ANADOLU ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ
ANADOLU UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
Cilt/Vol.:2 - Sayı/No: 1 : 183-189 (2001)

ARAŞTIRMA MAKALESİ/**RESEARCH ARTICLE**

KRUSKAL-WALLIS İSTATİSTİĞİNİN BAĞIMSIZ PARÇALARINA AYRILMASINDA YENİ BİR DÖNÜŞÜM YÖNTEMİ

Adil KORKMAZ[1], Süleyman GÜNAY[2]

ÖZ

k kitleden çekilen örneklem büyüklükleri $n_1,n_2,...,n_k$ ve örneklem sıra sayıları toplamları $R_1,R_2,...,R_k$ olduğuna göre bu göstergelere dayalı olarak tanımlanan Kruskal-Wallis H istatistiği asimptotik olarak k-1 serbestlik derecesinde ki-kare dağılımlı olacaktır. Bu teorem Kruskal tarafından tanıtlanmıştır. Bu tanıtlamada $n=n_1+n_2+...+n_k$ olmak üzere $(n+1)\cdot H/n$ istatistiği kullanılmıştır. Bu istatistiğin bir dönüşüm sonucunda standart normal raslantı değişkenlerinin bir karesel biçimi olarak yazılabileceği ve dolayısıyla da asimptotik olarak k-1 serbestlik derecesinde ki-kare dağılımlı olacağı gösterilmiştir. Buradan da $(n+1)\cdot H/n$ istatistiğinin asimptotik olarak H istatistiğine eşitleneceğini göz önünde bulundurularak H istatistiğinin dahi k-1 serbestlik derecesinde ki-kare dağılımlı olduğu sonucu çıkarılmıştır. Bu çalışmada farklı bir dönüşüm uygulanarak doğrudan doğruya H istatistiğinin k-1 sayıda standart normal raslantı değişkeninin karesel biçimi olarak yazılabileceği ve dolayısıyla da asimptotik olarak k-1 serbestlik derecesinde ki-kare dağılımlı olacağı gösterilmiştir.

Anahtar Kelimeler: Kruskal-Wallis İstatistiği, Dönüşüm Formulü.

A NEW TRANSFORMATION METHOD FOR KRUSKAL-WALLIS STATISTIC TO DECOMPOSE INDEPENDENT PARTS

ABSTRACT

The theorem is that well-known Kruskal-Wallis H statistic has asymptotically chi-square distribution with k-1 degrees of freedom while sample sizes are $n_1,n_2,...,n_k$ and sample rank sums are $R_1,R_2,...,R_k$ from k populations. Kruskal **(1952)** has used **a transformation formula,** which decompose the statistic $(n+1)\cdot H/n$ to independent parts, to prove the theorem. Then he has argued that the statistic $(n+1)\cdot H/n$ is a quadratic form of asymptotically normal random variables. Consequently he has argued that the statistic $(n+1)\cdot H/n$ and also the statistic H has asymptotically chi-square distribution. In this study to prove the theorem by using **new transformation formula,** which divide to independent parts of only the statistic H, it has been proved that the statistic H may be directly written as a quadratic form of asymptotically standard normal random variables. Consequently it has also been showed that the statistic H has asymptotically chi-square distribution with k-1 degrees of freedom.

Key Words: Kruskal-Wallis Statistic, Transformation Formula.

1. GİRİŞ

Bir yönlü varyans analizinin parametrik olmayan biçimlerinden biri de Kruskal-Wallis istatistiği ile yapılan hipotez testidir. Bir yönlü varyans analizinde "Kitleler belli bir niteliğin düzeyleri bakımından farklılaşmakta mıdır?" sorusuna yanıt aranmaktadır (Hollander–Wolfe, 1973). Kruskal-Wallis istatistiğine dayalı hipotez testinde de benzer bir soru geçerlidir. Çünkü $K_1,K_2,...,K_k$ kitlelerinin gerçekten farklı olup olmadıkları sorulmaktadır. William H. Kruskal'ın (1952) makalesi bu sorunun yanıtını araştırmaktadır.

1 Akdeniz Üniversitesi İktisadi ve İdari Bilimler Fakültesi İktisat Bölümü ANTALYA.
2 Hacettepe Üniversitesi Fen Bilimleri Enstitüsü İstatistik Bölümü Beytepe-ANKARA.

