

2 x 2 Tablolarında Bazı Örnek Genişlikleri ve I. Tip Hata Seviyeleri (α) İçin Kappa (K) İstatistiğine Ait Ampirik Olarak Gerçekleşen Kritik Değerler

Sıddık KESKİN¹

Geliş Tarihi: 24.06.2003

Özet: Kappa istatistiği, iki yönlü tablolarda uyum ölçüsü olarak kullanılmaktadır. Bu çalışmada, 2 x 2 tablolarında, bazı α (I. Tip hata) değerleri için yaygın olarak karşılaşılan örnek genişliklerine göre Kappa istatistiğinin kritik tablo değerlerinin elde edilmesi amaçlanmıştır. Bu amaçla, simülasyon metodu ile kesikli dağılım gösteren populasyondan örnek genişliği 6 ile 300 arasında değişen örnekler alınmıştır. Bu örneklerden 1.000.000 (bir milyon) simülasyon denemesi sonucunda, başlangıçta kararlaştırılan I. Tip hata; $\alpha = 0.25, 0.20, 0.10, 0.05$ ve 0.01 olduğunda, Kappa istatistiği için ampirik olarak gerçekleşen kritik değerler edilmiştir.

Anahtar Kelimeler : kategorik değişken, kappa katsayısı, uyum, kritik değer, I. tip hata

The Empirically Realized Critical Values for Kappa (K) Statistic for Several Sample Size and Type I Error Levels (α) in 2 x 2 Cross – Tables

Abstract: Kappa statistic (coefficient) is used for measurement of agreement in two-way tables. The aim of this study is to obtain empirically critical values of the Kappa statistic for several type I error rates (α values) and common sample size in 2 x 2 tables. For this reason, sample size ranged from 6 to 300 were drawn from discontinuous (discrete) population by the simulation method. When the type I error rates (α) are assumed as 0.25, 0.20, 0.10, 0.05 and 0.01, empirically realized critical values for Kappa statistics were obtained by the end of 1.000.000 (one million) simulation trials.

Keywords : categorical variable, kappa coefficient, agreement, critical value, type I error

Giriş

Üzerinde durulan özellik veya özellikler bakımından yeni bilgiler elde etmek amacıyla yapılan araştırma ve denemelerde, istatistik analizlerin doğru ve güvenilir olabilmesi için sayıların ya da rakamların çalışmaya konu olan deney ünitelerine uygun bir şekilde verilmesi gerekmektedir. Kısaca ölçme olarak adlandırılan rakamların deney ünitelerine verilmesi işlemi, üzerinde durulan özelliğe göre farklı şekillerde yapılmaktadır. Ölçmenin yapılışına göre de değişkenleri ölçülebilen değişkenler (measurement variables), sıralı değişkenler (ranked, ordinal variables) ve sınıflanan ya da kategorik değişkenler olmak üzere üç gruba ayırmak mümkündür (Sokal ve Rohlf 1995).

Kategorik değişken veya değişkenlerin kategorilerinin (seviyelerinin) sayısı, örnek genişliğinden az olduğu durumlarda, bu değişkenleri iki yönlü (yanlı) tablolarda özetlemek uygun bir yaklaşım olmaktadır. X ve Y gibi iki kategorik değişken, sırası ile i ve j adet seviye içerdiğinde, elde edilecek iki yönlü tablo $i \times j$ adet sınıfa sahip olacaktır. İki yönlü tabloların analiz edilmesinde, genellikle tabloda yer alan kategorik değişkenlerin birbirinden bağımsız olup olmadığı incelenir. İki (kategorik) değişkenin birbirinden bağımsız olup olmadığını belirlemek amacıyla, yaygın olarak kullanılan istatistikler arasında; χ^2 , Pearson

