

## Hayvancılıkta Kategorik Verilerde Ki-Kare ve G Testi

Ecevit EYDURAN<sup>1</sup>Taner ÖZDEMİR<sup>1</sup>Mürsel KÜÇÜK<sup>2</sup><sup>1</sup>Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi, Zootehni Anabilim Dalı, VAN<sup>2</sup>Yüzüncü Yıl Üniversitesi, Veteriner Fakültesi, Zootehni Anabilim Dalı, VAN

### ÖZET

Bu çalışmada kategorik hale dönüştürülen kirli yapağı verimi, transferin tipi, hemoglobin tipi ve cinsiyet özellikleri arasındaki ilişkiler, G ve Ki-kare istatistikleri ile belirlenmeye çalışılmıştır. Kirli yapağı verimi ile transferin tipi özellikleri G istatistiğine göre birbirine bağımlı ( $P<0.05$ ) bulunurken, ki-kare istatistiğine göre birbirinden bağımsız bulunmuştur. Diğer özellikler için ki-kare ve G istatistiği aynı sonuçları vermiştir. Sonuç olarak, iki yönlü tablolarda beklenen frekansların 5'ten az olması durumunda ki-kare istatistiğine ilaveten, G istatistiğinin de hesaplanarak buna göre karar vermenin daha uygun olacağı vurgulanmıştır.

**Anahtar Kelimeler :** G testi, Ki-kare Testi, İki yönlü tablo

### Chi-Square and G test in Categorical Data in Animal Science

### SUMMARY

In this study, it was studied to determine the relationships among dirty fleece yield, transferrin type, hemoglobin type and sex by using G and Chi-square statistics. In contrast to the Chi-square statistic, it was found that dirty fleece and transferrin type are dependent according to G statistic ( $P<0.05$ ). As a result, it was emphasized that it should be more suitable decision by G statistic as well as Chi square statistic when expected frequencies are less than 5 in two way tables.

**Key words:** G test, Chi-square test, two way tables

### GİRİŞ

Üzerinde durulan özellik veya özellikler bakımından yeni bilgiler elde etmek amacıyla yapılacak olan araştırma ve denemelerde, istatistik analizlerin doğru ve güvenilir bir şekilde yapılabilmesi için sayıların ve rakamların çalışmaya konu olan nesnelere uygun bir şekilde verilmesi gerekir. Kısaca ölçme olarak adlandırılan sayıların nesnelere verilmesi işlemi, üzerinde durulan özelliğe göre farklı şekillerde yapılmaktadır [8].

Ölçmenin yapılışına göre değişkenler genel olarak ölçülebilen değişkenler (measurement variables), sıralı değişkenler (ranked variables), ve kategorik değişkenler (nominal variables) olmak üzere üç gruba ayrılırlar [12,13]. Ölçülebilen değişkenler ise kendi arasında sürekli değişkenler (continuous variables) ve kesikli değişkenler (discontinuous variables) olmak üzere iki grup altında incelenir. Sürekli değişkenler, teorik olarak belirli bir tanım aralığında mümkün olan her değeri alabilen değişkenler olup birbirinden kesin sınırlarla ayrılmazken, kesikli değişkenler sadece tam sayılı değerleri alırlar ve birbirinden kesin sınırlarla ayrılırlar [8].

Canlı ağırlık, boy, sütteki yağ miktarı, koyunlardaki yapağı verimi, sürekli değişkenlere örnek olarak verilebilirken, ailedeki çocuk sayısı, cinsiyet, böcekteki segment sayısı kesikli değişkenlere örnek olarak verilebilir. Sıralı değişkenlerde; rakamlar veya sayılar, bireyler yada nesnelere, üzerinde durulan özelliğe sahip oluş derecelerine göre küçükten büyüğe yada büyükten küçüğe doğru verilir. Ailelerin gelir durumu (yüksek, orta, düşük), bir testin zorluğu (zor,

kolay, çok kolay) sıralı değişkenlere örnek olarak verilebilir. Kategorik değişkenlerde ise rakamlar birey yada nesnelere tamamen keyfi olarak verilir. Örneğin cinsiyet, ırk, doğum tipi gibi özellikler kategorik değişkenler grubunda yer alırlar.

Değişken yada değişkenlerin kategorilerinin sayısının değişken sayısından az olduğu durumlarda, kategorik verileri tablolar halinde özetlemek uygun bir yaklaşımdır. A ve B gibi iki kategorik değişken sırasıyla i ve j adet seviye içerdiği durumda nesnelere yada bireyleri sınıflandırmada mümkün olan sınıf sayısı  $i \times j$  kombinasyonu kadardır. Dolayısıyla  $i \times j$  adet sınıfa dönüşen nesnelere ait sayıların verilmiş olduğu tablolara iki yönlü (çapraz) tablolar adı verilir.

