Uludağ Üniversitesi
İktisat ve İdari Bilimler Dergisi
Cilt IV, Sayı 2, Kasım 1983

KORELASYON KATSAYILARININ GENELLEŞTİRİLMESİ

Mustafa AYTAÇ*

I. GİRİŞ

Değişkenler arasındaki ilişkiyi incelemek bilimin en önemli konularından birisidir. İstatistik kuramı ise değişkenler arası ilişkiyi incelemede bazı teknikler geliştirerek bunu anlamlı bir duruma getirmiştir. Değişkenler arasındaki ilişki değişik biçimlerde olabilir. Bu, doğrusal bir ilişki olabildiği kadar, doğrusal da olmayabilir.

Korelasyon genel anlamı ile iki veya daha çok sayıda değişken arasındaki ilişkiyi gösterir ve bu ilişkinin miktarını bir sayı ile belirtir. Bu sayıya korelasyon katsayısı adı verilir.

Ayrıca değişkenler arasındaki ilişkiyi saptarken kullanacağımız teknik, ilişkinin biçimine göre değiştiği gibi, aralarında ilişki bulunacak değişken sayısına ve ayrıca değişkenlerin sürekli veya süreksiz oluşlarına göre de değişebilir. İki değişken arasındaki ilişkiyi saptamak için kullanılan korelasyon tekniklerine basit korelasyon teknikleri denir. Aralarında ilişki aranarak değişken sayısı üç veya daha çoksa, bu durumda kullanılabilecek tekniklere de Kısmi Korelasyon Teknikleri adı verilir. Biz burada basit korelasyon tekniklerinden parametrik ve parametrik olmayan bazı korelasyon katsayılarının genel bir tanımdan nasıl elde edildikleri ve değerlendirmelerinin ne şekilde yapılması gerektiği üzerinde duracağız.

Fakat, daha önce çok kısa olarak incelememiz dışında bıraktığımız bazı korelasyon katsayılarını ele alalım.

Olağanlık katsayısı C, iki küme nitelik arasındaki ilişki derecesini ölçer. Özellikle niteliklerden en az birisi sınıflayıcı ölçek ile elde edilmişse, olağanlık katsayısı büyük faydalar sağlar[1].

Birbirini kapsamayan, iki sınıfa veya gruba ayrılmış iki değişken grubu arasındaki doğrusal korelasyon katsayısının tahmininde ise Tetrahoric korelasyon katsayısı kullanılır[2].

---

* *Yrd.Doç.Dr.; Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi.*

1 Siegel, S., (Terc: Topsever, Yurdal), Nonparametric Statistics for Behavioral Sciences, Mc Graw Hill, New York, 1956, s. 216-222.

2 M.Kemal Yoğurtçugil, Ki-Kare Üzerine Bir Deneme, İst.Üniv. İkt. Fak. İstanbul, 1978, s. 90.

İki değişken arasındaki ilişkinin doğrusal olup olmadığını test etmek için ise korelasyon oranından yararlanılır[3].

Bir sürekli değişken ile aslında sürekli olduğu halde yapay olarak süreksiz duruma getirilen başka bir değişken arasındaki doğrusal ilişki miktarını saptama Çift Serili Korelasyon Tekniği; bir sürekli ile bir de gerçek süreksiz iki değişken varsa, değişkenler arasındaki doğrusal ilişkiyi bulmak için de Nokta Serili Korelasyon Tekniğinden yararlanılır[4].

Bağlılık analizi ile değişkenler arasındaki ilişkinin ölçüsü veya derecesini anlatacak bir ölçmeyi hesaplamak isteriz. Bu tip ölçmeler örneklem verilerinden hesaplanır ve üç amaca hizmet eder. i) örneklem gözlemleri arasındaki ilişkilerin kuvvetini ölçerler. ii) yığındaki raslantı değişkenleri arasındaki ilişkinin kuvvetini ölçmek için bir nokta tahmini içerirler. iii) yığındaki raslantı değişkenleri arasındaki ilişkinin kuvvetini ölçmek için bir güven aralığı oluşturulmasına yardım ederler[5].

