

AN ALTERNATIVE RANK CORRELATION COEFFICIENT WITH GRAPHICAL APPROXIMATION

(Review)

Yaprak Arzu ÖZDEMİR* Meltem EKİZ
G.Ü.Fen Edebiyat Fakültesi İstatistik Bölümü 06500 Beşevler, Ankara
e-mail: yaprak@gazi.edu.tr

ABSTRACT

In this study, a rank correlation coefficient (v) proposed by Blest(2000), which does not require normality assumption in hypothesis testing of correlation coefficient for interval level or ratio scaled measured variables, is introduced. This rank correlation coefficient, which is examined based on graphical presentation and their comments, has a structure, which considers the interchanges in the early ranking of the data. When the sample size is five this introduced v rank correlation coefficient is considered and it has showed that it is more advantageous compared to other rank correlation coefficients since it gives different results even if for small changes in ranking. Furthermore, the critical values of v correlation coefficient are obtained for different significance levels and these values are presented in a table.

Key Words: Ranks, Kendall's τ rank correlation coefficient, Spearman's ρ rank correlation coefficient.

GRAFİKSEL YAKLAŞIMLA ALTERNATİF BİR SIRA KORELASYON KATSAYISI

(Derleme)

ÖZET

Bu çalışmada, iki değişkenin eşit aralıklı yada oranlama düzeyinde ölçüldüğü durumda, korelasyon katsayısına ilişkin hipotez testlerinde normal dağılım varsayımına ihtiyaç duymayan Blest(2000)'in önerdiği sıra korelasyon katsayısı(v) tanıtılmıştır. Bu katsayı, grafiksel gösterime ve bu grafiğin yorumlanmasına dayalı olup, sıralamalardaki ilk yer değişikliklerini önemseyen bir yapıya sahiptir. Tanıtılan v sıra korelasyon katsayısı, örnek çapı 5 iken incelenmiş ve mevcut sıra korelasyon katsayıları ile karşılaştırılarak, sıralamalardaki küçük yer değişikliklerinde bile farklı sonuçlar vermesi sebebiyle, daha avantajlı olduğu gösterilmiştir. Ayrıca farklı anlamlılık düzeyleri için v katsayısına ilişkin kritik değerler bulunarak tablo halinde verilmiştir.

Anahtar Kelimeler: Sıra Sayıları, Kendall'in τ sıra korelasyon katsayısı, Spearman'ın ρ sıra korelasyon katsayısı

1. GİRİŞ

Eşit aralıklı yada oranlama düzeyinde ölçülen iki değişken arasındaki ilişkinin, hipotez testinde normal dağılım varsayımına ihtiyaç duyulmayan sıra korelasyon katsayıları Kendall'in Tau (τ) ve Spearman'ın Ro (ρ) dur(1). Bilindiği üzere, bu sıra korelasyon katsayılarının elde edilmesinde temel "sıralama" üzerinedir. Korelasyon katsayısının genelleştirilmiş biçimi ;

1. INTRODUCTION

The rank correlation coefficients which don't require normal distribution in hypothesis testing for correlation of two variables that are measured interval level or ratio scale are Kendall's Tau (τ) and Spearman's Rho (ρ) (1). It is known that, the basis of obtaining these rank correlation coefficients are on "ordering". If the generalized form of the correlation coefficient writing as,

$$\Gamma = \frac{\sum_{i,j} c_{ij} d_{ij}}{\sqrt{\sum_{i,j} c_{ij}^2 \sum_{i,j} d_{ij}^2}} \quad [1.1]$$

olarak yazıldığında c_{ij} ve d_{ij} nin tanımlamalarına bağlı olarak bu istatistik farklı korelasyon katsayılarına dönüşür(2). $i, j \in \{1, 2, \dots, n\}$ olmak üzere, X değişkeni bakımından gözlenen değerlere atanan sıra sayıları p_i , Y değişkeni bakımından gözlenen değerlere atanan sıra sayıları ise q_i olsun. Hem X hem de Y değişkeni bakımından aynı sıra değerli gözlemlerin olmadığı varsayımı altında, yani ortak sıra sayısı atanması söz konusu değilken;

$$c_{ij} = \begin{cases} 1 & p_i < p_j \\ -1 & p_i > p_j \end{cases}$$

ve

$$d_{ij} = \begin{cases} 1 & q_i < q_j \\ -1 & q_i > q_j \end{cases}$$

olarak tanımlanırsa [1.1] eşitsizliği Kendall'ın τ sıra korelasyon katsayısını verir. Burada ayrıca $c_{ij} = -c_{ji}$, $c_{ii} = 0$, $d_{ij} = -d_{ji}$ ve $d_{ii} = 0$ özelliklerini taşır. Spearman'ın ρ sıra korelasyon katsayısına ise,

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

ve

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

iken ulaşılır(2). Bu durumda, [1.1] eşitsizliğindeki Γ istatistiği, $g_i = p_i - q_i$ olmak üzere,

$$\rho = 1 - \frac{6 \sum g_i^2}{n^3 - n} \quad [1.2]$$

şeklinde ifade edilir(2). Buz pateni ve kayak gibi bazı spor dallarında, ilk üç sıralamaya verilen önem son sıralamalara göre çok daha fazladır. Bu nedenle, ilk sıralamalardaki yer değişikliklerini daha iyi yansıtan yeni bir sıra korelasyon katsayısına ihtiyaç duyulabilir. Bu çalışmada Blest(2000) tarafından geliştirilmiş olup, grafiksel gösterime de açık olan, mevcut sıra korelasyon katsayılarına alternatif bir sıra korelasyon katsayısı tanıtılmıştır. Tanıtılan v sıra korelasyon katsayısının temel özelliği ise, sıralamalardaki ilk yer değişikliklerine daha fazla önem vermesidir. Bu özellik tüm $n \geq 4$ durumları için geçerlidir(3). Tanıtılan korelasyon katsayısının yapısını daha ayrıntılı bir biçimde görmek amacıyla, makul sayıda mümkün sıralamaya imkan tanıdığından sadece $n = 5$ durumu ele alınmıştır. $n = 5$ için, gözlemlerin mümkün sıralamalarındaki yer değişikliklerinin incelenmesi sonucunda, v sıra korelasyon katsayısının, sıralamalardaki küçük yer değişikliklerinde bile farklı sonuçlar verdiği ve bu nedenle sıralamalardaki yer değişikliklerini ρ ve τ 'ya göre daha iyi yansıttığı gösterilmiştir.

this statistic changes to different correlation coefficients which depends on the definitions of c_{ij} and d_{ij} (2).