Geliş: 24 Kasım 1999; **Düzeltme:** 09 Şubat 2000; **Kabul:** 15 Şubat 2001.

Çizelge : Örneklemlerin Sıraları.

K_1'den çekilen örneklem	K_2'den çekilen örneklem	...	K_k'den çekilen örneklem
R_{11}	R_{21}	...	R_{k1}
R_{12}	R_{22}	...	R_{k2}
...	...	...	...
R_{1n_1}	R_{2n_2}	...	R_{kn_k}
+____	+____		+____
R_1	R_2	...	R_k

2. İLK YAKLAŞIM

$K_1,K_2,...,K_k$ kitlelerinden çekilen örneklem öğelerinin birleştirilip küçükten büyüğe doğru dizilmesi sonucunda ortaya çıkan sıralar ve bunların toplamları aşağıdaki gibi olsun.

Kruskal (1952), örneklemlere ilişkin sıra ortalamaları $\overline{R_1},\overline{R_2},...,\overline{R_k}$, sıra toplamları da $R_1,R_2,...,R_k$ olmak üzere test istatistiğini şöyle tanımlamaktadır:

$$H = \frac{12}{n\cdot(n+1)} \cdot \sum_{g=1}^{k} n_g \cdot \left(\overline{R_g} - \frac{n+1}{2}\right)^2$$

$$= \frac{12}{n\cdot(n+1)} \cdot \sum_{g=1}^{k} \frac{R_g^2}{n_g} - 3\cdot(n+1) \qquad (1)$$

Kitleler homojen ise kitle ortalamaları birbirlerine yakın olacaktır. Kitle ortalamalarının birbirine yakın olması ise hepsinin $\frac{n+1}{2}$ değerine yakın olması anlamına gelir. Bu durumda H istatistiği küçük değerlere ulaşacaktır. Homojenlikten uzaklaşma durumunda ise kitle ortalamaları arasında büyük farklar ortaya çıkacağından H istatistiğinin değerini yükseltecektir. Dolayısıyla H istatistiğinin düşük değerlerinde kabul edilmesi gerekecek olan homojenlik hipotezinin, aynı istatistiğin yüksek değerlerinde reddedilmesi gerekecektir.

Wallis'in de katkıları ile biçimlendirilmiş olduğundan, (1) bağıntısı ile tanımlanan istatistik, yalnızca Kruskal'ın değil, onun yanı sıra Wallis'in de adıyla anılır. Kruskal-Wallis H istatistiğinin k-1 serbestlik derecesinde yaklaşık olarak ki-kare dağılımlı olduğu belirtilmektedir (Gibbons, 1971; Bickel-Doksum, 1977; Dixon-Massey, 1983; Sprent, 1989; Wackerly et al., 1996). Kruskal (1952), $\frac{n+1}{n}\cdot$ H diye tanımladığı istatistiğin karesel bir biçim olarak yazılabileceğini gösterdikten sonra $n\to\infty$ iken bunun da k-1 serbestlik derecesinde ki-kare dağılımlı olacağını tanıtlamakta ve buradan da H istatistiğinin aynı dağılımlı olacağı sonucunu çıkarmaktadır. Elbette $\frac{n+1}{n}\cdot$ H istatistiği bir karesel biçim olarak yazılabiliyorsa H istatistiği de yazılabilir. Üstelik H istatistiğinin asimptotik olarak k-1 serbestlik derecesinde ki-kare dağılımlı olacağını tanıtlamak daha kolay ise bu ikinci yolun daha tercih edilebilir olduğu dahi ileri sürülebilir. Ancak Kruskal (1952) bu ikinci yolu tercih etmemiştir. Sonraki dönemde de bu yaklaşım sürmüştür.

3. YENİ YAKLAŞIM

Bu çalışmanın amacı, $\frac{n+1}{n}\cdot$ H istatistiğinin değil, doğrudan doğruya H istatistiğini bir karesel biçim olarak yazıp asimtotik olarak k-1 serbestlik derecesinde ki-kare dağılımlı olacağı tanıtlamaktır. Bu amaçla $g=1,2,...,k$ olmak üzere R_g'nin varyans ve $g\neq h=1,2,...,k$ olmak üzere de R_g ile R_h arasındaki kovaryans bağıntılarını türetmek gerekmektedir.

