\mu_j})' \quad , \quad j=1,2 \qquad (15)$$

Hemen belirtelim ki, ayrışım analizinde grupların eşit dağılımlı oldukları (eşit varyans-kovaryansları olduğu) ve çok değişkenli normal yığından geldiği kabul edilmektedir.

Yine X vektör değişkeni itibariyle grup ortalamaları sırası ile;

$$\underset{\sim}{\mu_1} = E(\underset{\sim}{X} / \pi_1) \qquad (16)$$

$$\underset{\sim}{\mu_2} = E(\underset{\sim}{X} / \pi_2)$$

şeklinde yazılabilir.

Diğer taraftan, Y değişkeni itibariyle grup ortalamaları sırası ile;

$$\mu_{1Y} = E(Y/\pi_1) = E(\underset{\sim}{Z}'\underset{\sim}{X} / \pi_1) = \underset{\sim}{Z}'\underset{\sim}{\mu_1}$$

$$\mu_{2Y} = E(Y/\pi_2) = E(\underset{\sim}{Z}'\underset{\sim}{X} / \pi_2) = \underset{\sim}{Z}'\underset{\sim}{\mu_2} \qquad (17)$$

olarak yazılabilecektir.

Ayrıca grupların varyans-kovaryans matrislerinin eşit olduğu kabul edildiğinden, Y'nin varyansı;

$$\text{Var}\ (Y) = \text{Var}\ (\underset{\sim}{Z}'\underset{\sim}{X}) = \underset{\sim}{Z}'\ \text{Cov}(X)\ \underset{\sim}{Z} = \underset{\sim}{Z}'\ \underset{\sim}{\Sigma}\ \underset{\sim}{Z} \qquad (18)$$

şeklinde elde edilir.

Bu tanımlardan sonra, en büyüklenmek istenen oran aşağıdaki gibi yazılabilecektir.

$$Q = \frac{(Z' \mu_1 - Z' \mu_2)^2}{Z' \Sigma Z} \qquad (19)$$

Q'yu en büyükleyecek Z vektörünün elemanlarını bulmak için (19) no'lu eşitliğin bu elemanlar itibariyle kısmi türevleri alınıp elde edilen değer sıfıra eşitlenir. Türevin alınmasında kolaylık sağlaması bakımından (19) no'lu eşitlik aşağıdaki gibi yazılabilir.

$$Q = \frac{[\sum_{j=1}^{p} Z_j(\mu_{1j} - \mu_{2j})]^2}{\sum_{j=1}^{p}\sum_{i=1}^{p} Z_j Z_i \sigma_{ij}} \qquad (20)$$

(20) no'lu eşitlikte;

$$\sigma_{ij} = \begin{cases} \text{Var}(X_{ij}) & , \ i=j \ \text{için} \\ \text{Cov}(X_j X_i) & , \ i \neq j \ \text{için} \end{cases}$$

(20) no'lu eşitliğin türevinin alınıp sıfıra eşitlenmesinden sonra bazı matematiksel işlemler ve kısaltmaların yapılmasıyla;

$$\sum_{j=1}^{p} (\mu_{1j} - \mu_{2j}) = \frac{[\sum_{j=1}^{p} Z_j (\mu_{1j} - \mu_{2j})][\sum_{j,i=1}^{p} Z_i \sigma_{ij}]}{Z_j Z_i \sigma_{ij}} \qquad (21)$$

elde edilir. Buradan;

$$(\mu_1 - \mu_2) = \Sigma Z \frac{Z' (\mu_1 - \mu_2)}{Z' \Sigma Z} \qquad (22)$$

şeklinde yazılabilir.

(22) no'lu eşitliğin her iki tarafı Σ^{-1} ile çarpılarak;

$$\Sigma^{-1}\ (\mu_1-\mu_2) = Z \frac{Z'\ (\mu_1-\mu_2)}{Z'\ \Sigma\ Z} = Z\ C \qquad (23)$$

bulunur.