Taking $i, j \in \{1, 2, \dots, n\}$, let p_i be the assigned rank scores for observed values of variable X and q_i be assigned rank scores for observed values of variable Y. Assuming that both X and Y variables don't have tied observations, in other words there is no common rank scores for any two observations, the equation [1.1] gives Kendall's τ rank correlation coefficient when c_{ij} and d_{ij} are defined as,

and

Where c_{ij} and d_{ij} have the following properties, $c_{ij} = -c_{ji}$, $c_{ii} = 0$, $d_{ij} = -d_{ji}$ ve $d_{ii} = 0$. The rank correlation coefficient of Spearman's ρ can be obtained when

and

(2). In this case, Γ statistic in equation [1.1], when $g_i = p_i - q_i$, can be written as follows (2);

In some sport activities as ice skating and skiing the given importance of the first three place ranking is much more than the last place ranking. For this reason, a rank correlation coefficient that more reflects the interchanges in the first places ranking may be required. In this work, the rank correlation coefficient which is an alternative to existing rank correlation coefficients also suitable to graphical presentation is introduced, developed by Blest(2000). The basic property of the presented rank correlation coefficient is giving more importance to the first places of ordering. This property is valid in all situations for $n \geq 4$ (3). To see the structure of presented rank correlation coefficient in more detail, only the situation of $n=5$ is considered because the reasonable number of situations arise to deal with. In the result of investigation for the interchanges in possible orderings of the observations for the case $n=5$, it has been observed that v rank correlation coefficient gives different results even though in small interchanges in the orderings occurs and for this reason it reflects the interchanges in the orderings much better than ρ and τ .

2.SIRALANMIŞ VERİ İÇİN GRAFİKSEL GÖSTERİM VE S İSTATİSTİĞİ

Birinci bölümde tanıtılan Kendall'ın τ ve Spearman'ın ρ sıra korelasyon katsayılarının hesaplanmasında temel fikir, birimlerin gözlem değerlerinin küçükten büyüğe sıralandığı (doğal sıralama) durumdan sapmalarının belirlenmesidir. Bu sapmaları görmek için S istatistiğinin kullanılmasına dayanan grafiksel bir yaklaşımdan faydalanılabilir.

S istatistiğinin hesaplanmasında, öncelikli olarak hem X hem de Y değişkenleri bakımından gözlenen örnek birimlerine ayrı ayrı sıra sayıları atanır. Daha sonra X değişkenine atanan sıra sayıları doğal sıralamada olacak şekilde düzenlenir ve bunlara karşılık gelen Y deki sıra sayıları kullanılır. Böylece p_i ($i = 1, \dots, n$) ile gösterilen X in sıra sayıları artık "i" ile ifade edilebilir. Y için atanan sıra sayıları q_i olsun. Bu sıralama için sıra sayılarının kümülatif değerleri bulunur. X in doğal sırasına bağlı kalmak üzere, her birim için sırasıyla, Y gözlem birimine karşılık gelen kümülatif sıra sayısı değerinden, X gözlem biriminin kümülatif sıra sayısı değeri çıkartılır. n örneğe çıkan birimlerin toplam sayısını göstermek üzere, k . gözlem birimi için kümülatif sıra sayılarının fark istatistiği,

$$v_k = \sum_{i=1}^k q_i - \sum_{i=1}^k i = \sum_{i=1}^k (q_i - i) \quad k = 1, 2, \dots, n \quad [2.1]$$

biçiminde tanımlanır. [2.1] eşitliğinde, $\sum_{i=1}^k q_i$; Y değişkeni bakımından k . birimin kümülatif sıra sayısı değeri ve $\sum_{i=1}^k i$ ise doğal sıralama durumundaki X değişkeni bakımından k . birimin kümülatif sıra sayısı değeridir. Fark istatistiklerinin toplamı,

$$S = \sum_{k=1}^n v_k = \sum_{k=1}^n \sum_{i=1}^k (q_i - i) \quad [2.2]$$

istatistiğini verir. Bu istatistik örneğe çıkan birimlerin, doğal sıralamadan olan farklılığını gösteren bir ölçüdür. Örneğe çıkan birimler, doğal sıralamada olma özelliğine sahipse, yani tüm i ler için $q_i = i$ ise S istatistiği alabileceği minimum değeri alır ve $S_{\min} = 0$ olur. Dolayısıyla, bu istatistiğin alabileceği maksimum değer,

$$S_{\max} = \sum_{k=1}^n v_k^* = \sum_{k=1}^n \sum_{i=1}^k [(n+1-i) - i] = \frac{n(n^2-1)}{6} \quad [2.3]$$

ile bulunur. Burada v_k^* ; X in sıralaması doğal sıralamanın tersi iken, k . birime ait sıra sayılarının kümülatif farklarını gösterir. S istatistiğinden yararlanarak örnek birimlerinin doğal sıralamadan sapmaları grafiksel olarak görülebilir. Bunun için sıra sayılarının kümülatif farklarını gösteren

2. GRAPHICAL PRESENTATION AND STATISTIC S FOR RANKED DATA

The basic idea of calculate Kendall's τ and Spearman's ρ rank correlation coefficients which is presented in the first section is to determine the deviations of the observed values of units from the state that they are ranked from smallest to largest. For see this deviations, a graphical presentation which is based on the use of S statistic can be benefited.