Bunun yanında herhangi iki değişken arasındaki (X ve Y diyelim) korelasyon ölçüsü aşağıda sıralanan koşulları sağlar[6].

i) Korelasyon ölçüsünün yalnızca -1 ile $+1$ arasındaki değerleri aldığı varsayılmıştır.

ii) X in büyük değerleri ile Y nin büyük değerleri eş ise (ve X in küçük değerleri ile Y nin küçük değerleri eş ise) korelasyon ölçüsü artı değerde olup bir değerine yaklaşmalıdır. Bu gibi durumlarda X ve Y arasında doğrudan bir ilişkinin olduğunu söyleriz.

iii) X in küçük değerleri Y nin büyük değerleri ile (ve tersi) eş ise, korelasyon ölçüsünün aksi değerde olduğunu ve -1 değerine yaklaştığını söyleriz. Bu, ters bir ilişkinin var olduğunu simgeler.

iv) X in büyük değerleri, Y nin çok büyük değerleri yerine küçük değerleri ile eşleşmiş ise korelasyon ölçüsü sıfıra yaklaşır. X ve Y değişkenleri bağımsız iseler, korelasyon ölçüsü sıfır olmalıdır. Bu durumda X ve Y nin birbirleri ile bağımlı değil bağımsız olduğunu söyleriz. X ve Y bağımsız ise korelasyon sıfırdır, fakat korelasyonun sıfır çıkması onların bağımsız olduklarını önermez. Yalnızca aralarında doğrusal bir ilişki olmadığını belirtir.

II. GENEL KORELASYON KATSAYISI

Bu bölümde inceleyeceğimiz "Pearson Momentler Çarpımı Korelasyon Katsayısı" "Spearman Sıra Farkları Korelasyon Katsayısı" ve "Kendall Sıra Farkları Korelasyon Katsayısı"nın üçüde genel Korelasyon Katsayısının özel birer durumudurlar.

---

3 Bakınız Y. a.g.e., Bölüm VI.

4 Hüsnü Arıcı, İstatistik Yöntemler ve Uygulama, Hacettepe Üniv. Ankara, 1972, s. 127 ve 130.

5 Wayne W. Daniel, Applied Nonparametric Statistics Houghton Mifflin Comp. Boston, 1978, s. 299.

6 Conover, W.J., Pratical Nonparametric Statistics, John Wiley and Sons, New York, 1980, s. 250.

X ve Y ile gösterilen iki nitelikle ilişkili olarak n tane nesnenin bir kümesine sahip olduğumuzu varsayalım. Herhangi bir sırada nesneleri 1 den n'e kadar numaralıyalım. Yani, X e göre $x_1, x_2, x_3, \ldots\ldots x_n$ ve Y ye göre de $y_1, y_2, y_3, \ldots\ldots, y_n$ olsun. Bu değerlerde sıralama değişiklikleri olabilir.

Herhangi bir i.inci gözlem çifti için, a_{ij} tarafından gösterilen x-sıralama değerini tahsis edeceğiz. a_{ij} de indisler yer değiştirirse a_{ij}, eksi değerini alacaktır. Yani $a_{ij} = - a_{ji}$. Benzer şekilde b_{ij} ile gösterilen y-sıralama değerini de ayıracağız. Aynı zamanda $b_{ij} = - b_{ji}$. 1 den n'e kadar i ve j nin tüm değerleri üstüne, genel bir korelasyon katsayısı T tanımlıyabiliriz.

$$T = \frac{\sum_{i=1}^{n} a_{ij} b_{ij}}{\left(\sum_{i=1}^{n} a_{ij}^2 \sum_{i=1}^{n} b_{ij}^2\right)^{1/2}} \qquad (1)$$

Yukardaki eşitlikte i = j ise a_{ij}^2 sıfır olarak değerlendirilir[7].

(1) formülünün -1 ile $+1$ arasında değerler alabileceğini Cauchy-Schwentz eşitsizliği ile gösterilebilinir.

III. GENEL KORELASYON KATSAYISININ ÖZEL BİR HALİ OLARAK PEARSON MOMENTLER ÇARPIMI KORELASYON KATSAYISI

Bu korelasyon katsayısı adını İngiliz istatistikçisi Karl Pearson'dan almaktadır. ρ olarak gösterilen bu katsayı, korelasyon katsayıları içinde en çok kullanılan kavram olup parametrik bir yöntemdir. Çoğu kez bunun örneklemden hesaplanan tahmini olan r kullanılır. Ancak bunun kullanılabilmesi için, örneklem yığınının iki boyutlu normal olması gerekir. Bu koşul sağlanamaz ise bu korelasyon katsayısı kullanılamaz.

Ayrıca, yığın korelasyon katsayısı ρ nun anlamlılığını da — ρ nun aldığı değişik değerlere ve örneklem büyüklüğüne göre — test edebiliriz[8].

