İŞLETMELER İÇİN KORELASYON-REGRESYON ANALİZİ

Doç. Dr. M. Kemal YOĞURTÇUGİL

I

Bugün özellikle ilmi kanunlara bağlanabilen, yâni sebep - netice münasebeti ile izah olunabilen hadiseler müstesna, başta iktisat olmak üzere sosyal ilimler insanın gerek fert gerek bir gurubun üyesi sıfatıyle davranışlarını tayin eden temayül, prensip ve kanunları keşfetme hususunda büyük çabalar sarfetmektedirler(1). Bu konudaki gelişmelerde psikoloji, sosyoloji ve istatistik gibi ilimlerden ve bunların tâli kollarından faydalanılmaktadır. Elde edilen neticelere göre, bu ilimlerde herzaman ve heryerde kullanılacak adedi formüller şeklinde miktar kanunları tesisinin imkânsız olduğu, bununla beraber bazı istatistik tahliller ile tahminlerde bulunulabileceği ve bu tahminlerin muayyen bir ihtimâlle geçerlilik hudutlarının hesaplanabileceği kabul edilmiştir. Ancak bulunacak kanunların diğer bütün şartlar aynı kalmak üzere şu şartlar mevcutsa şu neticelerin doğmasını bekliyebiliriz şeklinde olacağı; diğer bir ifade ile bir kat'iyet değil, bir ihtimâl ifade edeceği ve bir bütünün cüz'lerinden her birini ayrı ayrı kat'î olarak bağlamaksızın bütün için geçerli sayılacağı da hatırlanmalıdır.

Bir hadise hakkında sahip olduğumuz bilgiye istinaden diğer bir hadise hakkında tahmin yapabiliyorsak bu iki hadise arasında bir münasebetin varlığından bahsedilebilir. Münasebet, fonksiyonel veya istatistikî olabilir. Belli bir hadisenin her kıymeti için diğer hadise ancak bir kıymet alabiliyorsa yapılan tahmin kat'î ve fonksiyoneldir. Aksine belli hadisenin her kıymeti için diğer hadise muhtelif kıymetler arasından herhangi birini alabiliyorsa münasebet istatistikî veya ihtimalîdir.

İki veya daha çok sayıda değişken arasındaki ilişkileri ortaya koymak için korelasyon ve regresyon analizlerine başvurulur. De-

(1) .— K. Tosun «İşletme ve Müesseselerde Sevk ve İdare» İstanbul 1961 sf. 559

ğişkenler arasındaki paralelliğin derecesi korelasyon katsayısı ile, sayısal bağıntısının biçimi ise regresyon denklemi ile bulunabilir(2). Bir X ve Y değişkenler gurubu arasındaki ilişkinin ölçülmesinde aşağıdaki sıra takip edilmelidir.

a. Münasebetin şekli «Regresyon denklemi» nin bulunması,

b. Dağılmanın derecesinin ölçülmesi «tahminin standart hatasının hesabı» ve

c. Bir nisbî münasebet ölçüsü olan korelasyon katsayısının hesabı.

Münasebetin şeklinin bulunması özellikle serpilme diyagramının çizimine dayanır. Bir kartezyen koordinat sisteminde değişkenlerden biri X ekseninde diğeri Y ekseninde gösterilmek suretiyle işaretlenen noktaların teşkil ettikleri şekle Serpilme Diyagramı denir(3). Bu diyagramda, iki değişken gurubunun karşılıklı olarak aldıkları değerlerin kesişme noktalarının yönü bir en küçük kareler doğrusu veya eğrisi ile belirtilebilir. Verilerin gösterdiği temayül doğrusal ise bunu gösteren eşitlik $Y = a + bX$ olup a ve b değerleri, verilen Y değerleri ile hesaplanacak Y′ değerleri arasındaki farkların kareleri toplamı minimum olacak şekilde tayin edilir. Bu ise $\Sigma n_i (Y_i - a - b \cdot X_i)^2 =$ Minimum eşitliğini sağlayan a ve b nin araştırılması demektir ve bu parametreler