Z yığını gruplara ayırma amacıyla kullanıldığından, arzu edilen herhangi bir sabit sayı ile (C) çarpılabilir. Böylece Z'nin $\Sigma^{-1}\ (\mu_1-\mu_2)$ ile orantılı olduğu gözönünde tutularak X vektör değişkeninin doğrusal birleşimi olarak Y, aşağıdaki gibi yazılabilecektir.

$$Y = Z'\ X = C\ (\mu_1-\mu_2)'\ \Sigma^{-1}\ X\ ,\ C \neq 0 \qquad (24)$$

Diğer taraftan C=1 olarak seçildiğinde;

$$Y = Z'\ X = (\mu_1-\mu_2)'\ \Sigma^{-1}\ X \qquad (25)$$

olarak elde edilir.

Kısaca belirtmek gerekirse, ayrışım katsayıları

$$\hat{Z} = \Sigma^{-1}\ d \qquad (26)$$

şeklinde yazılabilir.

Buradan (26) no'lu eşitliğin incelenmesi ve regresyon katsayılarının tahmini olan ve (11) no'lu eşitlik ile ifade edilen denklemin gözden geçirilmesiyle, ayrışım fonksiyonu katsayıları ile doğrusal olasılık fonksiyonu katsayılarının orantılı olduğu görülebilecektir.

Bilindiği gibi, yukarıda yer alan ayrışım analizi ile ilgili çözümde grupların eşit olan ve bilinen varyans-kovaryans matrisleri kullanılmıştır. Gruplara ilişkin parametre değerlerinin bilinmemesi durumunda ayrışım fonksiyonunun ne şekilde hesap edileceği aşağıda kısaca açıklanmaktadır.

π_1 ve π_2 gruplarından rassal olarak çekilen örnekler sırasiyle m_1 ve m_2 olsun. Ayrıca, bu örnekleri meydana getiren her bir birim üzerinde (p) sayıda değişken için ölçüm yapıldığı kabul edilmektedir ve örneklerdeki birim sayısı sırası ile n'_1 ve n'_2 dir.

$\overline{X}_1$ ve $\overline{X}_2$ örnek ortalama vektörleri, yığın ortalama vektörlerinin, $\underset{\sim}{}$ $\underset{\sim}{}$

S ise grup varyans-kovaryans matrislerinin sapmasız tahmin edicisi ola-
$\underset{\sim}{}$

rak kullanıldığında, ayrışım fonksiyonu (25) no'lu eşitlikte, parametre değerleri yerine tahmini parametre değerlerinin yerleştirilmesiyle aşağıdaki gibi elde edilir.

$$Y = \hat{Z}' X = (\overline{X}_1 - \overline{X}_2)' S^{-1} X \qquad (27)$$
$\underset{\sim}{}\ \underset{\sim}{}$ $\underset{\sim}{}$ $\underset{\sim}{}$ $\underset{\sim}{}$ $\underset{\sim}{}$ $\underset{\sim}{}$

veya

$$\hat{Z} = S^{-1} d \qquad (28)$$
$\underset{\sim}{}$ $\underset{\sim}{}$ $\underset{\sim}{}$

(26) no'lu ifade de, $\overline{X}_1$, $\overline{X}_2$ ve S değerleri şöyledir :
$\underset{\sim}{}$ $\underset{\sim}{}$ $\underset{\sim}{}$

$$\overline{X}_1 = \frac{1}{n'_1} \sum_{j=1}^{n'_1} X_{1j} \quad \text{ve} \quad \overline{X}_2 = \frac{1}{n'_2} \sum_{j=1}^{n'_2} X_{2j}$$
$\underset{\sim}{}$ $\underset{\sim}{}$

Burada, $X_1 = (X_{11}\ X_{12}\ \ldots\ X_{1p})$ ve $X_2 = (X_{21}\ X_{22}\ \ldots\ X_{2p})$
$\underset{\sim}{}$ $\underset{\sim}{}$

olarak alınmıştır.