Bu korelasyon katsayısı (1) formülünden şu şekilde bulunabilir: a_{ij} ve b_{ij} yi

$$a_{ij} = x_j - x_i$$

$$b_{ij} = y_j - y_i$$

olarak tanımlıyalım. Bu takdirde,

$$\frac{1}{2} \sum_{i,j} (x_j - x_i)(y_j - y_i) = n \sum_{i,j} x_i y_i - \sum x_i y_j = n \text{ Kov } (x, y)$$

$$\frac{1}{2} \sum_{i,j} (x_j x_i)^2 = n \sum x_i^2 - (\sum x_i)^2$$

---

7 Maurci, G. Kendall, Rank Correlation Methods, Hafner Publishing Comp, Third Edition, New York, 1962, s. 19.

8 M. Kemal Yoğurtçugil, Örnekleme, İstanbul, 1978, s. 122-130.

$$= n\,V(x)$$

$$\frac{1}{2}\sum_{i,j}(y_j - y_i) = n\,V(y)$$

Bulduğumuz değerleri (1) numaralı formülde yerine koyarsak,

$$\rho = \frac{\text{Kov}(X, y)}{\sqrt{V(X)\,V(Y)}}$$

değerini elde ederiz ki, bu da Pearson momentler çarpımı korelasyon katsayısından başka bir şey değildir[9]. Bu katsayıyı

$$\rho = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{[\sum_{1}^{n}(X_i - \bar{X})^2 \sum_{1}^{n}(Y_i - \bar{Y})^2]^{1/2}}$$

ya da

$$\rho = \frac{\Sigma X_i Y_i - n\bar{X}\bar{Y}}{(\Sigma X_i^2 - n\bar{X}^2)^{1/2}(\Sigma Y_i - n\bar{Y}^2)^{1/2}}$$

olarak formüle edebiliriz.

Görüldüğü gibi bu korelasyon katsayısını bulabilmemiz için dağılımın ne şekilde olduğunu bilmemiz gerekmektedir.

IV. GENEL KORELASYON KATSAYISININ ÖZEL BİR HALİ OLARAK SPEARMAN SIRA FARKLARI KORELASYON KATSAYISI

1904 yılında Spearman tarafından geliştirilen bu korelasyon katsayısı, en yaygın şekilde kullanılan parametrik olmayan korelasyon katsayısıdır. Özellikle gözlem değerleri sıralayıcı bir ölçek ile elde edilmişse çok iyi sonuçlar verir. Bazen "rho" olarak da bilinen bu katsayı "ρ_s" ile simgelenir.

Spearman korelasyon katsayısı, genel korelasyon katsayısından şu şekilde elde edilir: + 1 ve − 1 in yerine, q_i nin y-niteliğine göre i.inci üyesinin sıralaması olduğu yerlerde

$$a_{ij} = p_j - p_i$$

$$b_{ij} = q_j - q_i$$

yazalım[10]. p_i ve q_i değerleri 1 den n'e kadar olan değerleri alır. Bundan dolayı

$$\Sigma(p_j - p_i) = \Sigma(q_j - q_i)$$

9 Maurice, G. Kendall, a.g.e., s. 21.
10 Maurice, G. Kendall, a.g.e., s. 20-21.

O zaman (1) ile gösterdiğimiz formül

$$T = \frac{\Sigma (p_j - p_i)(q_j - q_i)}{\Sigma (p_j - p_i)^2}$$

olur.

$$\sum_{i,j=1}^{n} (p_j - p_i)(q_j - q_i) = 2n \sum_{1}^{n} p_i q_i - 2 \Sigma P_i \Sigma q_j \qquad (2)$$

Σp_i ve Σq_j, ilk n tane doğal sayısının toplamları olduğu için herbirisi $\frac{1}{2} n(n+1)$ değerine sahiptir. Dolayısı ile, yukardaki eşitliğin sağ tarafı

$$2n \, \Sigma p_i q_i - \frac{1}{2} n^2 (n+1)^2$$

şeklini alır.