3.1. Ortalama ve Varyans Bağıntısının Türetimi

$K_1,K_2,...,K_k$ kitlelerinden çekilen örneklemlerin sıralamadaki konum toplamlarına ilişkin vektör,

$$R'=[R_1\ R_2\ ...\ R_k]$$

olsun. Birleşik sıralı örneklemdeki konumların göstergesi olarak $i=1,2,...,n$ ve kitlelerin göstergesi olarak $g=1,2,...,k$ olmak üzere birleşik sıralı örneklemin i. konumdaki birimin K_g kitlesinin bir üyesi olması ya da olmaması ile ilgili bir raslantı değişkeni şöyle tanımlanabilir:

$$Z_{gi} = \begin{cases} 1, & \text{i.konumdaki öğe} \in K_g \\ 0, & \text{i.konumdaki öğe} \notin K_g \end{cases}$$

Z_{gi} raslantı değişkeni bir Bernoulli denemesidir. Anlaşılacağı gibi, birleşik sıralı örneklem n büyüklüğünde olduğu için K_g kitlesi açısından $Z_{g1},Z_{g2},...,Z_{gn}$ gibi Bernoulli raslantı değişkenleri olacaktır. K_g kitlesinden n_g büyüklüğünde örneklem çekildiği için anılan raslantı değişkenlerinin n_g tanesi 1, $n-n_g$ tanesi de 0'dır. Bernoulli denemesinin başarı olasılığı n_g/n olduğuna göre,

$$P(Z_{gi}=1)=n_g/n \qquad (2)$$

ve dolayısıyla da Z_{gi} raslantı değişkeninin ortalaması,

$$E(Z_{gi})=0\cdot P(Z_{gi}=0)+1\cdot P(Z_{gi}=1)=n_g/n$$

varyansı ise,

$$E(Z_{gi}i^2)=0^2 \cdot P(Z_{gi}=0)+1^2 \cdot P(Z_{gi}=1)=n_g/n,$$

$$Var(Z_{gi})=E(Z_{gi}^2)-E(Z_{gi})^2$$

bağıntılarından,

$$Var(Z_{gi}) = \frac{n_g \cdot (n - n_g)}{n^2}$$

bulunur. Bernoulli raslantı değişkenleri arasındaki kovaryans ise,

$$E(Z_{gi} \cdot Z_{gj})=P(Z_{gi}=1, Z_{gj}=1)=P(Z_{gi}=1) \cdot P(Z_{gj}=1/Zgi=1) \qquad (3)$$

bağıntısından yararlanılarak bulunabilir. Bunun için,

$$P(Z_{gj}=1/Z_{gi}=1)=(n_g-1)/(n-1) \qquad (4)$$

bağıntısı da gereklidir. (2) ve (4) bağıntıları (3) bağıntısında yerlerine konulunca,

$$E(Z_{gi} \cdot Z_{gi}) = -\frac{n_g \cdot (n - 1)}{n \cdot (n - 1)}$$

bulunur. Buradan,

$$Cov(Z_{gi}, Z_{gi}) = -\frac{n_g \cdot (n - n_g)}{n^2 \cdot (n - 1)}$$

elde edilir. Burada $Z_{g1}, Z_{g2}, \ldots, Z_{gn}$ raslantı değişkenlerinden oluşan vektör Z_g ile gösterilecek olursa,

$$Z_g' = [\, Z_{g1} \quad Z_{g2} \quad \cdots \quad Z_{gn} \,]$$

yazılabilir ve buradan da,

$$Var(Z_g) = \begin{bmatrix} Var(Z_{g1}) & Cov(Z_{g1}, Z_{g2}) & \ldots & Cov(Z_{g1}, Z_{gn}) \\ Cov(Z_{g2}, Z_{g1}) & Var(Z_{g2}) & \ldots & Cov(Z_{g2}, Z_{gn}) \\ \ldots & \ldots & \ldots & \ldots \\ Cov(Z_{gn}, Z_{g1}) & Cov(Z_{gn}, Z_{g2}) & \ldots & Var(Z_{gn}) \end{bmatrix}$$