$$\Sigma Y = n \cdot a + b \Sigma X$$

$$\Sigma Y.X = a \Sigma X + b \Sigma X^2$$

normal denklemlerinin birlikte çözülmesiyle elde edilirler. Bulunacak doğru denklemi belli bir (X) değeri için (Y) nin teorik değerini tahmin etmede kullanılır. Eğer değişkenler arasındaki ilişki mükemmel değilse gerçek değerler ile teorik değerler birbirlerinden farklı

---

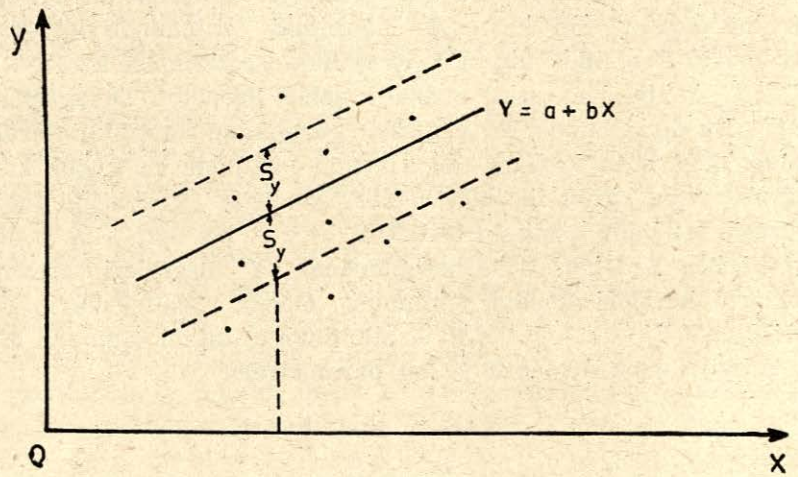
(2) .— Hesaplamalar değişkenler arasındaki kovariyanslarla değişkenlerin variyanslarına dayandığından bunların geçerliliği ancak X ve Y değişkenlerin tesadüfî değişkenler olması, bölünmelerinin normal bölünmeyi andırması veya hiç olmazsa fazla çarpık bulunmaması varsayımlarının mevcudiyetine bağlıdır. Bkz. S. Kendir. «Zaman Serilerinde Korelasyon ve Regresyon Analizleri» Türk İstatistik Derneği Dergisi Aralık 1968.

(3) .— H. Arkın - R. R. Colton (Terc. S. Kendir) «İstatistik Metodlar» Ankara 1968 sf. 82

olacaklardır. Regresyon doğrusu etrafındaki bu dağılmanın ölçüsü tahminin standart hatası adını alır ve en basit şekli

$$S_y = \sqrt{\frac{\Sigma (Y - Y')^2}{n}}$$

formülüyle yapılanıdır. Eğer gerçek değerler regresyon doğrusu etrafında normal dağılım gösteriyorlarsa, teorik olarak serpilme diyagramındaki noktaların % 68'i, elde edilen bu değer kadar regresyon doğrusunun altında ve üstünde çizilecek paralel doğrular arasında bulunacaktır (Grafik 1.). Paralel çizgiler 1.96 S_y mesafesinde iseler bu oranın 0.95 olması beklenecektir.