Aynı zamanda,

$$T = \sum_{1}^{n} (p_i - q_i)^2 = 2 \Sigma P_i^2 - 2 \Sigma p_i q_i$$

değerine de sahibiz. Buradan $2\Sigma p_i q_i$'yi çekip (2) numaralı eşitlikte yerine koyarsak

$$\Sigma (p_j - p_i)(q_j - q_i) = 2n \Sigma p_i^2 - \frac{1}{2} n^2 (n+1)^2 - n\,T$$

olur. Σp_i^2, ilk n doğal sayısının karelerinin toplamı olduğu için $\frac{1}{6} n(n+1)(2n+1)$ değerine eşittir ve

$$\Sigma (p_j - p_i)(q_j - q_i) = \frac{1}{6} n^2 (n^2 - 1) - nT$$

Bu takdirde de,

$$\begin{aligned} \sum_{1}^{n} (p_j - q_i)^2 &= 2n\, p_i^2 - 2\Sigma p_i q_i \\ &= 2n \Sigma p_i^2 - 2(\Sigma p_i)^2 \\ &= \frac{1}{6} n^2 (n^2 - 1) \end{aligned}$$ **olur.**

Bulduğumuz değerleri (1) formülünde yerine koyarsak

$$T = \rho_s = \frac{\frac{1}{6} n^2 (n^2 - 1) - nT}{\frac{1}{6} n^2 (n^2 - 1)} = 1 - \frac{6\,T}{n(n^2 - 1)}$$

Örneğin gözlem çiftlerimiz (X_i, Y_i) lerin değerleri (10,13), (5,15), (9,10) olsun. Bu durumda

$$R(X_i) = 1 \quad 2 \quad 3$$

$$R(Y_i) = 2 \quad 3 \quad 1$$

$$T = \sum_{1}^{4} [R(X_i) - R(Y_i)]^2 = (-1)^2 + (-1)^2 + (2)^2 = 6$$

$$\rho_s = 1 - \frac{6.6}{3.8} = 1 - \frac{36}{24} = -0.5$$

Görüldüğü gibi, gerek X_i gerekse Y_i ler içinde birbirine eşit değerler yoktur. Eğer kendi bünyelerinde eşit değerler — (10,13), (5,13) ve (10,15) olsa idi — varsa, ρ_s şöyle bulunur[11].

$$\rho_s = \frac{\Sigma x^2 + \Sigma y^2 - \Sigma T}{2(\Sigma x \, \Sigma y)\, Y^2}$$

Formülde,

$$x^2 = \frac{n(n^2-1)}{12} - \Sigma T_x$$

$$y^2 = \frac{n(n^2-1)}{12} - \Sigma T_y$$

$$T_x = \frac{t_x^3 - t_x}{12}$$

$$T_y = \frac{t_y^3 - t_y}{12}$$

olup t_x ve t_y sırası ile X ve Y nin kendi içlerindeki eşit gözlem değerlerinin sayısını vermektedir.

ρ da olduğu gibi ρ_s inde anlamlılık testini yapabilmek için yöntemler geliştirilmiştir[12].

Genel korelasyon katsayısına bağlı olmayan bir tane daha Spearman korelasyon katsayısı vardır[13]. Buna "Spearman Kuralı" (Spearman's footrule) denir. Bu değer T'ye bağlı olmayıp, $|R(X_i) - R(Y_i)|$ değerine bağlıdır. Başka bir deyişle sıralama farklarının mutlak değerine bağlıdır. Ve şöyle tanımlanır.

---

11 Wayne, W. Daniel, a.g.e. s. 303-304.

12 Daha fazla bilgi için bakınız.
Wayne W. Daniel, a.g.e., s. 300-306.
W.J. Conover, a.g.e., s. 252-256.
Jean D. Gibbons, Nonparametric Methods for Quantitative Analysis. Holt-Rine Hart-Winston Publ., New York, 1976, s. 276-284.

13 Maurice, G. Kendall, a.g.e. s. 32-33.

$$R = 1 - \frac{3|R(X_i) - R(Y_i)|}{n^2 - 1}$$

Görüldüğü gibi, iki sıralama aynı değerlere sahipse R = 1 olur. Ve n = 2 olmadıkça, R, — 1 değerine sahip olamaz.

V. GENEL KORELASYON KATSAYISININ ÖZEL BİR HALİ OLARAK KENDALL SIRA FARKLARI KORELASYON KATSAYISI

Kruskal 1899 yılında bu şekildeki bir korelasyon katsayısı üzerinde çalışmış; fakat onu kullanılabilir hale getiren kişi 1938 yılında Kendall olmuştur[14]. Bu korelasyon katsayısını göstermek için t, T veya τ simgeleri kullanılır. Fakat en çok tercih edileni τ dur. Bu korelasyon katsayısına Kendall's tau adı da verilmektedir.