For the calculation of the statistic S , first the rank scores are assigned for observation units separately for variable X and Y . Then the bivariate observations ordered by the assigned rank scores of variable X are arranged in natural order and the interrelated rank scores of Y are used. Thus the rank scores of X is p_i ($i = 1, \dots, n$)

could be renamed by "i". Let q_i be the scores of Y . The cumulative rank values for this ordering are found. Depending on the natural order of X , respectively for every unit, differ the cumulative rank score value of observed unit Y , from the cumulative rank score value of observed unit X . The summation of the units in the sample are presented as n and so the statistic of rank differences of cumulative rank scores for k -th unit is defined as,

In equation [2.1], $\sum_{i=1}^k q_i$ is the cumulative rank score value of k -th unit of variable Y and $\sum_{i=1}^k i$ is the cumulative rank score value of the k -th unit of naturally ordered variable X . The summation of difference statistics gives the statistic,

This statistic is a measure which denotes the difference in the observation units from the natural order. This statistic is a measure which denotes the difference in the sample units from the natural order. When the units in the sample are in the natural order in other words if $q_i = i$, for all i , then S takes the minimum value, $S_{\min} = 0$. From this, the maximum value of this statistic can be found by,

Here v_k^* is the cumulative difference of rank scores for k -th unit while X is in inverse natural order. By utilizing from the S statistic, the deviations of the sample units from natural order can be identify graphically. For this, the

V_k lar, doğal sıralamanın tersi olması durumunda alabileceği maksimum V_k^* lar ile grafiksel olarak karşılaştırılır.

$n = 5$ için grafiksel karşılaştırma yapıldığında, X in 1 2 3 4 5 doğal sıralamasına karşın Y deki örnek birimlerinin sıralaması, 4 3 1 5 2 olur. Bu durumda, S ve S_{mak} istatistiklerinin hesaplanmasında, Tablo1 de oluşturulan değerler kullanılmıştır. Böylece [2.2] eşitliğinden $S=12$ ve [2.3] eşitliğinden de $S_{mak}=20$ değerleri hesaplanır.

cumulative differences of rank scores V_k are compared graphically with its maximum V_k^* , when it's in inverse natural order.

Making a graphical comparison for $n=5$ according to the natural order of rank scores of X; 1 2 3 4 5, let the observed rank scores of Y be 4 3 1 5 2. In this state the values in Table1 are used for calculate the statistics S and S_{mak} . Thus from equation [2.2] $S=12$ and from equation [2.3] $S_{mak}=20$ values are calculated.

Table 1. An example for $n=5$.

Tablo 2. $n = 5$ için bir örnek

i	q_i	$\sum_{i=1}^k q_i$	$\sum_{i=1}^k i$	$v_k = \sum_{i=1}^k (q_i - i)$	$(n+1-i)$	$\sum_{i=1}^k (n+1-i)$	$v_k^* = \sum_{i=1}^k [(n+1-i) - i]$
1	4	4	1	3	5	5	4
2	3	7	3	4	4	9	6
3	1	8	6	2	3	12	6
4	5	13	10	3	2	14	4
5	2	15	15	0	1	15	0

Tablo 1 den yararlanarak, V_k ve V_k^* kümülatif farklarının k ya göre grafiksel gösterimi Şekil 1'de verilmiştir.

From using the Table 1 the graphical presentation of cumulative difference V_k and V_k^* with respect to k is given in Figure1.

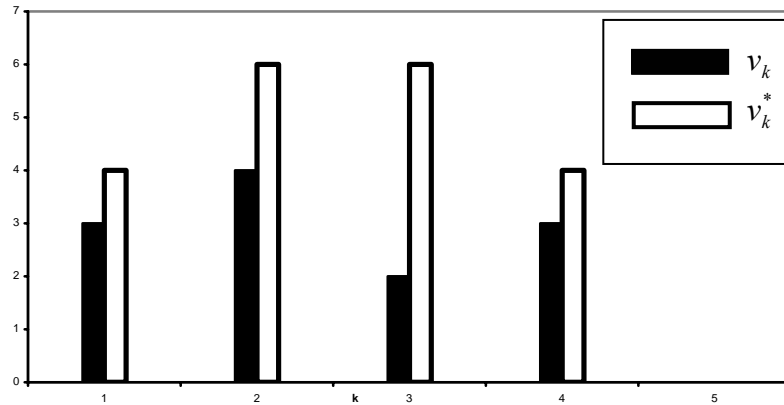


Figure1. V_k and V_k^* values for the sample size $n=5$

Şekil 1. $n = 5$ çaplı örnek için V_k ve V_k^* değerleri ($k=1,2,\dots,5$)

S ve S_{mak} istatistiklerine dayalı ve $[-1,1]$ aralığında değer alan γ sıra korelasyon katsayısı

Based on statistics S and S_{mak} and taking values in the range of $[-1,1]$ the correlation coefficient γ is defined as ,

$$\gamma = 1 - \frac{2S}{V_{mak}} = 1 - \frac{12S}{n(n^2 - 1)} \tag{2.4}$$

olarak tanımlanır(3). Burada S istatistiği,

(3). Here the statistic S can be introduced as,

$$S = \sum_{k=1}^n \sum_{i=1}^k (q_i - i) = \sum_{k=1}^n \sum_{i=1}^k (n + 1 - i)(q_i - i) = \frac{1}{2} \sum_{i=1}^k d_i^2 \quad [2.5]$$

biçiminde tanımlanabilir. Bu sonucun [2.4] de yerine yazılmasıyla Spearman'ın ρ sıra korelasyon katsayısı elde edilir. [2.5] eşitliği aynı zamanda ,

$$S = \sum_{i=1}^k (n + 1 - i)q_i - \frac{n(n + 1)(n + 2)}{6} \quad [2.6]$$

ile gösterilebilir(3).

3. v SIRA KORELASYON KATSAYISI

Tablo1 de verilen $n = 5$ çaplı örnek için, v_k ve v_k^* ların $\sum_{i=1}^k$ ye göre çizilen grafiği Şekil 2 de verilmiştir. Bu grafikte ortaya çıkan I II III IV ve V alanları doğal sıralamadan sapmaları ifade etmektedir. Ayrıca bu grafiksel gösterimde de görüldüğü gibi V alanı I alanından daha büyüktür. Yani, Y nin sıralamasında son sıralarda meydana gelen doğal sıralamadan sapmalar, ilk sıralarda meydana gelen doğal sıralamadan sapmalara göre daha büyük alanları ifade eder.