Grafik _1

Bu ölçünün büyüklüğü değişkenler arasındaki münasebetin bir ölçüsü olarak kullanılabilir. Ancak kıymeti Y değişkeninin birimine ve değişim sahasına bağlı olduğundan bu mahzurun giderilmesi için

$$\frac{\Sigma (Y - Y')^2}{\Sigma (Y - \overline{Y})^2} = \frac{S_y^2}{\sigma_y^2}$$

Y = Asli Değerler
Y' = Teorik Değerler
$\overline{Y}$ = Arit. Ortalama
Σ = Toplam

nisbî ölçüsü hesaplanmalıdır. Bu değere bağımsız değişkenin izah edemediği değişme nisbeti denir. Toplam değişme bir olduğu ve bağımsız değişkenin izah ettiği ve etmediği değişmelerin toplamından ibaret bulunduğuna göre, bağımsız değişkenin izah ettiği değişme nisbetine r^2 dersek

$$r^2 = 1 - \frac{S_y^2}{\sigma_y^2} \quad \text{bulunur.}$$

Determinasyon katsayısı adı verilen bu değerin kare kökü korelasyon katsayıdır. $r = 1$ ise iki regresyon doğrusu ($Y = a + bX$ ve $X = a + bY$) üst üste binmiş olup münasebet tamdır; $r = 0$ ise iki regresyon doğrusu birbirine diktir ve ele alınan hadiseler arasında ilişki yoktur denir.

İki değişken grubu arasındaki korelasyon katsayısı sıfıra çok yakın çıkarsa bu münasebet yoktur manâsına alınmamalıdır. Sadece doğrusal münasebetin olmadığı söylenebilir. Serpilme diyagramındaki noktaların umumî gidişine göre regresyon bağıntısının doğrusal olmadığı tesbit edilmişse bu takdirde regresyon eğrisi kullanılmalıdır. Bu bağıntıdan hesaplanan korelasyon ölçüsü korelasyon endeksi olarak tanımlanır ve

$$I^2 = 1 - \frac{\Sigma (Y - Y')^2}{\Sigma (Y - \bar{Y})^2}$$

formülüyle belirir.

Eğer $Y = a + bX + cX^2$ bir regresyon eğrisi ise a, b ve c parametreleri

$$\Sigma Y = na + b \Sigma X + c \Sigma X^2$$

$$\Sigma YX = a \Sigma X + b \Sigma X^2 + c \Sigma X^3$$

$$\Sigma YX^2 = a \Sigma X^2 + b \Sigma X^3 + c \Sigma X^4$$

denklemlerinin birlikte çözümü ile bulunacak, elde edilen kıymetler fonksiyonda yerlerine konmak suretiyle her X için Y' teorik değerlerinin hesaplanmasına girişilecektir.

Biz bundan sonra, aksi tasrih edilmedikçe korelasyon katsayısını doğrusal korelasyon katsayısı anlamında kullanacağız. Bu değerin hesabına dayanan değişik karekterdeki formüllerden bazıları aşağıda verilmiştir.

$$r = \frac{\frac{1}{n} \Sigma XY - \bar{X}\bar{Y}}{\sigma_X \cdot \sigma_y}$$

$$r = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\sigma_X \cdot \sigma_y}$$

$$r = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\sqrt{(\Sigma X^2 - n\bar{X}^2)\ (\Sigma Y^2 - n\bar{Y}^2)}}$$

$$r^2 = \frac{\Sigma Y'^2 - n\bar{Y}^2}{\Sigma Y^2 - n\bar{Y}^2}$$

$$r^2 = b_{yX} \cdot b_{Xy}$$

Regresyon ve korelasyon katsayılarının değeri hiç bir şekilde orijin değişikliğine bağlı değildir. Bu sebepten hesaplarda daha çok asli değerler yerine $x = X - \bar{X}$; $y = Y - \bar{Y}$ şeklinde bulunan değerler alınmakta ve neticede regresyon parametreleri

$$b_{yX} = \frac{\Sigma (X - \bar{X})\ (Y - \bar{Y})}{\Sigma (X - \bar{X})^2}$$ veya