olduğu göz önünde tutularak yukarıda ele edilen sonuçlardan,

$$Var(Z_g) = \begin{bmatrix} \frac{n_g \cdot (n - n_g)}{n^2} & -\frac{n_g \cdot (n - n_g)}{n^2 \cdot (n - 1)} & \ldots & -\frac{n_g \cdot (n - n_g)}{n^2 \cdot (n - 1)} \\ -\frac{n_g \cdot (n - n_g)}{n^2 \cdot (n - 1)} & \frac{n_g \cdot (n - n_g)}{n^2} & \ldots & -\frac{n_1 \cdot n_2}{n^2 \cdot (n - 1)} \\ \ldots & \ldots & \ldots & \ldots \\ -\frac{n_g \cdot (n - n_g)}{n^2 \cdot (n - 1)} & -\frac{n_g \cdot (n - n_g)}{n^2 \cdot (n - 1)} & \ldots & \frac{n_g \cdot (n - n_g)}{n^2} \end{bmatrix}$$

bulunur. Yapılacak düzenlemeler sonunda,

$$Var(Z_g) = \frac{n_g \cdot (n - n_g)}{n \cdot (n - 1)} \cdot \left\{ I - \frac{1}{n} \cdot \begin{bmatrix} 1 \\ 1 \\ \ldots \\ 1 \end{bmatrix} \cdot [\, 1 \quad 1 \ \ldots \ 1 \,] \right\}$$

yazılabilir. Şimdi

$$S' = [\, 1 \quad 2 \cdots n \,]$$

vektörü cinsinden

$$R_g = S' \cdot Z_g = 1 \cdot Z_{g1} + 2 \cdot Z_{g2} + \ldots + n \cdot Z_{gn}$$

biçiminde tanımlanan raslantı değişkeni, birleşik sıralı örneklemde K_g kitlesinden gelen birimlerin sıra sayıları toplamını yansıtır. Bu raslantı değişkeninin ortalaması şöyle hesaplanabilir:

$$E(R_g) = S' \cdot E(Z_g) = [\, 1 \quad 2 \cdots n \,] \cdot \begin{bmatrix} n_g/n \\ n_g/n \\ \ldots \\ n_g/n \end{bmatrix}$$

$$= 1 \cdot \frac{n_g}{n} + 2 \cdot \frac{n_g}{n} + \ldots + n \cdot \frac{n_g}{n} = \frac{n_g \cdot (n + 1)}{2}$$

Aynı raslantı değişkeninin varyansı da şöyle bulunabilir:

$$Var(R_g) = S' \cdot Var(Z_g) \cdot S = \frac{n_g \cdot (n - n_g)}{n^2 \cdot (n-1)}$$

$$\cdot [1\ 2 \dots n] \cdot \left\{ I - \frac{1}{n} \cdot \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix} \cdot [1\ 1 \dots 1] \right\} \cdot \begin{bmatrix} 1 \\ 2 \\ \dots \\ n \end{bmatrix}$$

Çarpma işleminin dağılım özelliğinden yararlanıldığında büyük ayraç içindeki terimlerden ilkinin, tamsayıların kareleri toplamı, ikincisinin de tamsayılar toplamının karesinin 1/n katı olduğu görülür. Böylece,

$$Var(R_g) = \frac{n_g \cdot (n - n_g)}{n^2 \cdot (n-1)} \cdot \left\{ \frac{n \cdot (n+1) \cdot (2n+1)}{6} - \frac{1}{n} \cdot \frac{n^2 \cdot (n+1)^2}{4} \right\}$$

yazılarak,

$$Var(R_g) = \frac{n_g \cdot (n - n_g) \cdot (n+1)}{12}, \qquad g = 1,2,\dots,k \tag{5}$$

elde edilir.