From writing this result in [2.4] Spearman's ρ correlation coefficient is obtained. The equation [2.5] can also be shown as,

3. RANK CORRELATION COEFFICIENT v

For the given example in Table1, with sample size $n=5$ the lined graph of v_k and v_k^* with respect to $\sum_{i=1}^k$ is given. The areas I, II, III, IV and V which occur in this graph expresses the deviations from the natural ordering. Furthermore as be shown in the graphical representation the area V is much bigger than area I. In other words, in the ordering of Y the deviations of the natural ordering occur in tail end of the ranking, expresses much bigger areas with respect to early rankings.

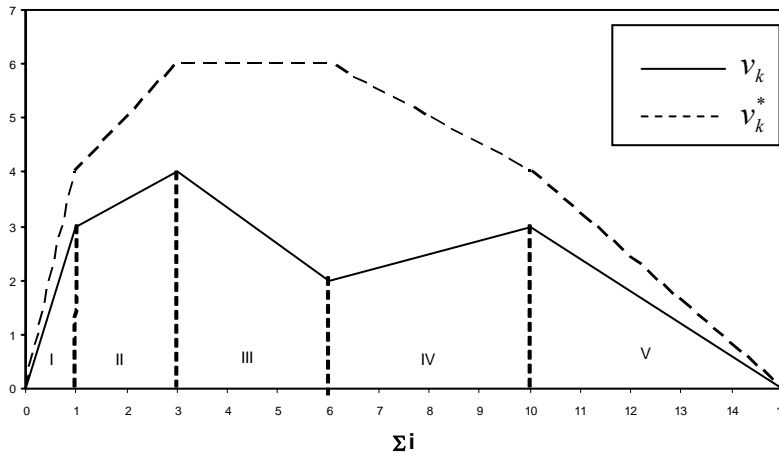


Figure 2. Cumulative differences of v_k and v_k^* with respect to $\sum_{i=1}^k$

Şekil 2. $\sum_{i=1}^k$ değerlerine göre v_k ve v_k^* kümülatif farkları

$\sum_{i=1}^k$ değerlerine göre v_k ve v_k^* kümülatif farkları çizildiğinde, oluşan alanların büyüklüğü doğal sıralamadan olan sapmaların ne denli önemli olduğunu gösterir. Örneğin; $n = 5$ iken Y nin yalnızca ilk iki sırası yer değiştiğinde oluşan 2 1 3 4 5 sıralamasına ilişkin kümülatif farkların grafiği çizildiğinde, alan büyüklüğü 1.5 olarak hesaplanırken, son iki sırası yer değiştiğinde oluşan 1 2 3 5 4 sıralamasına ilişkin alan 4.5 olur. Bu alanlar hesaplanırken bilinen üçgen alanı hesaplama

If the cumulative differences of v_k and v_k^* with respect to $\sum_{i=1}^k$ is going to line, the bigness of the occurred areas represents how much important the deviation from natural ordering is. For example when $n = 5$, the size of the area is 1.5 when the first two rank scores of Y is interchanged relative to natural order namely 2 1 3 4 5 and it's 4.5 when the last two rank scores of Y is interchanged namely 1 2 3 5 4. For the calculation of the areas, the

formüllerinden yararlanır. Örnekten de görüldüğü gibi ilk sıralarda meydana gelen yer değişikliği, son sıralarda meydana gelen yer değişikliğine göre daha küçük bir alana sahiptir. Fakat bazı spor dallarında (Buz pateni, jimnastik, dalma v.b.) final sıralaması bir dizi farklı kategorilere göre sıralamaya dayanır. Ancak farklı kategorilerdeki ilk üç sıralama, final sıralamasını belirleyebilecek özelliğe sahip olduğundan, ilk sıralamalardaki yer değişiklikleri daha önemli olacaktır. n nin büyük değerleri için puanlamalar yapılırken, ilk sıralamaların daha özenli belirlenmesi mümkünken, daha sonraki sıralamalara verilen önemin konsantrasyon ve dikkat eksikliğinden dolayı azaldığı görülebilir (4). Bu durumda, ilk sıralamalardaki yer değişikliklerini daha çok önemseyen yeni bir sıra korelasyon katsayısına ihtiyaç duyulabilir. Tanıtılan v sıra

korelasyon katsayısı ise, doğal sıralamanın kümülatifi $\sum_{i=1}^k i$ yerine, doğal sıralamanın tersinin kümülatifi $\sum_{i=1}^k (n+1-i)$

kullanılarak V_k ve V_k^* kümülatif farklarının grafiksel olarak incelenmesine dayanır. Böylece oluşacak alanlar ilk sıralamaya önem veren alternatif bir sıra korelasyon katsayısının oluşumuna izin verir. Bu durum ikinci bölümde verilen 4 3 5 1 2 örneğinden görülebilir(Şekil3).

known triangle area calculation formulizations are used. As it can be seen in the example the interchanges occur in the early rankings has much more a smaller area than in the last ranking. However in some sport activities (ice skating, gymnastic, diving, e.g.) the final ranking is based on a set of different categories of ranking. Especially, because of the first three ranking in different categories has a property that it can be determine the final ranking, interchanges in the first ranking is much more important. Making scores for larger values of n, while it's possible to perform the first ranking more carefully, the given importance to the ranks afterwards is reduced because of deficiency of concentrations and attention(4). In this case, a new rank correlation coefficient which gives much more importance to interchanges in the early ranking may be required. The introduced correlation coefficient v, which uses the cumulative of inverse natural order $\sum_{i=1}^k (n+1-i)$,

instead of the cumulative of natural order $\sum_{i=1}^k i$, is based on

graphical investigation of cumulative differences of V_k ve

V_k^* . From this the appearing areas gives a permission to form an alternative rank correlation coefficient which gives importance to the early ranking. This case can be seen from the example 4 3 5 1 2 given in the second section (Figure 3).