τ yu (1) formülünden şu şekilde elde edebiliriz. p_i, x niteliğine göre i.inci elemanın sıralaması olduğu yerlerde $P_j > P_i$ ise + 1 sıralama değerini; $P_j < P_i$ ise — 1 sıralama değerini verdiğimizi düşünelim. O zaman

$$a_{ij} = \begin{cases} +1 \ , & P_i < P_j \text{ için} \\ -1 \ , & P_i > P_j \text{ için} \end{cases}$$

Aynı şekilde b içinde bu değerleri oluşturursak,

$$b_{ij} = \begin{cases} +1 \ , & P_i < P_j \text{ için} \\ -1 \ , & P_i > P_j \text{ için} \end{cases}$$

$\Sigma\, a_{ij}.b_{ij} = 2S$ dir. Çünkü elde ettiğimiz gözlem değer çiftlerinin herbirisini (i,j) — doğal sırada olmak üzere — ve (j,i) — ters doğal sırada olmak üzere — iki defa toplamını alıyoruz. Bundan başka Σa_{ij}^2 ve Σb_{ij}^2, a_{ij} ve b_{ij} terimlerinin sayıları olup her ikisi de n(n — 1)'e eşittirler. Bulduğumuz bu değerleri (1) formülünde yerine koyarsak,

$$T = \frac{\Sigma\, a_{ij}\, b_{ij}}{\sqrt{(\Sigma a_{ij}^2)\,(\Sigma b_{ij}^2)}}$$

$$= \frac{2S}{\sqrt{[n(n-1)]\,[n(n-1)]}}$$

---

14 Kruskal, W.H., "Ordinal measures of association", Journ. Amer. Statist. Ass. 1958, Vol. 53, s. 814-861. Bu makalede, τ nun gelişmesinde yapılan bağımsız çalışmalar ile bağımsızlık üzerinde varolabilecek genel sorunlar ele alınarak incelenmiştir.

$$= \frac{2S}{n(n-1)}$$

değerini elde ediriz.

τ, verilerin en azından sıralayıcı bir ölçek düzeyinde ölçülmeleri durumunda kullanılır. Ayrıca sıfır hipotezi altında τ nun tahmini olan, τ nun örnekleme dağılımı bilindiği için anlamlılık testi de yapılabilir[15].

Eş gözlem değerleri çok fazla olduğu zaman, τ yu bulabilmek çok zaman alır ve o kadar da sıkıcı olmaktadır. Bunun önüne geçmek için bazı çalışmalar yapılmıştır.

Bunlardan birincisi bilgisayarın getirdiği olanaklardan yararlanarak τ yu hesaplama tekniğidir. Knight, makalesinde τ yu bulabilmek için bir akış şeması verir[16]. Böylece, τ nun bulunması aşamasında büyük bir kolaylık sağlanır.

İkincisi ise τ yu grafik olarak bulmaktır. Griffin, nasıl bulunacağını grafik yöntemi ile açıklamış ve Shah'da ona bazı katkılar yapmıştır. Gerek grafik gerekse bilgisayar yöntemi ile τ nun bulunması durumunda değer olarak bir farklılık olmamaktadır[17].

Bilindiği gibi, regresyon katsayısı normal dağılım koşulların sağlanamadığı durumlarda çok büyük hatalara neden olabilmektedir. Bu gibi durumlarda τ ya dayanarak β nın etkin ve o derecede de basit bir tahminin elde edilmesine çalışılmıştır. Gerek β gerekse diğer regresyon parametrelerinin tahmin edilmesinde Sen ve Aidichie bir yöntem geliştirmişlerdir[18].

Kuramsal olarak değişkenler sürekli varsayılmasına rağmen, herbir gözlem değerlerinin kendi işlerinde eşit değerlere sahip olması sözkonusu olabilir. Bu durumda, Kendall's tau aşağıdaki formül yardımı ile hesaplanır:

$$\tau = \frac{S}{\left[\frac{n(n-1) - 2T_x}{2}\right]^{1/2} \left[\frac{n(n-1) - 2T_y}{2}\right]^{1/2}}$$

Formülde;

$$T_x = \frac{1}{2} \Sigma\, t_x(t_x - 1)$$

---

15 Daha fazla bilgi için bakınız.
Wayne, W. Daniel, a.g.e., s. 306-315. Myles Hollander and Douglas A-Wolfe, Nonparametric Statistical Methods, John and Wiley, New York, 1973, s.185-190.

16 Knight, W.R., "A Computer Method for Calculating Kendall's tau with ungrauped data", Jour, Amer. Stat, Ass., 1966, Vol: 61, s. 436-439.