$$S = \frac{(n'_1 - 1) S_1 + (n'_2 - 1) S_2}{(n'_1 + n'_2 - 2)} \qquad (29)$$
$\underset{\sim}{}$ $\underset{\sim}{}$ $\underset{\sim}{}$ $\underset{\sim}{}$

(29) no'lu ifadede yeralan S_1 ve S_2 aşağıdaki gibidir.
$\underset{\sim}{}$ $\underset{\sim}{}$

$$S_1 = \frac{1}{(n'_1 - 1)} \sum_{j=1}^{n'_1} (X_{1j} - \overline{X}_1)(X_{1j} - \overline{X}_1)' \qquad (30)$$
$\underset{\sim}{}$ $\underset{\sim}{}$ $\underset{\sim}{}$ $\underset{\sim}{}$

$$\underset{\sim}{S}_2 = \frac{1}{(n'_2-1)} \sum_{j=1}^{n'_2} (\underset{\sim}{X}_{2j}-\overline{\underset{\sim}{X}}_2)\,(\underset{\sim}{X}_{2j}-\overline{\underset{\sim}{X}}_2)' \qquad (31)$$

Çalışmanın bütününde gerek ayrışım fonksiyonu gerekse regresyon fonksiyonuna ilişkin olarak elde edilmiş olan katsayıların tahminleri aşağıda topluca sunulmaktadır.

1. Σ'lı ayrışım fonksiyonu (26) no'lu eşitlik ile gösterildiği gibi,

$$\hat{\underset{\sim}{Z}} = \underset{\sim}{\Sigma}^{-1}\,(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2) = \underset{\sim}{\Sigma}^{-1}\,\underset{\sim}{d}$$

2. S'li ayrışım fonksiyonu (28) no'lu eşitlik ile gösterildiği gibi,

$$\hat{\underset{\sim}{Z}} = \underset{\sim}{S}^{-1}\,(\overline{\underset{\sim}{X}}_1-\overline{\underset{\sim}{X}}_2) = \underset{\sim}{S}^{-1}\,\underset{\sim}{d}$$

3. S'li regresyon fonksiyonu (11) no'lu eşitlik ile gösterildiği gibi,

$$\hat{\underset{\sim}{k}} = \underset{\sim}{S}^{-1}\,\frac{n_1\,n_2}{n}\,(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2) = \underset{\sim}{S}^{-1}\,\underset{\sim}{r}\,\underset{\sim}{d}$$

2. Önemlilik Testleri

Bilindiği gibi klasik regresyon analizinde, regresyonun istatistik bakımından önemliliği test edilebilmektedir. Böyle bir hipotez testinde yokluk ve alternatif hipotezler aşağıdaki gibi ifade edlebilir.

$H_o : R = 0$

$H_A : R \neq 0$

$R = 0$ şeklnde ifade edilen yokluk hipotezini test etmek için geliştirilen test istatistiği, (32) no'lu eşitlik ile ifade edilmektedir.

$$F = \frac{R^2}{1-R^2}\,\frac{n-p}{p-1} \qquad (32)$$

Yukarıda yer alan eşitlikte n; yığındaki birim sayısını, p; değişken

sayısını ifade etmekte olup $R^2 = \hat{\beta} X'Y'/Y'Y$ ile hesap edilmektedir (Ladd, 1966, s. 881). Bu teoriye paralel olarak R^2'nin hesap edilmesinde $\hat{k}$ kullanılabilir. $\hat{k}$'nın $\hat{\beta}$ yerine kullanılmasıyla R^2 aşağıdaki gibi bulunur.

$$R^2 = \frac{\hat{k}' \frac{n_1 n_2}{n} d}{\frac{n_1 n_2}{n}} = d' S^{-1} d = \hat{Z}' \frac{n_1 n_2}{n} d \qquad (33)$$