Kategorik değişkenlerin yada kategorik hale dönüştürülebilen değişkenlerin özetlenmesinde genellikle iki yönlü tablolar kullanılmaktadır. İki yönlü tablolarda yer alan kategorik değişkenler arasındaki ilişkiyi belirlemede yaygın olarak kullanılan istatistikler arasında; Fisher'in Kesin olasılık testi, Ki kare testi, Oran testi, G testi, Log-doğrusal modeller sayılabilir.

Bu çalışmada, cinsiyet, transferin tipi, hemoglobin tipi ve kategorik hale dönüştürülmüş kirli yapağı verimi özellikleri arasındaki ilişkilerin G testi ve Ki-kare testi ile irdelenmesi amaçlanmıştır.

### MATERYAL ve METOT

Çalışmanın materyalini Doğu ve Güneydoğu Anadolu Bölgesinde yetiştirilen 20 baş renkli Tiftik keçisi oluşturmuştur. Keçilerden alınan kan örnekleri Etlik Merkez Veteriner Kontrol ve Araştırma Enstitüsü'nde analiz yapılarak Transferin ve Hemoglobin değerlerine bakılmıştır. Buna ek olarak yapağı verimine ait bir takım ölçümler

yapılmıştır. Elde edilen bu değerler, iki yönlü tablo haline getirilerek G ve ki-kare istatistiği hesaplanmıştır[11].

### Test İstatistikleri

İki yönlü tablolarda analize konu olan iki kategorik değişkene ait genel notasyon Tablo 1’de özetlenmiştir.

Tablo 1. İki yönlü tablolara ilişkin notasyon

i/j	1	2	...	c	Toplam
1	$f_{11}$	$f_{12}$	.....	$f_{1c}$	$f_{1.}$
2	$f_{21}$	$f_{22}$	.....	$f_{2c}$	$f_{2.}$
3	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
r	$f_{r1}$	$f_{r2}$	....	$f_{rc}$	$f_{r.}$
Toplam	$f_{.1}$	$f_{.2}$	.....	$f_{.c}$	$f_{..}$

Tablo 1’de 1. kategorik değişkenin yani satır değişkeninin i. seviyesi ile 2. kategorik değişkenin yani sütun değişkeninin j. seviyesinde bulunan bireylerin yada nesnelere frekansı  $f_{ij}$  ile i.satırın toplam frekansı  $f_{i.}$  ile, 1.sütunun toplam frekansı  $f_{.j}$  ile ve toplam frekans  $f_{..}$  ile gösterilmiştir.

Tablo 1’de i.satır j.sütuna ait beklenen frekans ( $f_B$ ) ise  $f_B = (f_{i.})(f_{.j}) / (f_{..})$  eşitliği ile bulunur. Gözlenen frekans değeri ise ( $f_G$ ) = i. satır ve j. sütunda bulunan nesne yada gözlemlerin sayısı kadardır[4]. Her hücre için beklenen frekanslar ile gözlenen frekanslar arasındaki farklar tesadüfen ileri gelmekte ise bu durumda satırda yer alan kategorik değişken ile sütunda yer alan kategorik değişkenin birbirinden bağımsız olduğuna karar verilir. Diğer bir ifade ile “iki kategorik değişken arasında ilişki olmadığı” kabul edilir. Bunu test etmek için Ki-Kare istatistiği ise

$$\chi^2 = \sum_{i=1}^k \frac{(f_G - f_B)^2}{f_B} \quad (1)$$

eşitliği ile hesaplanmıştır[2,3,4,5,6,7,10,12,13]. (1) nolu eşitlikte  $f_G$  : gözlenen frekansı,  $f_B$  : beklenen frekansı ve k ise iki yönlü tablolarda toplam hücre sayısını (rxc kadar) göstermektedir.

### G İstatistiği : rxc tablolarında G istatistiği

$(\sum f_G \ln f_G)_{\text{hücre}} - (\sum f_G \ln f_G)_{\text{sıra}} - (\sum f_G \ln f_G)_{\text{sütun}} + (M \ln N)$  (2) nolu eşitliğine göre hesaplanmıştır [4,10,12,13].

4 kısımdan oluşan (2) no’lu eşitlikteki hücre indisinin olduğu 1. kısımda  $f_G$  : i. satır ve j. sütundaki gözlem sayısını, satır indisinin olduğu 2. kısımda  $f_G$  : satır toplamlarındaki frekansları, sütun indisinin olduğu 3. kısımda  $f_G$  : sütun toplamlarındaki frekansları ifade etmektedir.

G istatistiğinin Ki-kare dağılımına yaklaşımını artırmak amacıyla William Düzeltmesi yapılmıştır. Bunun için q değeri,

$$q = 1 + \frac{\left( n \sum_{i=1}^r \frac{1}{\sum_{c=1}^c f} - 1 \right) \left( n \sum_{i=1}^c \frac{1}{\sum_{r=1}^r f} - 1 \right)}{6 n (r - 1)(c - 1)}$$

olarak hesaplanmış ve buradan düzeltilmiş G istatistiği  $G_{\text{Düzeltmiş}} = G/q$  olarak elde edilmiştir [4,12,13]. Hesaplamalarda SAS istatistik paket programı kullanılmıştır[11].