C, Phi (Φ) ve Cramer V gibi istatistikler sayılabilir. İki yönlü tablolarda bağımlılığın özel bir hali olan uyumun (agreement) belirlenmesinde ise en yaygın kullanılan istatistik Kappa (K) istatistiğidir. İlk olarak Cohen (1968) tarafından geliştirilen bu istatistik Cohen Kappa istatistiği olarak da bilinir. Kappa istatistiği, iki yönlü tablolarda satırda ve sütunda yer alan değişkenlerin seviyelerinin sayısı eşit olduğu durumda, yani $2 \times 2, 3 \times 3, \dots, k \times k$ olduğu durumlarda, iki değişken arasında uyum ölçüsü olarak kullanılmaktadır. Bu istatistik, aynı nesnelere (deney ünitelerinin) bazı özellik veya özellikleri dikkate alınarak iki farklı hakem metot veya eksper tarafından değerlendirilerek sınıflandırılması sonucunda; hakemler, metotlar veya eksperler arasındaki uyumu belirlemede veya aynı hakemin ya da eksperin farklı zamanlarda veya farklı yerlerde aynı nesnelere için yapmış olduğu sınıflandırmalarda uyum ölçüsü olarak kullanılabilir. Bunun yanı sıra, gerçek sınıfları belli olan nesnelere bazı özellikleri dikkate alınarak; hakem, metot ya da eksper tarafından sınıflandırılması sonucunda güvenilirlik (reliability) ölçüsü olarak da kullanılabilir (Keskin 2001). Kappa istatistiği, -1 ile 1 arasında değerler alan bir istatistik olup, bu istatistiğin 0' dan küçük olması; uyumun olmadığını, 1 olması ise tam bir uyumun olduğunu belirtir.

¹ Yüzüncü Yıl Üniv. Ziraat Fak. Zootekni Bölümü - Van

Cohen (1968) büyük örneklerde Kappa istatistiğine ait örneklem dağılımın yaklaşık olarak normal dağılım gösterdiğini ve böylece normal dağılım temeline dayalı olarak, Kappa istatistiğine ait hipotez testi yapılabileceğini belirtmiştir. Ancak, büyük örneğin ne kadar olduğuna ilişkin bilgi vermemiştir. Fleiss ve ark (1969), Kappa istatistiği için Cohen (1960, 1968) tarafından yaklaşık olarak verilen standart sapmaya dayalı yapılacak hipotez testlerinin ve hesaplanan güven aralıklarının tutucu (conservative) olduğunu belirtmişlerdir. Ayrıca, Fleiss ve ark. (1979) daha önce Cohen (1960, 1968) tarafından verilen Kappa istatistiğine ait standart sapmanın hatalı olduğunu ve bunun Landis ve Koch (1977a) tarafından yapılan bir simülasyon çalışması ile de doğrulandığını belirtmişlerdir. Bu gibi nedenlerden dolayı, Kappa istatistiğinin değerlendirilmesinde; hipotez testi yerine daha çok subjektif yaklaşım (0-0.10; Hiç yok, 0.11-0.40; Zayıf, 0.41-0.60; Belirgin, 0.61-0.80; Güçlü, 0.81-1.00; Mükemmel veya 0-0.20; Zayıf, 0.21-0.40; Düşük, 0.41-0.60; Belirgin, 0.61-1.00; Mükemmel (Güçlü)) tercih edilmiştir. Ancak, Rigby (2000) Kappa istatistiğinin hangi değeri için "iyi bir uyum" var denilebileceğindeki belirsizliğe dikkat çekerek, yukarıdaki sınıflandırmanın tamamıyla subjektif olduğunu ve bu yaklaşım ile verilen kararların, örnek genişliğine ve iki yönlü tablonun satır ve sütun sayısına bağlı olmaksızın, doğrudan doğruya araştırmacıya bağlı olarak değişebileceğini vurgulamıştır. Her ne kadar son yıllarda yapılan çalışmalarda; (Donner 1998, Lui ve Kelly 1999, Eriksson ve ark. 2000, Blackman ve Koval 2000) Kappa istatistiği için bazı yöntemlerle güven aralığı hesaplanmışsa da, bu çalışmalarda verilen güven aralıklarının sadece bazı özel koşullarda tatminkar sonuçlar verdiği belirtilmiş ve bu çalışmalarda Kappa istatistiği için hipotez testinden kaçınılmıştır.