(33) no'lu eşitlik ile gösterilen R^2 ile hesaplanan F oranı aşağıda gösterildiği gibidir ve ayrışım fonksiyonunun önemini test etmede kullanılır. F'nin serbestlik derecesi p ve n−p−1 olan F dağılımına sahip olduğu gösterilebilir.

$$F = \frac{R^2}{1-R^2} \quad \frac{n-p-1}{p} \qquad (34)$$

Hatırlanacağı gibi klasik regresyon analizinde aşağıda kısaca açıklanan varsayımlar yapılmaktadır:

1. Bağımsız değişkenler sabit değerlerdir.
2. Bağımlı değişken normal dağılıma sahiptir.
3. Normal dağıldığı kabul edilen hata teriminin varyansı sabittir.

Ayrışım analizinde ise, bağımlı ve bağımsız değişkenlerin rolleri değişmiştir (Kshirsagar, 1972, s. 211). Regresyon analizinde yapılan varsayımın tersine olarak, ayrışım analizinde X'ler rassal değişken olarak işlem görürken Y'lerin sabit olduğu kabul edilir. Bununla birlikte, X'lerin çok değişkenli normal yığından geldiğinin kabul edilmesi ve grupların eşit dağılımlı olmalarının sağlanması durumunda (34) no'lu eşitlikle ifade edilen F oranı hala en uygun test istatistiğidir (Ladd, 1966, s. 881). Bir başka deyişle ayrışım fonksiyonunun istatistik bakımdan önemliliğini test etmede kullanılabilir. Hemen belirtelim ki bu çalışmada açıklanan regresyon analizinde kullanılan bağımlı değişken normal dağılıma sahip değildir. Rassal bir değişken olduğu da söylenemez. Bununla birlikte, kullanılan yöntem bu sapmalara karşı oldukça sağlamdır.

SONUÇ

1. Doğrusal olasılık fonksiyonu katsayıları ayrışım fonksiyonu katsayılarıyla orantılıdır.

2. Doğrusal olasılık fonksiyonu için hesaplanan determinasyon katsayısı (R^2) ile ayrışım fonksiyonu için hesaplanan ve (33) no'lu eşitlik ile gösterilen R^2 aynıdır.

3. Gerek regresyon fonksiyonunun gerekse ayrışım fonksiyonunun istatistik bakımdan önemliliğini test etmede kullanılan F oranları birbirine eşittir.

Bilindiği gibi bu çalışmada bağımlı değişkenin «0» ve «1» değerlerini aldığı regresyon analizi üzerinde durulmuştur. Bununla birlikte bağımlı değişkenin «1» ve «-1» veya «n_2/n» ve «$-n_1/n$» değerlerini aldığı regresyon analizleri üzerinde de çalışılabilir. Nitekim Fisher «n_2/n» ve «$-n_1/n$» değerlerini alan bağımlı değişken ile regresyon analizi üzerinde durarak regresyon katsayıları ile ayrışım katsayılarının orantılı olduklarını göstermiştir (Fisher, 1936).

UYGULAMA

Regresyon Analizi Yöntemi ve Ayrışım Analizi Yöntemlerinin Minyatür Veri Üzerinde Gösterilmesi.

Bu bölümde 3 değişkenli $\underset{\sim}{X}_2$ vektör değişkenine ait 2 gruplu ve her grupta 5 birim bulunan toplam 10 birimlik basit bir araştırma planı kabul edilerek ilk önce «doğrusal olasılık fonksiyonu» daha sonra aynı veri kullanılarak «ayrışım fonksiyonu» elde edilecektir. Böylece simge ve formüllerle anlatılan yöntemlere daha çok açıklık kazandırılmaya çalışılacaktır. Bulunan fonksiyonların benzer yanlarının incelenmesi uygulamanın son bölümünü oluşturmaktadır.