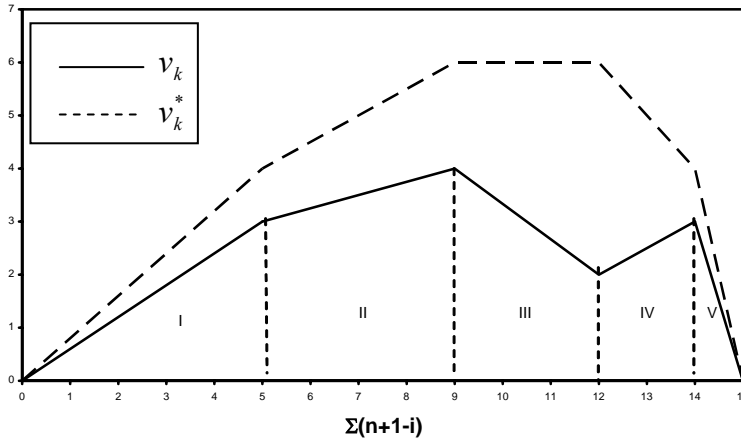


Figure 3. V_k and V_k^* cumulative difference with respect to $\sum_{i=1}^k (n+1-i)$.

Şekil 3. $\sum_{i=1}^k (n+1-i)$ değerlerine göre V_k ve V_k^* kümülatif farkları

Herhangi bir n değeri için, $\sum_{i=1}^k (n+1-i)$ 'ye ve V_k kümülatif fark istatistiklerine göre oluşan grafiğin toplam alanı;

For any value of n, the summation of the graph areas which occurs for the cumulative difference statistics V_k ,

with respect to $\sum_{i=1}^k (n+1-i)$ can be found with,

$$W = \frac{1}{2} \sum_{i=1}^n (n+1-i)^2 q_i - \frac{n(n+1)^2(n+2)}{24} \quad [3.1]$$

ile bulunur. Y değişkeni doğal sıralamada olduğunda, W alabileceği en küçük değeri almış olur. Yani $W_{\min}=0$ dır. Doğal sıralamanın tersi durumunda ise, W nin alabileceği en büyük değer

When variable Y is in natural order W takes as minimum value as it can be take. In other words $W_{\min}=0$. If in the state of inverse natural order the biggest value which W will be take is obtained from,

$$W_{\max} = \frac{n(n+1)^2(n-1)}{12} \quad [3.2]$$

den elde edilir. Tablo1 deki veri kullanılarak, Şekil 3 için $W= 37$ ve $W_{\max}= 60$ olur. Y değişkeninin herhangi bir sıralamasının doğal sıralamadan olan sapmasını, W ve W_{\max} alanlarına dayalı olarak belirlemek mümkündür. Bu alanları kullanarak v sıra korelasyon katsayısı,

For Figure 3, using the sample given in Table 1 $W=37$ and $W_{\max}=60$ is obtained. The deviation of any ordering of variable Y from the natural ordering will be expressed from the areas W and W_{\max} . After using this areas the rank correlation coefficient v is calculated from,

$$v = 1 - \frac{2W}{W_{\max}} = 1 - \frac{24W}{n(n+1)^2(n-1)} \quad [3.3]$$

ile hesaplanır. Bu katsayının Spearman'ın ρ ve Kendall'in τ sıra korelasyon katsayılarına göre daha güçlü ayırt edici özelliğe sahip olduğu görülmüştür(3). Yani Y değişkeni için oluşan mümkün dizilişlerde, ρ ve τ da tekrar eden değerlerin v sıra korelasyon katsayısından daha fazla olduğu gözlenmiştir.

It has seen that, this rank correlation coefficient have more distinguishable property compared to Spearman's ρ and Kendall's τ rank correlation coefficients (3). In other words, it is observed that in the possible orderings for variable Y, there are more repeatable values for ρ and τ than the rank correlation coefficient v.

$n=5$ iken Y değişkeni bakımından oluşturulabilecek tüm mümkün 120 farklı sıralama için ρ , τ ve v sıra korelasyon katsayılarının değerleri Ekl de verilmiştir. Ayrıca n in küçük değerleri için bu katsayı değerlerinden faydalanılarak v ve ρ yada v ve τ arasındaki çarpım korelasyon katsayılarının hesaplanması mümkünken, n in 12 yi aştığı durumlarda mümkün sıralamaların sayısı çok fazla olduğundan dolayı hesaplama zorluğu ortaya çıkar. Spearman'ın ρ ve tanıtilen v sıra korelasyon katsayısı arasındaki çarpım moment korelasyon katsayısını hesaplamak için [2.6] ve [3.1] formüllerinden yararlanılabilir. Ancak sabit terimler önemsiz olduğundan ρ ve v katsayısı için sırasıyla,

In Appendix 1, ρ , τ and v correlation coefficients are given for every 120 possible order to be formed by variable of Y when $n=5$. Moreover, while for small values of n, v and ρ or v and τ product moment correlation coefficients can be calculated using coefficient values easily, however because of the possible orderings much more in the case of n greater than 12, there is a calculation problem appear. To calculate moment product correlation coefficient between Spearman's ρ and Blest's v, equation [2.6] and equation [3.1] can be used. Since the constant terms are trivial, using

$$u = \sum_{i=1}^n (n+1-i)q_i \quad [3.4]$$

ve

and

$$w = \sum_{i=1}^n (n+1-i)^2 q_i \quad [3.5]$$

terimlerini kullanmak yeterli olacaktır. Böylece

terms are enough for ρ and v coefficients respectively.Hence,

$$r_{pv} = r_{uw} = \frac{E(uw) - E(u)E(w)}{\sqrt{(E(u^2) - E(u)^2)(E(w^2) - E(w)^2)}} \quad [3.6]$$

olmak üzere, [3.6] eşitliği

is rewritten as

$$r_{pv} = \frac{(n+1)\sqrt{15}}{\sqrt{(2n+1)(8n+11)}}$$

şeklinde elde edilir. Burada,

Where

$$E(u)=E\left(\sum_{i=1}^n(n+1-i)q_i\right)=\frac{n(n+1)^2}{4}$$

$$E(u^2)=E\left(\left[\sum_{i=1}^n(n+1-i)\right]^2\right)=\frac{n^2(n+1)^2(9n^2+19n+8)}{144}$$

$$E(w)=E\left(\sum_{i=1}^n(n+1-i)^2q_i\right)=\frac{n(n+1)^2(2n+1)}{12} \tag{3.7}$$

$$E(w^2)=E\left(\left[\sum_{i=1}^n(n+1-i)^2q_i\right]^2\right)=\frac{n^2(n+1)^2(2n+1)(30n^3+83n^2+63n+4)}{2160} \tag{3.8}$$