17 Griffin, H.D., "Graphic Computation of Tau as a Coefficient of disarray" Jour. Amer. Stat. Ass, 1958, Vol. 53, s. 441-447, Shah, S.M., A note on Griffin's Paper Jour. Amer. Stat. Ass., 1961, Vol: 56, s. 736.

18 Sen, P.K., "Estimates of the regression coefficient based on Kendall's tau", Jour. Amer. Stat. Ass., 1968, Vol: 63, s. 1379-1389. Aidichie, J.N., "Estimates of regression parameters based on rank test", Annals of Math. Statis., 1967, Vol: 38, s. 894-904.

$$T_y = \frac{1}{2} \Sigma t_y(t_y - 1)$$

t_x = verilen sıralamada birbirine eşit olan X gözlemlerinin sayısı
t_y = verilen sıralamada birbirine eşit olan Y gözlemlerinin sayısıdır.

Eşit gözlem değerlerinin sayısı 1 ile 10 arasında olduğu zaman, τ'nun tahmininin örnekleme dağılımı için ayrı bir tablo oluşturulmuştur[19].

VI. KORELASYON KATSAYILARININ NORMAL DAĞILIM İLE BAĞLANTILARI

Bu korelasyon katsayılarının örneklemden hesaplanan tahminleri, örneklem büyüklüğü 30 dan büyük olduğu zaman, belli tanımlamalarla standart normal dağılıma yaklaştığı gösterilebilir. Bunu yapmak aynı zamanda zorunludur; çünkü örneklem büyüklüğünün 30'u geçmesi halinde tablo kritik değerleri oluşturulmamıştır. Bunun yerine normal dağılım tablosu kullanılır.

i) ρ nun Normal Dağılıma Yaklaşımı:

ρ nun tahminini r olarak gösterelim. O zaman $E(r) = r$ ve $V(r) = (1 - r^2)^2 / (n - 2)$'dir. n çok büyük ise

$$Z = \frac{r - R}{\sigma_r}$$

dönüşümü ile Z'nin normal dağılıma yaklaşır.

ii) ρ_s'un Normal Dağılıma Yaklaşımı:

ρ_s'in tahmini de r_s olsun. $V(r_s) = 1/(n-1)$ değerine sahiptir. $n > 30$ olduğu durumlarda aşağıdaki gibi tanımlanan Z,

$$Z = r_s\,(n - 1)^{1/2}$$

normal dağılıma yaklaşır.

Ayrıca örneklemden bulunan r_s değerinin varyansının alabileceği en büyük değer $\frac{3}{n}(1 - \rho_s^2)$'dir[20]. Yani,

$$V(r_s) \leqslant \frac{3}{n}(1 - \rho_s^2)$$

iii) τ nun Normal Dağılıma Yaklaşımı:

Örneklem büyüklüğü arttığı zaman aşağıda tanımlanan Z, standart normal dağılıma yaklaşır.

19 Daha fazla bilgi için bakınız.
Silverstone, H., "A note on the cumulants of Kendall's distribution", Biometrica, 1950, Vol: 37, s. 231-235.
20 Maurice G.Kendall, a.g.e., s. 63.

$$Z = \frac{3\hat{\tau}\,[n(n-1)]^{1/2}}{[2(2n+5)]^{1/2}} = \frac{S}{[n(n+1)(2n+5)/18]^{1/2}}$$

Z, büyük örneklem büyüklüklerinde olduğu kadar, eşit gözlem değerleri varolduğu zamanda başarı ile kullanılabilir.

τ nun varyansının alabileceği en büyük değer de şu şekilde sınırlanmıştır.

$$V(\tau) \leqslant \frac{2}{n}(1-\tau^2)$$

VII. KORELASYON KATSAYILARI ARASINDAKİ BAĞINTILAR

Hernekadar açıkladığımız üç korelasyon katsayısını, genel bir korelasyon katsayısından elde ettiğimizi belirttiysek de, kendi aralarındaki ilişkilere değinmedik. Özellikle ρ_s ve τ arasındaki ilişkiyi gösteren iki tane önemli eşitsizlik vardır.

i) Daniel Eşitsizliği:

n sıralamadaki sayımız olsun. O zaman,

$$-1 \leqslant \frac{3(n+2)}{n-2}\tau - \frac{2(n+1)}{n-2}\rho_s \leqslant 1$$

olur. n'nin büyük değerleri için bu eşitsizlik

$$-1 \leqslant 3\tau - 2\rho_s \leqslant 1$$

değerini alır[21]. τ sıfırdan büyük olduğu zaman, alt limit değil ama üst limit elde edilebilir. τ sıfırdan küçük olduğu zaman üst limit belirlenemez, fakat alt limit belirlenebilir. $\tau = 0$ durumunda ise, iki limit de hesaplanabilir.

ii) Durbin-Stuart Eşitsizliği:

Bu eşitsizlik belli bir τ için ρ_s'in, alt ve üst limit değerlerinin bulunmasında kullanılır[22].