π_1 ve π_2 grupları göstermek üzere minyatür veri aşağıdaki gibidir:

π_1			π_2		
X_1	X_2	X_3	X_1	X_2	X_3
12	12	2	16	12	5
14	10	4	18	10	6
10	18	3	20	13	7
12	16	2	22	8	7
12	14	4	19	12	5
Toplam 60	70	15	Toplam 95	55	30
Ortalama 12	14	3	Ortalama 19	11	6

3 değişken itibariyle hesap edilen grup ortalamaları arasındaki farklar 6 no'lu eşitliğin kullanılmasıyla

$$\underset{\sim}{d} = (12-19 \quad 14-11 \quad 3-6) = (-7 \quad 3 \quad -3)$$

şeklinde elde edilir.

Elemanları 12 no'lu eşitlik ile elde edilen $\underset{\sim}{S}$ matrisi ise aşağıda gösterilmektedir.

$$S = \begin{bmatrix} 3.5 & -3.125 & 1.125 \\ -3.125 & 7 & -0.875 \\ 1.125 & 0.875 & 1 \end{bmatrix}$$

S'nin tersinin alınmasıyla elde edilen S^{-1}, $\underset{\sim}{d}$ ve r'nin 11 no'lu eşitlikte yerine konmasıyla $\hat{k}$ aşağıdaki şekilde hesap edilecektir.

$$\hat{k} = S^{-1} \cdot r \cdot d = \begin{bmatrix} 0.6663 & 0.22897 & -0.54986 \\ 0.22897 & 0.23896 & -0.04845 \\ -0.54986 & -0.04845 & 1.57599 \end{bmatrix} \frac{25}{10} \begin{bmatrix} -7 \\ 3 \\ -3 \end{bmatrix}$$

$$= \begin{bmatrix} -5.8295 \\ -1.8514 \\ -1.8343 \end{bmatrix}$$

Minyatür veri aynı kalmak koşuluyla ayrışım fonksiyonun elde edilmesine ilişkin sayısal işlemler aşağıdaki gibidir.

30 no'lu ifade ile tanımlanan S_1 aşağıdaki gibi elde edilmiştir.

$$S_1 = \begin{bmatrix} 2 & -4 & 0.5 \\ -4 & 10 & -1 \\ 0.5 & -1 & 1 \end{bmatrix}$$

Benzer şekilde $\underset{\sim}{S}_2 = \begin{bmatrix} 5 & -2.25 & 1.75 \\ -2.25 & 4 & -0.75 \\ 1.75 & -0.75 & 1 \end{bmatrix}$ olacaktır.

$\underset{\sim}{S}_1$ ve $\underset{\sim}{S}_2$'nin 29 no'lu formülde yerine konmasıyla $\underset{\sim}{S}$,

$$\underset{\sim}{S} = \begin{bmatrix} 3.5 & -3.125 & 1.125 \\ -3.125 & 7 & -0.875 \\ 1.125 & -0.875 & 1 \end{bmatrix}$$

şeklinde elde edilir. Hatırlanacağı gibi bu matris $(\underset{\sim}{S})$ 12 no'lu eşitliğin kullanılmasıyla hesap edilen matris ile aynıdır. Buradan; $\underset{\sim}{S}^{-1}$ ve $\underset{\sim}{d}$'nin ayrışım fonksiyonu katsayılarının tahmini olan (28) no'lu eşitlikte yerine konmasıyla,

$$\hat{\underset{\sim}{Z}} = \underset{\sim}{S}^{-1} \cdot \underset{\sim}{d} = \begin{bmatrix} 0.6669 & 0.22897 & -0.54986 \\ 0.22897 & 0.23896 & -0.04845 \\ -0.54986 & -0.04845 & 1.57599 \end{bmatrix} \begin{bmatrix} -7 \\ 3 \\ -3 \end{bmatrix}$$

$$= \begin{bmatrix} -2.3318 \\ -0.74056 \\ -0.7336 \end{bmatrix}$$

olarak elde edilir.

Doğrusal olasılık fonksiyonu ile ayrışım fonksiyonu katsayılarının tahmin edilmesinden sonra bu fonksiyonların istatistik bakımdan önemlilikleri test edilebilir. Şimdi sırasıyla ilk önce d.O.f.'nun sonra da ayrışım fonksiyonunun istatistik bakımdan önemliliğini test edelim.