Bu çalışmada; sadece 2 x 2 tablolarında ve bazı α (I. Tip hata) seviyelerinde Kappa istatistiğine ait önem kontrolü için, örnek genişliğine bağlı olarak araştırmacılara yardımcı olabilecek, kritik alanların başladığı tablo değerlerinin ampirik olarak geliştirilmesi amaçlanmıştır.

Materyal ve Yöntem

Bu çalışmanın materyalini IMSL LIBRARY desteği ile FORTRAN programlama dilinde yazılmış simülasyon programı ile üretilen tesadüf sayıları oluşturmaktadır (Anonymous 1994). Bu tesadüf sayıları kesikli (integer) sayılar olarak üretilmiş ve üretilen bu sayılar 2 x 2 tablosuna aktarılmıştır. 2 x 2 tablosunda örnek genişliği 6 ile 300 arasında muhtelif değerler olarak belirlenmiştir. Belirlenen her bir örnek genişliği için 1.000.000 (bir milyon) simülasyon denemesi yapılarak, Kappa istatistiği hesaplanmış ve bu istatistik için $\alpha = 0.25, 0.20, 0.10, 0.05$ ve 0.01 'deki kritik tablo değerleri oluşturulmuştur. Bu değerler oluşturulurken, her bir örnek genişliği için hesaplanan Kappa istatistikleri küçükten büyüğe doğru sıraya dizilmiş ve en büyük 0.25, 0.20, 0.10, 0.05 ve 0.01' lik alanların başladığı değerler, kritik tablo değeri olarak kabul edilmiştir. Çalışma, sadece 2 x 2 çapraz (iki yönlü) tablosu ile sınırlı tutulmuştur.

Çalışmada Kappa istatistiği;

$$K = \frac{P_G - P_B}{1 - P_B} \quad (1)$$

eşitliğine göre hesaplanmıştır. (1) no'lu eşitlikte; P_G ve P_B , iki yönlü tablolarda sırası ile gözlenen ve beklenen olasılıkları göstermekte olup bu değerler;

$$P_G = \frac{\sum_i G_{ij}}{N} \quad \text{ve} \quad P_B = \frac{\sum_i R_i C_i}{N^2} \quad \text{eşitliklerine göre}$$

hesaplanmıştır. Bu eşitliklerde; N tablodaki toplam gözlem sayısı olmak üzere; G_{ij} değeri: i. satır ve j. sütundaki gözlenen frekansı, R_i : değeri i. satırdaki toplam frekansı ve C_j değeri ise j. sütunun toplam frekansı göstermektedir.