$$b_{yX} = \frac{\Sigma xy - n\bar{y}\,\bar{x}}{\Sigma x^2 - n\,\bar{x}^2}$$ ile

$$a = \bar{Y} - b \cdot \bar{X}$$

den süratle hesaplanmaktadır. $\bar{X}$ ve $\bar{Y}$ değerlerinin küsüratlı değerler olmaları halinde farkların hesabı yorucu işlemleri gerektirdiğinden, $\bar{X}$ ve $\bar{Y}$ yerine A ve B gibi herhangi sabit iki değerde kullanılabilir. Hatta X ve Y serilerinin sabit sınıf fasılaları c_X ve c_y ise

$$u_X = \frac{X - A}{c_X} \; ; \; u_y = \frac{Y - B}{c_y}$$

değerleri hesaplara esas alınmalıdır. Münasebetin derecesi tayin edildikten sonra regresyon doğruları

$$\bar{X} = A + c_x \cdot \bar{u}_x$$

$$\bar{Y} = B + c_y \cdot \bar{u}_y \quad \text{kıymetlerinin}$$

$$Y - \bar{Y} = r \cdot \frac{\sigma_y}{\sigma_x} (X - \bar{X})$$

$$X - \bar{X} = r \cdot \frac{\sigma_x}{\sigma_y} (Y - \bar{Y}) \quad \text{denklemlerinde yerlerine}$$

konulmaları ile bulunacaktır.

II.

Korelasyon katsayısının mutlak değeri grubun değişme istidadına bağlı olduğundan, mukayese ancak aynı değişme istidadı gösteren gruplar arasında yapılmalıdır. Bu sebepten ait olduğu grubu belirtmeksizin iki değişken arasındaki korelasyondan bahsetmek manâsızdır. Ayrıca bu katsayının mutlak değerine bakarak münasebeti yüksek, orta veya zayıf diye tasrih etmek hatalı sonuçlar verebilir. Meselâ biri bedenî diğeri zihnî iki nitelik arasında % 50 korelasyona çok az rastlanır. r = 0.60 çıkınca aslında münasebet için orta demek lâzımken bu konuda çok yüksek bir ilişkinin, mucizenin mevcut olduğu söylenebilir.

Korelasyon katsayısının manâlandırılmasında düşülen ikinci ciddi hata bu iki nitelik arasındaki ilişkiye dayanarak bunlardan birinin diğerinin sebebi olduğu sonucunu çıkartmaktır. Bazen bu doğru olabilir. Ancak hangisinin sebep hangisinin netice olduğu çıkarılamaz. Aksine hiç bir sebep netice bağıntısı olmayan iki hadise arasında korelasyonun varlığı bulunabilir. Bu korelasyon her iki hadise ile ilgili bir üçüncü etkene bağlanabilir. Kısacası, korelasyon ancak bir sebep netice bağıntısının varlığını hatıra getirebilir, fakat bu bağıntının varlığını ispat etmez. Meselâ, Yule 1806 - 1911 yılları arasında İngilteredeki ölüm oranı ile Anglikan kilisesindeki evlenmeler arasında + 0.95'e varan çok yüksek bir münasebet bulmuştur. Daha uzun ömür için Anglikan kilisesi dışında evlenme tavsiyesi gülünç bir sonuç olmalıdır.

X ve Y serilerinin gruplanmış olduğu problemler için tatbikatta genellikle korelasyon tablosu kullanılmakta, böylece bir tablonun tanzimi halinde basit doğrusal korelasyon katsayısı

$$r = \frac{n \Sigma f . u_x . u_y - (\Sigma f_x . u_x)(\Sigma f_y . u_y)}{\sqrt{[n . \Sigma f_x . u_x^2 - (\Sigma f_x . u_x)^2][n \Sigma f_y . u_y^2 - (\Sigma f_y . u_y)^2]}}$$

ile süratle bulunabilmektedir. Bu konuda teferruatlı bir misal aşağıda verilmiştir.

$$r = \frac{(248)(611) - (48)(164)}{\sqrt{[248 . 734 - (48)^2][248 . 720 - (164)^2]}} = 0.87$$