Hatırlanacağı gibi $R=0$ şeklinde ifade edilen yokluk hipotezini test etmek için geliştirilen test istatistiği 33 no'lu eşitlik ile tanımlanmaktadır. İfadede yer alan R^2, doğrusal olasılık fonksiyonu için;

$$R^2 = \frac{[\,-5.8295 \quad -1.8514 \quad -1.834\,] \quad (2.5) \quad \begin{bmatrix} -7 \\ 3 \\ -3 \end{bmatrix}}{(2.5)} = 46.3085$$

olarak bulunur. R^2 için bulunan bu değerin (33) no'lu eşitlikte yerine konmasıyla elde edilen F değeri ise şöyle olacaktır.

$$F = \frac{46.3085}{1-46.3085} \cdot \frac{10-3-1}{3} = -2.284832$$

R^2 ve F değerleri ayrışım fonksiyonu için aşağıdaki gibi hesap edilmişlerdir.

$$R^2 = \hat{Z} \cdot r \cdot d = [\,-2.3318 \quad -0.74056 \quad -0.7336\,] \quad (2.5) \quad \begin{bmatrix} -7 \\ 3 \\ -3 \end{bmatrix}$$

$$F = \frac{46.3085}{1-46.3085} \cdot \frac{7}{3} = -2.384832$$

Görüldüğü gibi biri doğrusal olasılık fonksiyonu için diğeri ayrışım fonksiyonu için olma üzere hesaplanan iki F oranı birbirine eşittir.

$\underset{\sim}{\hat{k}}$ ve $\underset{\sim}{\hat{Z}}$ ile gösterilen matrislerin karşılaştırılması sonucu açıkça görüleceği gibi $\hat{k}$ ile $\hat{Z}$ birbirleriyle orantılıdır. Başka bir deyişle $\hat{k}$ biliniyor iken bu matrisin elemanlarının r'ye bölünmesiyle $\underset{\sim}{\hat{Z}}$ elde edilebilir. Bunun tersi de doğrudur. Yani $\underset{\sim}{\hat{Z}}$ bilindiği zaman da bu matrisin elemanlarının r ile çarpımı $\underset{\sim}{\hat{k}}$'yı verir.

K A Y N A K Ç A

Akdeniz, Fikri; Olasılık ve İstatistik, Ankara, 1976.

Cooley, William Wand Paul R. Lohnes; Multivariate Data Analysis. New York: John Wiley and Sons, 1962.

Fisher, R.A.; «The Use of Multiple Measurementi in Taxonomic Problems», Ann. Eugen. Vol. 7 (1936). s. 179-188.

Kshirsagar, Anant. M; Multivariate Analysis, Marcal Dekker Inc. New York, 1972.

Lachenbruch, Peter A; Discriminant Analysis, New York: Hafner Press A Division of MacMillian Publishing Co. Inc. 1975.

Ladd, George W; «Linear Probability Functions and Discriminant Functions», Econometrica. Vol 34 (1966), s. 873-885.

Mardia.K.V., Kent.J.T and Biby.J.M.; Multivariate Analysis, New York: Academic Press, 1979.

REGRESSION AND DISCRIMINANT ANALYSIS AS STATISTICAL METHODS FOR THE COMPARISON OF GROUPS

In this article, regression analysis in two groups case and discriminant analysis are briefly discussed. Certain identities are obtained.

It is assumed that the population consists of two sub_populations (group). Discriminant analysis can be handled either two groups or mul tiple groups (three or more). When two groups are involved the method is referred to as two_group discriminant analysis. If a regression function in which dependent variable has only two values it is referred to as linear probability function.

The aim of this article is to show the relation between discriminant function and linear probability function. To do this linear probability functions are briefly discussed. The theory of discriminant analysis in two group case is then presented.