Bulgular ve Tartışma

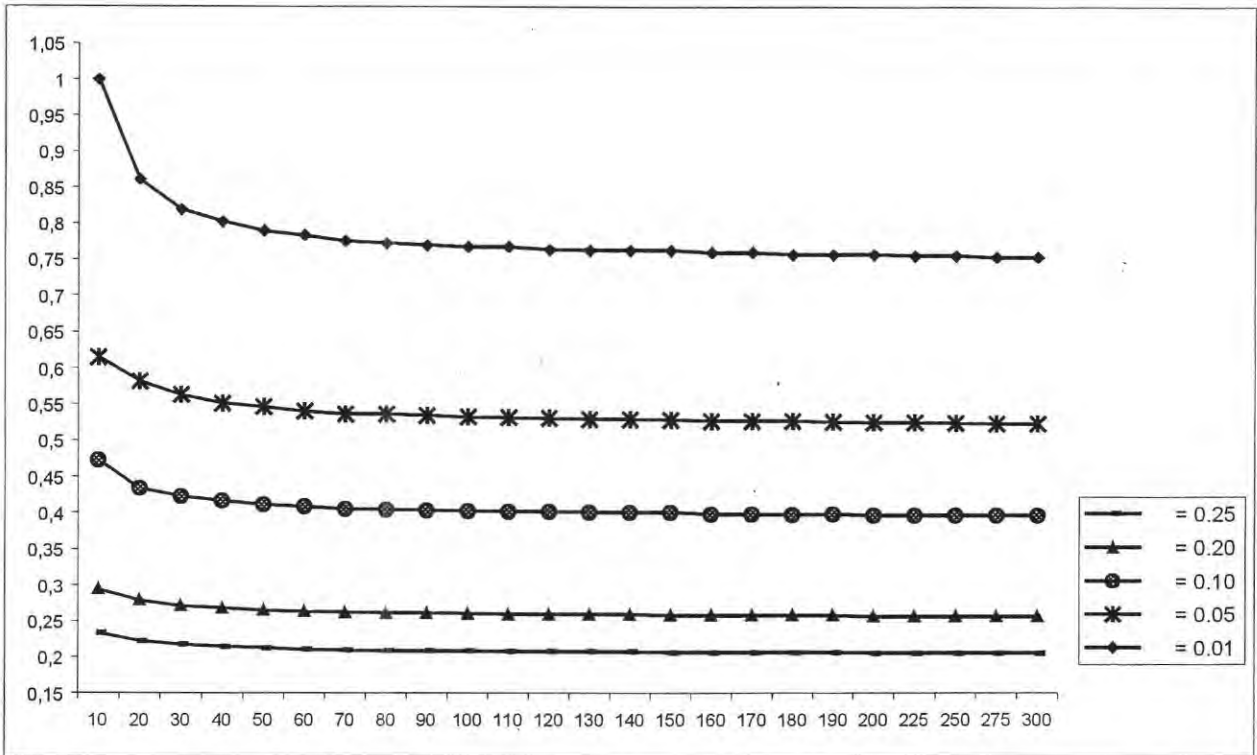
Bazı örnek genişliklerinde; $\alpha = 0.25, 0.20, 0.10, 0.05$ ve 0.01 'de Kappa istatistiği için ampirik olarak gerçekleşen kritik değerler Çizelge 1' de verilmiştir. Çizelge 1 incelendiğinde; örnek genişliğinin 6' dan başlayarak 300' e kadar artırıldığı görülür. $\alpha = 0.25$, yani I. tip hata (gerçekte kabul edilmesi gereken test hipotezinin, hipotez kontrolü sonucunda ret edilme olasılığı) %25 olarak alındığında; Kappa istatistiğine ait ampirik olarak gerçekleşen kritik tablo değerleri; 0.250 ile 0.205 arasında değişirken, $\alpha = 0.20$ 'ye düştüğünde gerçekleşen kritik değerler 0.333 ile 0.256 arasında değişmektedir. α 'nın %10' a düşmesi ile birlikte, Kappa istatistiğine ait ampirik olarak gerçekleşen kritik değerlerde değişim genişliği belirgin bir şekilde artmaktadır (0.500-0.396).

Ele alınan α değerlerinde örnek genişliğine göre Kappa istatistiği için ampirik olarak gerçekleşen kritik değerlerin değişimini daha iyi görebilmek amacıyla, Şekil 1 incelendiğinde; en geniş değişim aralığının $\alpha = 0.01$ 'de gerçekleştiği, en dar değişim aralığının ise $\alpha = 0.25$ 'te gerçekleştiği görülür. Dolayısıyla, Kappa istatistiğine ait kritik değerlerin değişim genişliğinin, α 'nın artması ile birlikte, daraldığı söylenebilir.