ve

and

$$E(uw)=E\left(\left[\sum_{i=1}^n(n+1-i)q_i\right]\left[\sum_{i=1}^n(n+1-i)^2q_i\right]\right)=\frac{n^2(n+1)^3(3n^2+5n+1)}{24}$$

dır(3). Ayrıca,

(3). Moreover when

$$r_{\tau p}=\frac{2(n+1)}{\sqrt{2n(2n+5)}}$$

olmak üzere,

and

$$r_{\tau v}\approx r_{\tau p}\times r_{pv}$$

olarak ifade edildiğinden, τ ve v katsayısı arasındaki çarpım moment korelasyon katsayısı $r_{\tau v}$ için yaklaşık bir çözüm,

an approximate solution of product moment correlation coefficient $r_{\tau v}$ between Kendall's τ and Blest's v is obtained as follow,

$$r_{\tau v}\approx\frac{2\sqrt{15}(n+1)^2}{\sqrt{2n(2n+1)(2n+5)(8n+11)}}$$

şeklinde bulunur. $n=4(1)12$ olmak üzere, r_{pv} , r_{pv} ve $r_{\tau v}$ korelasyon katsayılarına ilişkin değerler Tablo 2 de verilmiştir.

For $n=4(1)12$, product moment correlation coefficients $r_{\rho\tau}$, $r_{\rho v}$ and $r_{\tau v}$ are given in Table2.

Table2: The product moment correlation coefficient values among rank correlation coefficients.

Table2: Sıra korelasyon katsayıları arasındaki çarpım moment korelasyon değerleri

n	$r_{\rho\tau}$	$r_{\rho v}$	$r_{\tau v}$
4	0.9806	0.9844	0.9653
5	0.9798	0.9811	0.9613
6	0.9802	0.9789	0.9595
7	0.9810	0.9774	0.9588
8	0.9820	0.9762	0.9586
9	0.9829	0.9753	0.9586
10	0.9839	0.9746	0.9588
11	0.9848	0.9740	0.9591
12	0.9855	0.9735	0.9594

Tablo2 den de görüldüğü gibi, üç sıra korelasyon katsayısı arasındaki çarpım moment korelasyon katsayıları 1 e yakın değerler almıştır. Bu yüzden de, genel anlamda değerlendirildiğinde, katsayılar arasında güçlü bir ilişki olduğu söylenebilir. Ayrıca r_{pv} değerlerinin n arttıkça azaldığı gözlenmiştir.

As it can be seen in Table 2, the values of three product moment correlation coefficients are closed to one. For this reason, in general, it can be said that there is a strong correlation between these coefficients. Furthermore, it is observed that r_{pv} increases as n decreases.

4. v İSTATİSTİĞİNİN DAĞILIM ÖZELLİKLERİ

Üçüncü bölümde [3.7] eşitliğindeki beklenen değer ifadesi kullanılarak [3.3] eşitliğinde tanımlanan v katsayısının beklenen değeri

$$E(v) = 1 - \frac{n(n+1)^2(n-1)}{n(n+1)^2(n-1)} = 0 \tag{4.2}$$

olarak elde edilir. [3.7] ve [3.8] eşitliklerinin

$$\text{Var}(w) = E(w^2) - [E(w)]^2$$

varyans formülünde yerine yazılmasıyla, v sıra korelasyon katsayısı için varyans ifadesi

$$\text{Var}(v) = \frac{576(2n+1)(8n+11)}{8640(n+1)^2(n-1)} \tag{4.3}$$

olacaktır. n=5 için, Ek1 de verilen v'nin örnekleme dağılımı kullanılarak oluşturulan olasılık dağılımı incelendiğinde, bu istatistiğin 0 ortalama etrafında simetrik dağıldığı gözlenmiştir. Yani, v istatistiğinin hesaplanmasında yer alan q_i sıra sayıları yerine n+1-q_i değerlerinin kullanılması ile oluşan V' değerleri arasında, V'=-v ilişkisi vardır. Buna göre tüm n değerleri için v istatistiği simetrik dağılıma sahip olacaktır. Ayrıca n örnek çapının 12 den büyük değerleri için v istatistiği normal dağılıma yaklaşır(3). Böylece v istatistiğinin beklenen değeri 0 varyansı da $\frac{576(2n+1)(8n+11)}{8640(n+1)^2(n-1)}$ olmak üzere hipotez testlerinde,

$$z = \frac{v-0}{\sqrt{\frac{576(2n+1)(8n+11)}{8640(n+1)^2(n-1)}}} \sim N(0,1)$$

test istatistiğinden yararlanır.v istatistiğine ilişkin hipotez testlerinin gerçekleştirilmesi amacıyla, farklı α anlamlılık düzeylerinde n=4,5,...,12 için v* kritik değerleri, Blest'in w ye ilişkin kritik değerlerinden yararlanılarak hesaplanmıştır(3). Hesaplanan bu değerler Tablo 3 de verilmiştir.

Table 3. The v* critical values satisfying Pr(v ≥ v*)=α
Tablo 3. Pr(v ≥ v*)=α'yı sağlayan v* kritik değerleri

n	α Significance Levels α Anlamlılık Düzeyleri					
	0.001	0.005	0.01	0.02	0.05	0.10
4	-----	-----	-----	-----	0.9200	0.8400
5	-----	-----	0.9666	0.9333	0.8333	0.7166
6	-----	0.9510	0.9184	0.8694	0.7551	0.6408
7	0.9598	0.9063	0.8616	0.8036	0.6964	0.5714
8	0.9365	0.8651	0.8201	0.7593	0.6455	0.5212
9	0.9050	0.8767	0.7783	0.7167	0.6017	0.4833
10	0.8766	0.7950	0.7444	0.6804	0.5669	0.4545
11	0.8500	0.7644	0.7121	0.6477	0.5114	0.4295
12	0.8257	0.7364	0.6837	0.6197	0.5126	0.4083

4. DISTRIBUTION PROPERTIES OF v STATISTIC

Using equation [3.7] given in the third section, expected value of the v statistic in equation [3.3] can be obtained as,

When substituting equation [3.7] and equation [3.8] in the following formulation of variance,

then the variance of v rank correlation coefficient is obtained as,

Investigating the probability distribution which is obtained from using v sampling distribution of the example given in Appendix 1, for n=5, it is seen that this statistic has a symmetric distribution around zero mean. In other words there is a V' = -v relationship between v and V' which is occurred from the calculation of v statistic using n+1-q_i instead of q_i. Thus, v statistic has a symmetric distribution for all value of n. In addition, v statistic approximates to normal distribution for the sample size n greater than 12. Since v has zero mean and $\frac{576(2n+1)(8n+11)}{8640(n+1)^2(n-1)}$ variance, the following test statistic used in hypothesis testing

For the realization of hypothesis testing about the v statistic, the critical values v* are calculated by using the critical values of Blest's w for different α significance levels when n=4,5,...,12 (3). These calculated values are given in Table 3.