Regresyon denklemlerine gelince

$$\bar{Y} = \frac{164}{248} X\ 5 + 147.50 = 150.80$$

$$\bar{X} = \frac{48}{248} X\ 5 + 57.50 = 67.00 \text{ dır ve}$$

$$\sigma_y^2 = \frac{\Sigma f_y . u_y^2}{\Sigma f_y} - \bar{u}_y^2 \quad \text{olduğundan}$$

$$\sigma_y^2 = \frac{720}{248} - (\frac{164}{248})^2 \qquad \sigma_y = 1.5705 \text{ ve}$$

$$\sigma_X^2 = \frac{734}{248} - (\frac{48}{248})^2 \qquad \sigma_x = 1.7095$$

bulunmakta, bu değerler

$$Y - \bar{Y} = r . \frac{\sigma_y}{\sigma_x} (X - \bar{X}) \quad \text{de yerine konmak suretiyle}$$

$$Y - 150.80 = 0.87 \cdot \frac{1.5705}{1.7095} (X - 67.00)$$

$$Y = 97.27 + 0.799\ X$$

elde edilmektedir. Benzer şekilde

$$X - 67.00 = 0.87 \cdot \frac{1.7095}{1.5705} \cdot (Y. - 150.80)$$

$$X = -75.80 + 0.947\ Y$$

yazılabilir.

Buraya kadarki tahlillerimiz, serideki dalgalanmaların bir tek sebebe bağlı olduğu faraziyesine dayandırılmıştır. Çoğu kere ilişki, bir bağlı değişkenle iki veya daha fazla bağımsız değişken arasında «çoklu korelasyon» veya bağımsız değişkenle bağlı değişken arasında diğer bağımsız değişkenlerin etkisinden arıtılmış şekilde «kısmî korelasyon» hesaplanabilir. Bu takdirde yine müşahede değerlerinden bir regresyon eğrisi veya yüzeyi geçirilecek, bunu takiben değişkenler arasındaki bağlılık izah edilmiş variyansın toplam variyansa nisbeti yardımiyle ölçülecektir.

Tablo 1.

19 - 16 Yaşları arasındaki 248 öğrencinin Boy ve Ağırlıkları itibariyle bölünmeleri

Ağırlık (kg)

Boy (cm)	40—45	45—50	50—55	55—60	60—65	65—70	70—75	75—80	80—85
130—135	—	1	—	—	—	—	—	—	—
135—140	4	12	4	—	—	—	—	—	—
140—145	—	22	21	3	—	—	—	—	—
145—150	—	6	28	14	1	1	—	—	—
150—155	—	1	1	24	17	4	—	—	1
155—160	—	—	—	4	22	18	2	3	—
160—165	—	—	—	—	1	14	10	4	—
165—170	—	—	—	—	—	2	2	1	—
170—175	—	—	—	—	—	—	—	—	—

Tablo 2.
Gruplanmış seriler için korelasyon tablosu
Ağırlık (u_x)

Boy (u_y)	— 3	— 2	— 1	0	1	2	3	4	5	f_y	$f_y . u_y$	$f_y . u_y^2$
— 3	—	1	—	—	—	—	—	—	—	1	— 3	9
— 2	4	12	4	—	—	—	—	—	—	20	—40	80
— 1	—	22	21	3	—	—	—	—	—	46	—46	46
0	—	6	28	14	1	1	—	—	—	50	0	0
1	—	1	1	24	17	4	—	—	1	48	48	48
2	—	—	—	4	22	18	2	3	—	49	98	196
3	—	—	—	—	1	14	10	4	—	29	87	261
4	—	—	—	—	—	2	2	1	—	5	20	80
f_x	4	42	54	45	41	39	14	8	1	248	164	720
$f_x . u_x$	—12	—84	—54	0	41	78	42	32	5	48		
$f_x . u_x^2$	36	168	54	0	41	156	126	128	25	734		
$f . u_x . u_y$	24	96	28	0	64	180	126	88	5	611		