Sosyal bilimlerde yapılan bilimsel çalışmalarda; genellikle başlangıçta kararlaştırılan I. tip hatanın %10 olarak alındığı kabul edilirse; örnek genişliğinin 6 olması durumunda bile, 0.50 olarak hesaplanan Kappa istatistiğinin %10 düzeyinde istatistik olarak önemli (anlamlı) olduğu görülür. Şüphesiz ki, bilimsel çalışmalarda, örnek genişliğinin 6 olması durumunda, Kappa istatistiği hesaplamak ve buna göre akıl yürütmek doğru bir yaklaşım değildir. Örnek genişliğinin 50 olması durumunda; %41.1 (0.411) ve daha büyük olan Kappa istatistikleri %10 düzeyinde istatistik olarak önemli bulunurken, örnek genişliğinin 100 olması durumunda %40.2 ve daha büyükler, 200 olması durumunda ise %39.6 ve daha büyük olan Kappa istatistikleri %10

Çizelge 1. Bazı örnek genişliklerinde Kappa istatistiği için $\alpha = 0.25, 0.20, 0.10, 0.05$ ve 0.01 'deki gerçekleşen kritik değerler

N	$\alpha = 0.25$	$\alpha = 0.20$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
6	0.250	0.333	0.500	0.667	1.000
8	0.233	0.314	0.491	0.641	1.000
10	0.233	0.294	0.472	0.615	1.000
12	0.233	0.290	0.461	0.607	1.000
14	0.232	0.286	0.450	0.602	1.000
16	0.224	0.284	0.444	0.595	0.936
18	0.222	0.282	0.438	0.587	0.871
20	0.222	0.279	0.434	0.582	0.861
22	0.222	0.276	0.431	0.576	0.851
24	0.220	0.274	0.428	0.573	0.842
26	0.218	0.273	0.426	0.569	0.833
28	0.218	0.270	0.420	0.559	0.815
30	0.217	0.271	0.423	0.563	0.819
32	0.216	0.270	0.420	0.559	0.815
34	0.215	0.269	0.418	0.556	0.811
36	0.215	0.269	0.417	0.554	0.807
38	0.214	0.269	0.417	0.553	0.803
40	0.214	0.268	0.416	0.551	0.802
42	0.213	0.267	0.415	0.550	0.800
44	0.212	0.266	0.413	0.548	0.796
46	0.212	0.265	0.412	0.546	0.792
48	0.212	0.265	0.411	0.546	0.790
50	0.212	0.265	0.411	0.546	0.789
54	0.212	0.264	0.410	0.544	0.788
56	0.212	0.263	0.409	0.543	0.787
58	0.211	0.263	0.409	0.542	0.785
60	0.210	0.263	0.408	0.540	0.783
62	0.210	0.263	0.407	0.539	0.782
64	0.210	0.262	0.406	0.538	0.781
66	0.210	0.262	0.406	0.537	0.778
68	0.209	0.262	0.405	0.536	0.775
70	0.209	0.262	0.405	0.536	0.775
72	0.209	0.261	0.405	0.536	0.775
74	0.209	0.261	0.404	0.536	0.774
76	0.209	0.261	0.404	0.536	0.774
80	0.208	0.261	0.404	0.536	0.772
90	0.208	0.261	0.403	0.534	0.769
100	0.208	0.260	0.402	0.532	0.767
110	0.207	0.259	0.401	0.531	0.767
120	0.207	0.259	0.401	0.530	0.763
130	0.207	0.259	0.400	0.529	0.762
140	0.207	0.259	0.400	0.529	0.762
150	0.206	0.258	0.400	0.528	0.762
160	0.206	0.258	0.398	0.526	0.759
170	0.206	0.258	0.398	0.526	0.759
180	0.206	0.258	0.397	0.526	0.756
190	0.206	0.258	0.398	0.525	0.756
200	0.205	0.256	0.396	0.524	0.756
225	0.205	0.256	0.396	0.524	0.754
250	0.205	0.256	0.396	0.523	0.754
275	0.205	0.256	0.396	0.522	0.752
300	0.205	0.256	0.396	0.522	0.752



Şekil 1. Bazı α değerlerinde Kappa istatistiği için ampirik olarak gerçekleşen kritik değerlerin örnek genişliğine göre değişimi

düzeyinde istatistik olarak önemli olmaktadır. α 'nın % 5'e düşmesi ile birlikte; belirlenen örnek genişliklerinde Kappa istatistiğine ait ampirik olarak gerçekleşen kritik değerler 0.667 ile 0.522 arasında değişmektedir.