5. SONUÇ

İki değişken arası ilişkinin belirlenmesinde, test işlemleri için normal dağılım varsayımına ihtiyaç duymayan sıra korelasyon katsayıları Kendall'ın τ ve Spearman'ın ρ sudur. Örneğe çıkan birimlerin sıralanmasına dayalı bu istatistikler, sıralamadaki küçük yer değişikliklerini iyi yansıtmazlar. Ancak bazı spor dallarında ilk üç sıralamanın, final sıralamasını etkileyebilecek özelliğe sahip olması nedeniyle, sıralamalardaki yer değişikliklerini daha çok önemseyen alternatif bir sıra korelasyon katsayısına ihtiyaç duyulabilir. Bu katsayı, doğal sıranın tersinin kümülatifi olan $\sum_{i=1}^k (n+1-i)$ 'yi temel alarak, v_k ve v_k^* kümülatif farklarının grafiksel olarak incelenmesine dayanır. Bu çalışmada, $n=5$ iken tüm mümkün sıralamalar üzerinden τ , ρ ve v sıra korelasyon katsayıları hesaplanarak tanıtılan v istatistiğinin, sıralamalardaki küçük yer değişikliklerinde bile daha az tekrar eden sonuçlar verdiği gözlenmiştir (Ek 1). Bu durum $n \geq 4$ iken oluşturulacak tüm mümkün sıralamalar için de geçerlidir. Ayrıca bu istatistiğin 0 ortalama etrafında simetrik dağıldığı görülmüş ve $n=4,5,\dots,12$ için farklı anlamlılık düzeylerine karşılık gelen v^* kritik değerleri elde edilmiştir.

5. CONCLUSION

The correlation coefficients which don't require normality assumption for determining the correlation between two variables are Kendall's τ and Spearman's ρ . These statistics based on ordering the sample values does not reflect the interchanges in the orderings well. Since in some of the sport activities first three places have an importance in the final ranking, an alternative rank correlation coefficient which gives more emphasize to interchanges in the ranking can be required. This coefficient is based on the graphical investigation of cumulative differences v_k and v_k^* on the basis of cumulative for inverse natural order $\sum_{i=1}^k (n+1-i)$. In this study, for the all possible rank score orderings, τ , ρ and v values are calculated and it is observed that v statistic gives less repeatable results than τ and ρ for the interchanges of orderings (appendix 1). This situation is also valid for constituting of all possible orderings when $n \geq 4$. Moreover it is seen that this statistic has a symmetric distribution around zero mean and v^* critical values are obtained corresponding to different α significance levels for $n=4, 5, \dots, 12$.

KAYNAKLAR/ REFERENCES

1. Conover, W.J., *Practical Nonparametric Statistics*, **John Wiley & Sons**, 250-263, (1980).
2. Kendall, M.G., Gibbons, J.D., *Rank Correlation Methods*, 5th edn.London: **Edward Arnold**, 25-27, (1990).
3. Blest, D. "Rank Correlation-An Alternative Measure", *Austral. and New Zealand J.Statist.*,42(1):101-111, (2000).
4. Miller, G.A. "The Magical Number Seven, Plus or Minuse Two: Some Limits on Our Capacity for Processing Information", *The Psychological Review*,63:81-97, (1956).

Received/ Geliş Tarihi: 23.03.,2004 Accepted/Kabul Tarihi: 20.10.2004

EK 1:
n=5 için sıra korelasyon katsayıları

APPENDIX I:
Rank correlation coefficients for n=5.