$\tau \geqslant 0$ olduğu durumlarda, verilen belli bir τ için ρ_s'in üst limiti şöyledir:

$$\rho_s \leqslant 1 - \frac{1-\tau}{2(n+1)} \quad (n-1)(n-2)+4$$

Alt limit ise $\tau \geqslant 0$ durumunda şöyledir:

$$\rho_s \geqslant \frac{3n\tau - (n-2)}{2(n+1)}$$

21 İspatı için bakınız. Maurice G. Kendall, y.a.g.e., s. 28-29.
22 İspatı için bakınız. Maurice G. Kendall, y.a.g.e., s. 29-30.

$\tau \geqslant 0$ ve n'nin çok büyük olduğu durumlarda bunlar aşağıdaki şekle dönüşürler.

$$\frac{3}{2}\tau - \frac{1}{2} \leqslant \rho_s \leqslant \frac{1}{2} + \tau - \frac{1}{2}\tau^2$$

Kendall's tam korelasyon katsayısının sıfırdan küçük durumları için bu limit değerleri şöyledir:

$$\frac{1}{2}\tau^2 + \tau - \frac{1}{2} \leqslant \rho_s \leqslant \frac{3}{2}\tau + \frac{1}{2}$$

VIII. KORELASYON KATSAYILARININ KARŞILAŞTIRILMASI

İncelediğimiz korelasyon katsayılarından ρ parametrik bir yöntem olmasına karşılık, ρ_S ve τ parametrik olmayan yöntemlerdir.

ρ'nun kullanılabilmesi verilerin aralıkla bir ölçekle ölçülmesine bağlı olduğu kadar, bazı önemli varsayımlara da bağlıdır (Normal dağılım olması gibi).

ρ_S ve τ, gerek sınıflayıcı gerekse sıralayıcı veriler için uygun olup, elde edilişleri kolaydır. Gözlem değerlerinin alındığı yığınların dağılımı üzerinde herhangi bir varsayımda bulunmaması ve küçük örneklem büyüklüklerinde bile başarı ile ilişkileri ortaya koyması bakımından, parametrik olmayan korelasyon katsayıları günümüzde sık sık uygulanan yöntemler olmuşlardır.

r gözlem değerlerinin sıralamalarına bağlı olmadığı halde, r_S ve τ gözlem değerlerinin sıralanmalarından etkilenir. r_S ve τ aynı verilere uygulandığı zaman çoğu kez değişik sayısal veriler verirler. Bu doğaldır; çünkü her iki korelasyon katsayısı da değişik yöntemlerle sonuca ulaşır.

r ve τ, yığın korelasyon katsayılarının yansız bir tahmini iken; r_S yığın korelasyon katsayısının yanlı bir tahminidir.

Aralarındaki temel fark burada açıkca gözler önüne serilir. Yani,

$$E(r) = \rho$$

$$E(\tau) = \tau$$

fakat

$$E(r_S) \neq \rho_S$$

r_S ve τ'nın değişkenler sıralayıcı bir ölçek düzeyi ile ölçüldüğü zaman çok sık kullanılan iki korelasyon katsayısı olduğunu daha önce belirtmiştik. Yığındaki bir bağlantının var olup olmadığını test etmede r ye dayanan t-testine göre r_S ve τ nun ARE'si (Bağıl Asimtotik Etkinlik) 0.912'dir[23].

---

23 Bunların daha fazla ayrıntılı incelenmesi — diğer test yöntemleri — ni de içeren — aşağıdaki makalelerde ele alınmıştır.
Bhattacharyya, G.K., Johnson, R.A. and Neave, H.R., "Percentage points of non-parametrik tests for independence and empirical power comparisons", Jour. Amer. Statis. Ass., 1970, Vol: 67, s. 976-983. Farlie, D.J.G., "The asymptotic efficiency of Daniel's generalized correlation coefficient", Jour. Royal. Stat. Soc., 1961, Vol. B-23, s. 128-142.