$\alpha = 0.01$ olduğunda, yani, I. tip hata % 1 olarak alındığında; örnek genişliği 14' ten az olduğu durumlarda, ancak 1.00 (%100) olarak hesaplanan Kappa istatistikleri % 1 düzeyinde istatistik olarak önemli olmaktadır. Örnek genişliğinin artması ile birlikte ampirik olarak gerçekleşen kritik değerler düşmekte ve örnek genişliği 300 olduğu durumda kritik değer, 0.752 olarak gerçekleşmektedir. Biyolojik çalışmalarda, başlangıçta kararlaştırılan I. tip hatanın genellikle %5 olarak kabul edildiği düşünülürse; yaklaşık olarak 0.67' den daha büyük olarak hesaplanan Kappa istatistiklerinin % 5 düzeyinde istatistik olarak önemli olduğu söylenebilir. Ayrıca, örnek genişliği 50' nin altına düşmediği sürece, 0.789 ve daha büyük olan Kappa istatistiklerinin % 1 düzeyinde istatistik olarak önemli oldukları da dikkat çekmektedir.

Örnek genişliğine bağlı olmadan subjektif değerlendirme kriterleri dikkate alındığında; hesaplanan Kappa istatistiği; 0.41-0.60 arasından ise "Belirgin" bir uyumun olduğundan söz edilmektedir. Oysa ki, Çizelge 1 incelendiğinde; örnek genişliği 16 'dan büyük olduğu durumda, 0.60 olarak hesaplanan Kappa istatistiğinin, % 5 düzeyinde önemli olduğu görülür. Benzer şekilde, yine bu kriterlere göre "Mükemmel" bir uyumdan söz etmek için

Kappa istatistiğinin 0.81'den daha büyük olması gerekmektedir. Oysa ki, örnek genişliği 50'den büyük olduğu durumlarda, %79 olarak hesaplanan Kappa istatistiği %1 düzeyinde önemli olmaktadır.

Rigby (2000) Kappa istatistiğine ait asimtotik standart hatanın (sapmanın) hesaplanması için, örnek genişliğine bağlı olarak farklı istatistik eşitliklerin bulunduğu dikkat çekerek, Kappa istatistiğine ait güven aralığında; alt sınır değerinin sıfırdan uzaklaşması durumunda, Kappa istatistiğinin sıfırdan önemli derecede farklı olduğunu, alt sınır değerinin sıfıra yaklaşması durumunda ise Kappa istatistiğinin sıfırdan önemli derecede farklı olmadığını vurgulamıştır. Ayrıca, 81 bireylik bir örnekte (2 x 2 tablosu) 0.80 olarak hesaplanmış olduğu Kappa istatistiğine ait alt ve üst sınırların, sırasıyla; 0.67 ve 0.93 olduğunu ve bu katsayının sıfırdan önemli derecede farklı olduğunu belirtmiştir.

Blackman ve Koval (2000) Kappa istatistiği için BK, FCE ve Jackknife (Garner metodu) olmak üzere 3 adet güven aralığı bulma yöntemlerini karşılaştırmışlar ve her koşulda uygun olan yöntemin bulunmadığını belirtmişlerdir.

Dikkat edileceği üzere, Kappa istatistiği için güven aralığı bulmada, kullanılacak standart hatanın asimtotik oluşu ve farklı şekillerde hesaplanabilmesi, Kappa istatistiğinin güven sınırlarını değiştirebilmektedir. Bunun

yanı sıra, bulunan güven sınırlarına göre karar vermede yine sıfıra yakınlığın ölçüsü belirtilmemekte ve her koşulda kullanılabilirlik güven aralığı yönteminin bulunmayışı vurgulanmaktadır.