3.3. R_g'nin Asimtotik Dağılımı

$i \neq j = 1,2,\dots,n$ olmak üzere Z_{gi} ve Z_{gj} raslantı değişkenlerinin $n \to \infty$ için bağımsız olduğu şöyle tanıtlanabilir:

$$P(Z_g = 1, Z_{gj} = 1) - P(Z_{gi} = 1) \cdot P(Z_{gj} = 1)$$

$$= \frac{n_g}{n} \cdot \frac{n_g - 1}{n-1} - \frac{n_g}{n} \cdot \frac{n_g}{n}$$

$n \to \infty$ için sağ yan sıfır olacağından asimtotik olarak şu bağıntı yazılabilir:

$$\lim_{n \to \infty} P(Z_{gi} = 1, Z_{gj} = 1) = P(Z_{gi} = 1) \cdot P(Z_{gj} = 1) \tag{6}$$

Benzer bağıntılar Z_{gi} ve Z_{gj} raslantı değişkenlerinin değerlerine ilişkin öteki seçenekler için de gerçekleştirilebilir. Bu da söz konusu raslantı değişkenlerinin asimtotik olarak bağımsız oldukları anlamına gelir. Öyleyse $W_i = i \cdot Z_{gi}$ ve $W_j = j \cdot Z_{gj}$ raslantı değişkenleri de asimtotik olarak bağımsızdır. Öte yandan $W_1, W_2, \dots$ raslantı değişkenleri bağımsız olmaları durumunda Lyapunov koşulunu sağlarlar. Bu da şöyle tanıtlanabilir (Gnedenko, 1976).

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} E[W_i - E\,W_i]^{2+\delta}}{\left\{ \sum_{i=1}^{n} E[W_i - E\,W_i]^2 \right\}^{1+\frac{\delta}{2}}}$$

$$= \lim_{n \to \infty} \frac{\sum_{i=1}^{n} \left\{ i^{2+\delta} \cdot E[Z_{gi} - E(Z_{gi})]^{2+\delta} \right\}}{\left\{ \frac{n_g \cdot (n - n_g) \cdot (n+1)}{12} \right\}^{1+\frac{\delta}{2}}} \tag{7}$$

Burada oranı 1'den küçük olduğundan,

$$E[Z_g - E(Z_{gi})]^{2+\delta} = \left(0 - \frac{n_g}{n}\right)^{2+\delta} \cdot \left(1 - \frac{n_g}{n}\right) + \left(1 - \frac{n_g}{n}\right)^{2+\delta} \cdot \frac{n_g}{n} \tag{8}$$

anlatımının sonlu bir değere eşit olacağı kolayca gösterilebilir. Ayrıca

$$\sum_{i=1}^{n} i^{2+\delta} = On^{3+\delta} \tag{9}$$

anlatımının geçerliliği de kolayca tanıtlanabilir. (8) ve (9) bağıntıları (7) bağıntısında yerlerine konulursa,

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} E[W_i - E(W_i)]^{2+\delta}}{\left\{ \sum_{i=1}^{n} E[W_i - E(W_i)]^2 \right\}^{1+\frac{\delta}{2}}}$$

$$= \lim_{n \to \infty} \frac{On^{3+\delta}}{\left\{On^{3/2}\right\}^{1+\frac{\delta}{2}}} = On^{-1/2}$$

olduğu ve $n \to \infty$ için sağ yanın sıfır olduğu görülerek $W_1, W_2, \dots$ raslantı değişkenlerinin Lyapunov koşulunu sağladıkları görülür. Bu raslantı değişkenlerinin asimtotik olarak bağımsız oldukları da (6) bağıntısıyla gösterilmiştir. Öyleyse $W_1, W_2, \dots$ raslantı değişkenleri merkezi limit teoremine uyacaktır. Dolayısıyla bunların toplamından oluşan R_g raslantı değişkeni asimtotik olarak normal dağılımlı olacaktır.

3.3. Kovaryans Bağıntısının Türetimi

$g \neq h$ olmak üzere K_g ve K_h kitlelerinden çekilen örneklemlerin birleşik sıralı örneklemdeki konumlar toplamı R_g ve R_h olduğuna göre,

$$R_g = \sum_{i=1}^{n} i \cdot Z_{gi}, \tag{10}$$

$$R_h = \sum_{j=1}^{n} j \cdot Z_{hj} \quad \textbf{(11)}$$

olacaktır. (10) ve (11) bağıntılarında sol yan sol yan ile, sağ yan da sağ yan ile çarpılarak,

$$E\left(R_g \cdot R_h\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} i \cdot j \cdot E\left(Z_{gi} \cdot Z_{hj}\right)$$

elde edilir. Belli bir konumda aynı anda iki öğe olamayacağından iki Bernoulli raslantı değişkeni aynı anda 1 olamayacaktır. Dolayısıyla,