Bir başka deyişle, X ve Y değişkenleri arasında varolan ilişkiyi, r, 912 gözlemin incelenmesi ile ortaya çıkarırken; r_s ve τ, aynı ilişkiyi ancak 1000 gözlem değerinin incelenmesi ile ortaya koyabilmektedir.

r_s ve τ, trend testleri içinde kullanılabilirler. Parametrik olmayan trend testi, Cox-Stuart trend testidir. Raslantı değişkenlerinin normal dağılıma sahip olduğu bilindiği zaman; regresyon katsayısına bağlı testin, Cox-Stuart testine göre ARE'si 0.78'dir. Koşullar aynı kaldığı sürece τ ve r_s in aynı trend testine göre ARE'si ise 0.98'dir. Fakat buna rağmen, değişik neuenierden dolayı τ ve r_s, trend testlerinde Cox-Stuart testi kadar yaygın bir kullanıma sahip değildirler[24].

Son olarak iki parametrik olmayan korelasyon katsayısından hangisinin tercih edileceği üzerinde duralım[25].

i) Örneklem büyüklükleri fazla ve hesaplamalar elde yapılıyorsa; r_s, τ ya tercih edilmelidir. Çünkü τ'i bulmak hem zor hem de can sıkıcı olur.

ii) τ'nun dağılımı, r_s den daha hızlı olarak normal dağılıma yaklaştığı için normal yaklaşım kullanılacağı zaman τ ele alınmalıdır.

iii) Genelde r_s ve τ aynı verilerden hesaplandığı zaman farklı sayısal değerlere sahip olurlar. Fakat buna rağmen oluşturulan hipotez testi için aynı sonucu verirler.

iv) τ, yansız bir tahmin olduğu için birçok araştırıcı tarafından tercih edilerek kullanılabilir.

YARARLANILAN KAYNAKLAR

Aidichie, J.N., Estimates of regression parameters based rank test, Ann. Math. Statis. 1967, Vol: 38, s. 894-904.

Arıcı, Hüsnü, İstatistik, Yöntemler ve Uygulama, Hacettepe Üniv., Ankara, 1972.

Bhattacharyya, G.K., Percentage points of non parametrik tests for Johnson, R.A. and Neave H.R. independence and empirical power comparison. JASA 1970, Vol: 67, s. 976-983.

Conever, W.J., Practial Nonparametric Statistics John Wiley and Sons, Newyork, 1980.

Daniel, Wayne, W., Applied Nonparametric Statistics, Houghton Mifflin Co., Boston 1978.

Farlie, D.J.G., The asyymptotic efficieniy of Daniels, generalized correlation coefficient, Jour Royal Stat. Soc., 1961, Vol: B-23, s. 128-142.

Gibbons, Jean, D., Nonparametric Methods for Quantative Analysis, Holt-Rinehart-Winston Publ., New York, 1976.

Griffin, H.D., Graphic computation of Tau as a coefficient of disorray, JASA, 1958 Vol. 53, s. 441-447.

Hollander, M. and Wolfe, D.A., Nonparametric Statistical Methods, John and Wiley, New York, 1973.

---

24 Stuart, A., "The efficiencies of test of randomness against normal regression", Jour. Amer. Stat. Ass., 1956, Vol: 51, s. 285-287.

25 Wayne, W. Daniel, a.g.e., s. 315-316.

Kendall, M.G., Rank Correlation Methods, Hafner Publishing Comp., Third Edition, New York, 1962.

Knight, W.R., Accomputer method for calculating Kendall's tau with ungrouped data, JASA, 1966, Vol: 61, s. 436-439.

Kruskal, W.H., Ordinal measures of association, JASA, 1958, Vol: 53, s. 814-861.

Sen, P.K., Estimates of the regression coefficent based on Kendall's tau, JASA 1968, Vol: 63, s. 1379-1389.

Shah, S.M., A note on Griffin's paper, JASA, 1961, Vol: 56, s. 736.

Siegel, S. (Çev. Topsever Yurdal), Nonparametric Statistics for Behavioral Sciences, Mc Graw Hill, New York, 1956.

Silverstone, H., Anote on the cumulants of Kendall's distribution Biometrica, 1950 Vol: 37, s. 231-235.

Stuart, A., The efficiencies of test of, randomness againsts normal regression, JASA, 1956, Vol: 51, s. 285-287.

Yoğurçugil, M. Kemal, Ki-Kare Üzerine Bir Deneme, İst. Üniv. İktisat Fak., İstanbul, 1978.

————, Örnekleme, İst. Üniv. İktisat Fak. İstanbul, 1978.