Rank Scores/ Sıra Sayıları	τ	ρ	v	Rank Scores/ Sıra Sayıları	τ	ρ	v
5 4 3 2 1	-1.0000	-1.0000	-1.0000	3 2 5 4 1	-0.2000	-0.2000	-0.0667
5 4 3 1 2	-0.8000	-0.9000	-0.9500	3 2 5 1 4	0.0000	0.1000	0.0833
5 4 2 3 1	-0.8000	-0.9000	-0.9167	3 2 4 5 1	0.0000	-0.1000	0.0167
5 4 2 1 3	-0.6000	-0.7000	-0.8167	3 2 4 1 5	0.2000	0.3000	0.2167
5 4 1 2 3	-0.4000	-0.6000	-0.7333	3 2 1 4 5	0.4000	0.6000	0.4667
5 4 1 3 2	-0.6000	-0.7000	-0.7833	3 2 1 5 4	0.2000	0.5000	0.4167
5 3 4 2 1	-0.8000	-0.9000	-0.8833	3 1 5 2 4	0.2000	0.3000	0.2833
5 3 4 1 2	-0.6000	-0.8000	-0.8333	3 1 5 4 2	0.0000	0.1000	0.1833
5 3 2 4 1	-0.6000	-0.7000	-0.7167	3 1 2 5 4	0.4000	0.6000	0.5333
5 3 2 1 4	-0.4000	-0.4000	-0.5667	3 1 2 4 5	0.6000	0.7000	0.5833
5 3 1 2 4	-0.2000	-0.3000	-0.4833	3 1 4 2 5	0.4000	0.5000	0.4167
5 3 1 4 2	-0.4000	-0.5000	-0.5833	3 1 4 5 2	0.2000	0.2000	0.2667
5 2 3 4 1	-0.4000	-0.6000	-0.6000	2 4 3 5 1	0.0000	-0.1000	0.0500
5 2 3 1 4	-0.2000	-0.3000	-0.4500	2 4 3 1 5	0.2000	0.3000	0.2500
5 2 4 3 1	-0.6000	-0.7000	-0.6833	2 4 5 3 1	-0.2000	-0.3000	-0.1167
5 2 4 1 3	-0.4000	-0.5000	-0.5833	2 4 5 1 3	0.0000	-0.1000	-0.0167
5 2 1 4 3	-0.2000	-0.2000	-0.3333	2 4 1 5 3	0.2000	0.3000	0.3167
5 2 1 3 4	0.0000	-0.1000	-0.2833	2 4 1 3 5	0.4000	0.5000	0.4167
5 1 3 2 4	0.0000	-0.1000	-0.2500	2 3 4 5 1	0.2000	0.0000	0.1667
5 1 3 4 2	-0.2000	-0.3000	-0.3500	2 3 4 1 5	0.4000	0.4000	0.3667
5 1 2 3 4	0.2000	0.0000	-0.1667	2 3 5 4 1	0.0000	-0.1000	0.0833
5 1 2 4 3	0.0000	-0.1000	-0.2167	2 3 5 1 4	0.2000	0.2000	0.2333
5 1 4 2 3	-0.2000	-0.3000	-0.3833	2 3 1 5 4	0.4000	0.6000	0.5667
5 1 4 3 2	-0.4000	-0.4000	-0.4333	2 3 1 4 5	0.7000	0.7000	0.6167
4 5 3 2 1	-0.8000	-0.9000	-0.8500	2 5 3 4 1	-0.2000	-0.3000	-0.1500
4 5 3 1 2	-0.6000	-0.8000	-0.8000	2 5 3 1 4	0.0000	0.0000	0.0000
4 5 2 3 1	-0.6000	-0.8000	-0.7667	2 5 4 3 1	-0.4000	-0.4000	-0.2333
4 5 2 1 3	-0.4000	-0.6000	-0.6667	2 5 4 1 3	-0.2000	-0.2000	-0.1333
4 5 1 2 3	-0.2000	-0.5000	-0.5833	2 5 1 4 3	0.0000	0.1000	0.1167
4 5 1 3 2	-0.4000	-0.6000	-0.6333	2 5 1 3 4	0.2000	0.2000	0.1667
4 3 5 2 1	-0.6000	-0.7000	-0.6167	2 1 3 5 4	0.6000	0.8000	0.8000
4 3 5 1 2	-0.4000	-0.6000	-0.5667	2 1 3 4 5	0.8000	0.9000	0.8500
4 3 2 5 1	-0.4000	-0.4000	-0.3667	2 1 5 3 4	0.4000	0.6000	0.6333
4 3 2 1 5	-0.2000	0.0000	-0.1667	2 1 5 4 3	0.2000	0.5000	0.5833
4 3 1 2 5	0.0000	0.1000	-0.0833	2 1 4 5 3	0.4000	0.6000	0.6667
4 3 1 5 2	-0.2000	-0.2000	-0.2333	2 1 4 3 5	0.6000	0.8000	0.7667
4 2 3 5 1	-0.2000	-0.3000	-0.2500	1 4 3 2 5	0.4000	0.6000	0.6000
4 2 3 1 5	0.0000	0.1000	-0.0500	1 4 3 5 2	0.2000	0.3000	0.4500
4 2 5 3 1	-0.4000	-0.5000	-0.4167	1 4 2 3 5	0.6000	0.7000	0.6833
4 2 5 1 3	-0.2000	-0.3000	-0.3167	1 4 2 5 3	0.4000	0.5000	0.5833
4 2 1 5 3	0.0000	0.1000	0.0167	1 4 5 2 3	0.2000	0.2000	0.3333
4 2 1 3 5	0.2000	0.3000	0.1167	1 4 5 3 2	0.0000	0.1000	0.2833
4 1 3 2 5	0.2000	0.3000	0.1500	1 3 4 2 5	0.6000	0.7000	0.7167
4 1 3 5 2	0.0000	0.0000	0.0000	1 3 4 5 2	0.4000	0.4000	0.5667
4 1 2 3 5	0.4000	0.4000	0.2333	1 3 2 4 5	0.8000	0.9000	0.8833
4 1 2 5 3	0.2000	0.2000	0.1333	1 3 2 5 4	0.6000	0.8000	0.8333
4 1 5 2 3	0.0000	-0.1000	-0.1167	1 3 5 2 4	0.4000	0.5000	0.5833
4 1 5 3 2	-0.2000	-0.2000	-0.1667	1 3 5 4 2	0.2000	0.3000	0.4833
3 4 5 2 1	-0.4000	-0.6000	-0.4667	1 2 3 4 5	1.0000	1.0000	1.0000
3 4 5 1 2	-0.2000	-0.5000	-0.4167	1 2 3 5 4	0.8000	0.9000	0.9500
3 4 2 5 1	-0.2000	-0.3000	-0.2167	1 2 4 3 5	0.8000	0.9000	0.9167
3 4 2 1 5	0.0000	0.1000	-0.0167	1 2 4 5 3	0.6000	0.7000	0.8167
3 4 1 2 5	0.2000	0.2000	0.0667	1 2 5 4 3	0.4000	0.6000	0.7333
3 4 1 5 2	0.0000	-0.1000	-0.0833	1 2 5 3 4	0.6000	0.7000	0.7833
3 5 4 2 1	-0.6000	-0.7000	-0.5833	1 5 3 2 4	0.2000	0.3000	0.3500
3 5 4 1 2	-0.4000	-0.6000	-0.5333	1 5 3 4 2	0.0000	0.1000	0.2500
3 5 2 4 1	-0.4000	-0.5000	-0.4167	1 5 2 3 4	0.4000	0.4000	0.4333
3 5 2 1 4	-0.2000	-0.2000	-0.2667	1 5 2 4 3	0.2000	0.3000	0.3833
3 5 1 2 4	0.0000	-0.1000	-0.1833	1 5 4 2 3	0.0000	0.1000	0.2167
3 5 1 4 2	-0.2000	-0.3000	-0.2833	1 5 4 3 2	-0.2000	0.0000	0.1667