$$E\left(R_g \cdot R_h\right) = \sum_{i \neq j=1}^{n} i \cdot j \cdot E\left(Z_{gi} \cdot Z_{hj}\right)$$

olur. Buradan da,

$$E\left(R_g \cdot R_h\right) = \frac{n_g}{n} \cdot \frac{n_h}{n-1} \cdot \sum_{i \neq j=1}^{n} i \cdot j \quad \textbf{(12)}$$

bulunur. Öte yandan,

$$\sum_{i \neq j=1}^{n} i \cdot j = \frac{n \cdot \left(n^2 - 1\right) \cdot \left(3 \cdot n + 2\right)}{12} \quad \textbf{(13)}$$

olduğu kolayca tanıtlanabilir. (13) bağıntısı, (12) bağıntısında yerine konulacak olursa,

$$E\left(R_g \cdot R_h\right) = \frac{n_g}{n} \cdot \frac{n_h}{n-1} \cdot \frac{n \cdot \left(n^2 - 1\right) \cdot \left(3 \cdot n + 2\right)}{12}$$

elde edilir. Gerekli kısaltmalardan sonra,

$$E\left(R_g \cdot R_h\right) = \frac{n_g \cdot n_h(n+1) \cdot \left(3 \cdot n + 2\right)}{12}$$

sonucuna ulaşılır. Kovaryans tanımından,

$$Cov\left(R_g, R_h\right) = \frac{n_g \cdot n_h \cdot (n+1) \cdot \left(3 \cdot n + 2\right)}{12} - \frac{n_g \cdot (n+1)}{2} \cdot \frac{n_h \cdot (n+1)}{2}$$

yazılabileceğinden,

$$Cov\left(R_g, R_h\right) = -\frac{n_g \cdot n_h \cdot (n+1)}{12} \qquad g \neq h = 1, 2, \ldots, k \quad \textbf{(14)}$$

bağıntısına ulaşılır. Kruskal (1952), $\frac{n+1}{n} \cdot H$ istatistiğini bir karesel biçime dönüştürürken,

$$T_g = \sqrt{12} \cdot \frac{R_g - E\left(R_g\right)}{n^{3/2} \cdot \sqrt{n_g/n}}, \qquad g = 1, 2, \ldots, k$$

biçimindeki bir değişken dönüşümünden yararlanmış idi. $\sum_{g=1}^{k} T_g^2$'nin asimtotik olarak ki-kare dağılımlı olacağını, bunun aynı zamanda $\frac{n+1}{n} \cdot H$ değerine eşitleneceğini ve dolayısıyla da H istatistiğinin asimtotik olarak ki-kare dağılımı göstereceğini belirtmiş idi. Burada ona benzeyen, ancak çok az farklı olan bir başka değişken dönüşümünden yararlanılacaktır:

$$Y_g = \sqrt{\frac{12}{n_g \cdot n \cdot (n+1)}} \cdot \left(R_g - \frac{n_g \cdot n \cdot (n+1)}{2}\right), \quad g = 1, 2, \ldots, k$$

Y_g ile T_g arasındaki ilişkinin $Y_g = \sqrt{\frac{n}{n+1}} \cdot T_g$ biçiminde olduğu göz önünde tutulursa $\sum_{g=1}^{k} T_g^2 = \frac{n+1}{n} \cdot \sum_{g=1}^{k} Y_g^2$ bulunur. Bu bağıntıdan yola çıkılarak $\sum_{g=1}^{k} T_g^2$ istatistiğinin asimtotik olarak k-1 serbestlik derecesinde ki-kare dağılımlı olduğu tanıtlanmış ise, $\frac{n}{n+1}$ oranı asimtotik olarak 1'e eşit olduğundan, $\sum_{g=1}^{k} Y_g^2$ nin de asimtotik olarak k-1 serbestlik derecesinde ki-kare dağılımlı olduğu söylenebilir. 1952 yılındaki makalesinde Kruskal'ın izlediği yol budur. Ancak $\sum_{g=1}^{k} Y_g^2$ istatistiğinin asimtotik olarak ki-kare dağılımlı olduğu tanıtlanabiliyorsa o zaman böyle dolayılı bir yol izlemenin gereksiz olduğu kabul edilebilir. Bu ikinci yaklaşımı sergileyebilmek için $Y_1, Y_2, \ldots, Y_k$ raslantı değişkenlerine ilişkin vektörün kısa gösterimiyle işe başlamak gerekmektedir:

$$Y' = [Y_1 \; Y_2 \; \ldots \; Y_k]$$

Ayrıca,

$$\upsilon_g = n_g/n, \quad g = 1, 2, \ldots, k$$

biçiminde sabiteler tanımlansın. (5) ve (14) göz önünde tutulduğunda $Y_1, Y_2, \ldots, Y_k$ raslantı değişkenleri için yukarıda tanımlanan sabiteler türünden aşağıdaki bağıntılar yazılabilir:

$$Var(Y_g) = 1 - \upsilon_g, \qquad g = 1, 2, \ldots, k$$

$$Cov(Y_g, Y_h) = -(\upsilon_g \cdot \upsilon_h), \qquad g \neq h = 1, 2, \ldots, k$$

Bu iki bağıntıdan yararlanılarak,

$$\mathrm{Var}(Y)=\begin{bmatrix} 1-\upsilon & -\sqrt{\upsilon_1\cdot\upsilon_2} & \cdots & -\sqrt{\upsilon_1\cdot\upsilon_k} \\ -\sqrt{\upsilon_2\cdot\upsilon_1} & 1-\upsilon_2 & \cdots & -\sqrt{\upsilon_2\cdot\upsilon_k} \\ \cdots & \cdots & \cdots & \cdots \\ -\sqrt{\upsilon_k\cdot\upsilon_1} & -\sqrt{\upsilon_k\cdot\upsilon_2} & \cdots & 1-\upsilon_k \end{bmatrix}$$

yazılabilir. I, kxk-boyutlu birim matris olduğuna göre Var(Y) şöyle de dile getirilebilir:

$$\mathrm{Var}(Y)=\begin{bmatrix} \upsilon_1 \\ \upsilon_2 \\ \\ \upsilon_k \end{bmatrix}\cdot\begin{bmatrix} \upsilon_1 & \upsilon_2 & & \upsilon_k \end{bmatrix}$$

$\upsilon_1+\upsilon_2+\ldots+\upsilon_k=1$ bağıntısı göz önünde tutulduğunda Var(Y) matrisinin bakışımlı ve denkgüçlü olduğu kolayca tanıtlanabilir. Böyle bir matrisin rankı ise izine eşittir. Bu nedenle Var(Y) matrisinin rankı k-1'dir. Dolayısıyla söz konusu matrisin k sayıdaki öz değerinin k-1 tanesi 1, 1 tanesi de 0 olacaktır. Öyleyse P, Var(Y) matrisinin öz vektörlerinden oluşan dik matris olduğuna göre,

$$P'\cdot \mathrm{Var}(Y)\cdot P=\mathrm{diag}(1,1,\ldots,1,0) \qquad \textbf{(15)}$$

olacaktır. Öte yandan Kruskal-Wallis H istatistiği aşağıdaki gibi yazılabilir:

$$H=Y'\cdot Y$$

$P\cdot P'=I$ olduğundan yukarıdaki bağıntı,

$$H=Y'\cdot P\cdot P'\cdot Y$$

biçimine dönüştürülebilir. $U=P'\cdot Y$ denilecek olursa, bir yandan

$$H=U'\cdot U$$

öbür yandan da,

$$\mathrm{Var}(U)=P'\cdot \mathrm{Var}(Y)\cdot P=\mathrm{diag}(1,1,\ldots,1,0)$$

bağıntıları elde edilir. Bütün bunlardan çıkan sonuçlar şöyle özetlenebilir: $R_1,R_2,\ldots,R_k$ raslantı değişkenleri asimtotik olarak normal dağılım gösterdiğinden bunların doğrusal dönüşümüyle elde edilen $Y_1,Y_2,\ldots,Y_k$ ve dolayısıyla da -sabit değilseler- $U_1,U_2,\ldots,U_k$ raslantı değişkenleri de asimtotik olarak normal dağılımlıdır. (15) sayılı bağıntı gereğince bu ikincilerden $U_1,U_2,\ldots,U_{k-1}$ raslantı değişkenlerinin bağımsız ve standart biçimde; varyansı sıfır olan U_k raslantı değişkeninin ise sıfıra özdeş olduğu söylenebilir. Öyleyse $U_1^2+U_2^2+\ldots+U_{k-1}^2$ biçiminde olmak üzere k-1 sayıda bağımsız standart normal raslantı değişkeninin kareleri toplamı olarak yazılabilecek olan $H=U'\cdot U$ istatistiği asimtotik olarak k-1 serbestlik derecesinde ki-kare dağılımlıdır.