Sonuç

İki yönlü tablolarda, bağımlılığın özel hali olan uyumun belirlenmesinde, yaygın olarak kullanılan Kappa istatistiğine ait önem kontrolünde, örnek genişliğine bağlı olmaksızın, sübjektif yaklaşımla (Mükemmel, Güçlü, Belirgin vs.) karar vermek doğrudan doğruya araştırmacıya bağlı olmaktadır. Bunun sonucu olarak da, bu sınıflandırmanın alt ve üst sınırları değişebilmekte ve literatürde farklı aralıklar belirtilmektedir. Bunun yanı sıra, güven aralığına dayalı yapılan hesaplamalara göre verilecek kararlar da standart sapmaya göre değişebilmektedir. Ayrıca, Kappa istatistiğinin hangi durumda önemli, hangi durumda ise önemsiz olduğuna karar vermede belirsizlikler bulunmaktadır. Bu gibi nedenlerden dolayı; 2 x 2 tablolarında, bazı α değerleri için yaygın olarak karşılaşılan örnek genişliklerine göre düzenlenmiş olan bu tablo değerlerinin kullanılmasının, Kappa istatistiğinin sıfırdan önemli derecede farklı olup olmadığının test edilmesinde, araştırmacılara büyük kolaylıklar sağlayacağı düşünülmektedir.

Satır ve sütun sayısı 2' den büyük olan iki yönlü tablolarda; (3 x 3, 4 x 4, ..., k x k) Kappa istatistiğine ait kritik değerlerin nasıl değiştiği başka çalışmalara bırakılmıştır.

Kaynaklar

- Anonymous, 1994. IMSL MATH/LIBRARY FORTRAN subroutines for mathematical applications. Vol. 1, 2, Visual Numerics Inc. USA.
- Blackman, N. J. M. and J. J. Koval, 2000. Interval estimation for Cohen's kappa as a measure of agreement, *Statistics in Medicine*, (19) 723-741.

Cohen, J. 1968. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit, *Psychological Bulletin*, 70 (4) 213- 220.

Donner, A. 1998. Sample size requirements for the comparison of two or more coefficient of inter - observer agreement, *Statistics in Medicine*, (17) 1157-1168.

Eriksson, E. M., M. Mokhtari, L. Pourmotamed, L. Holmdahl and H. Eriksson, 2000. Inter- rater reliability in a resource oriented physiotherapeutic examination, *Physiotherapy Theory and Practice*, (16) 95-103.

Fleiss, J. L., J. Cohen and B. S. Everitt, 1969. Large sample standard errors of kappa and weighted kappa, *Psychological Bulletin*, 72 (5) 323 - 327.

Fleiss, J. L., J. C. M. Nee, and J. R. Landis, 1979. Large sample variance of kappa in the case of different sets of raters *Psychological Bulletin*, 86 (5) 974 - 977.

Keskın, S. 2001. İki yönlü (contingency) tablolarda kappa (K) istatistiğinin kullanımı, *Biyoteknoloji (Kükem) Dergisi*, 25 (1) 53-57.

Landis J. R. and G. G. Koch, 1977a. The measurement of observer agreement for categorical data, *biometrics*, (33) 159-174.

Lui, K. J. and C. Kelly, 1999. A note on interval estimation of kappa in a series of 2 x 2 tables, *Statistics in Medicine*, (18) 2041-2049.

Rigby, A. S., 2000. Statistical methods in epidemiology. v. towards an understanding of the kappa coefficient, *Disability and Rehabilitation*, 22 (8) 339-344.

Sokal, R. R. and F. J. Rohlf, 1995. *Biometry*. W. H. Freeman and Company. 887p., New York.

İletişim adresi:

Siddik KESKİN
Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi
Zootekni Bölümü-Van
E-mail: skeskin973@hotmail.com