4. SONUÇ

Görüldüğü gibi, $n\to\infty$ iken H istatistiğinin ki-kare dağılımlı olacağı, dolayılı bir yol izlenmeksizin de tanıtlanabilmektedir. Bu nedenle H istatistiğinin $n\to\infty$ iken ki-kare dağılımlı olduğunu tanıtlamak amacıyla $\frac{n+1}{n}\cdot H$ istatistiğinden yola çıkmanın uzun bir yol olduğu söylenebilir. İstatistik yazınında bilinen sonuçları farklı yollardan bulma geleneği vardır. Bu farklı yollar bilinen sonuçları daha yalın, kolay ya da açık biçimde anlama konusunda etkili olabildiklerinde boş bir çaba olmaktan da uzak görünürler. Kruskal'ın izlediği yolu kısaltan bu çalışma da, bilinen sonuçları böyle bir anlayışla yeniden türetme geleneğinin bir uzantısı olarak değerlendirilebilir.

KAYNAKÇA

Bickel, P. J. ve Doksum, K. A. (1977). *Mathematical Statistics,* Prentice Hall, s. 344-406, New Jersey.

Dixon, W. J. ve Massey, F. J. (1983). *Introduction to Statistical Analysis,* McGraw-Hill, s. 395-397, New York.

Gibbons, J. D. (1971). *Nonparametric Statistical Inference,* McGraw Hill, s. 6, 149, Tokyo.

Gnedenko, B. V. (1976). *The Theory of Probability,* Mir Publishers, s. 76, Moscow.

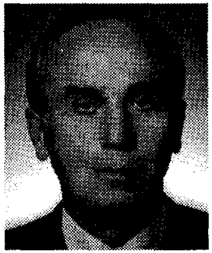
Hollander, M. ve Wolfe, A. D. (1973). *Nonparametric Statistical Methods,* John Wiley and Sons, New York.

Kruskal, W. H. (1952). *"A Nonparametric Test for the Several Sample Problem", The Annals of Mathematical Statistics,* 23, 525-540.

Sprent, P. (1989). *Applied Nonparametric Statistical Methods,* Chapman and Hall, 2. 113, London.

Wackerly, D. D., Mendenhall, ve W.-Scheaffer, R.L. (1996). *Mathematical Statistics with Applications,* Duxbury Press, s. 290, 671-672, Belmont.

Adil Korkmaz, 1956 yılında Trabzon'da (Yomra'da) doğdu. İstanbul Teknik Üniversitesinin Elektrik Fakültesini ve İstanbul Üniversitesinin İktisat Fakültesini bitirdi. Ankara Üniversitesi Sosyal Bilimler Enstitüsü İktisat Anabilim Dalında yüksek lisans, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalında doktora derecesi aldı. Halen 2000 yılında göreve başladığı Akdeniz Üniversitesinde çalışmaktadır.



Süleyman Günay, 1949 yılında Düzce'de doğdu. Ankara Üniversitesi Matematik Bölümü'nden 1968 yılında mezun oldu. Aynı yıl, Hacettepe Üniversitesi İstatistik Bölümü'nde asistan olarak göreve başladı. 1972 yılında İstatistik dalında doktora, 1978'de de doçentlik unvanını aldı. 1980 yılında İstatistik Bölümü Başkanlığı'na atandı. TÜBİTAK doktora sonrası bursuyla, Warwick Üniversitesi'nde bir yıl çalıştı. Profesörlüğe yükseltildiği 1988 yılında, Fen Fakültesi Dekan Yardımcılığı'na atandı. *Hacettepe Natural Sciences* and *Engineering* dergisinde üç yıl boyunca editörlük yaptı. Halen, İstatistik bölümünde, Uygulamalı İstatistik Anabilim Dalında Öğretim Üyesi olarak görev yapmaktadır.