### BULGULAR

İki yönlü tablolarda cinsiyet ve hemoglobin, kirli yapağı verimi ve hemoglobin tipi, kirli yapağı verimi ve transferin tipi, cinsiyet ve transferin özelliklerine ilişkin iki yönlü tablolar ve analiz sonuçları sırasıyla Tablo 2’de verilmiştir. Tablo 2, genel olarak incelendiğinde, hücrelerdeki frekansların 5’ten az olduğu ve bazı hücrelerde sıfır frekans olduğu dikkati çekmektedir. Elde edilen sonuçlar

Tablo 2. Cinsiyet, Hemoglobin Tipi, Transferin Tipi ve Kirli Yapağı Verimine İlişkin Sonuçlar

			Hemoglobin Tipi			Toplam
			1,00	2,00	3,00	
Cinsiyet	1,00	Gözlenen	2	10	1	13
		Beklenen	2,6	9,1	1,3	13,0
	2,00	Gözlenen	2	4	1	7
		Beklenen	1,4	4,9	,7	7,0
G- testi 0,829 P=0,661      X <sup>2</sup> Testi 0,848 P=0,655						
			Hemoglobin Tipi			Toplam
			1,00	2,00	3,00	
Kirli Yapağı	1,00	Gözlenen	2	8	0	10
		Beklenen	2,0	7,0	1,0	10,0
	2,00	Gözlenen	2	6	2	10
		Beklenen	2,0	7,0	1,0	10,0
G- testi 3,059 P=0,217      X <sup>2</sup> Testi 2,286 P= 0,212						

Tablo 2 : (Devam)

			Transferin		Toplam
			1	2	
Kirli Yapağı	1	Gözlenen	7	3	10
		Beklenen	8,5	1,5	10,0
	2	Gözlenen	10	0	10
		Beklenen	8,5	1,5	10,0
G- testi 4,691 P= 0,030      X <sup>2</sup> Testi 3,529 P= 0,060					
			Transferin		Toplam
			1	2	
Cinsiyet	1	Gözlenen	12	1	13
		Beklenen	11,1	2,0	13,0
	2	Gözlenen	5	2	7
		Beklenen	5,9	1,0	7,0
G- testi 1,482 P= 0,224      X <sup>2</sup> Testi 1,556 P= 0,212					

ki-kare istatistiği bakımından incelendiğinde; en yüksek değerin 3.529 ile kirli yapağı ile transferin iki yönlü tablosuna ait olduğu ancak hesaplanan bu değerin istatistik olarak önemli olmadığı görülür. Benzer şekilde diğer tablolara ait Ki-kare istatistik değerlerinin de önemli olmadığı görülmektedir. Tablo 2, G istatistiği sonuçları bakımından incelendiğinde; en yüksek değerin 4.691 ile, kirli yapağı ve transferin tablosuna ait olduğu ve bu değerin istatistik olarak önemli olduğu görülür ( $P < 0.05$ ). Diğer tablolara ait G istatistiği değerleri ise istatistik olarak önemli değildir.

### TARTIŞMA ve SONUÇ

İki yönlü tablolarda ( $r \times c$ ) yaygın olarak kullanılan Ki-kare ve G istatistikleri genel olarak benzer sonuçlar vermekle birlikte [12,13], G istatistiğinin Ki-kare dağılımına uygunluğu daha iyidir [2]. Bununla birlikte G istatistiğinin dağılımının teorik olarak daha sağlam temellere dayanması, çapraz tablolarda hücrelerdeki frekansların 5'ten küçük olduğu hallerde bile hesaplanan G istatistiğinin Ki-kare dağılımına uygunluğunun pek az etkilenmesi [8,9], ayrıca  $r \times c$  tablolarında hesaplanmanın kolay olması gibi nedenlerden dolayı G istatistiğinin yaygın olduğu söylenebilir [2,4,12,13]. Tablo 2'den görüldüğü gibi G testine göre Kirli yapağı ile transferin tipleri birbirinden bağımsız değildir. Yani kirli yapağı miktarının 0.76 kg'dan az yada çok oluşu transferin tipine göre değişmekte, ancak hemoglobinin tipine göre değişmemektedir.

Bununla birlikte, Tablo 2'de hemoglobinin ve transferin tiplerinin cinsiyete göre değişim göstermediği yani hemoglobinin ve transferin tiplerinin cinsiyetten bağımsız oldukları dikkati çekmektedir.

Sonuç olarak, Ki-kare ve G istatistiklerinin genel olarak benzer sonuçlar verdiği ancak, ki-kare istatistiğine göre yapılacak hesaplamalarda test istatistiği değerinin kritik değere yakın olduğu durumlarda, G istatistiğinin de hesaplanarak karar vermenin daha doğru olacağı

söylenebilir. Ancak her iki istatistik arasındaki farklılığı daha belirgin hale getirebilmek için daha kapsamlı çalışmaların yapılması önerilir.

### KAYNAKLAR

- 1-Berry, D.A., B.W.Lindgran, 1990.** Statistics. Theory and Methods. Pacific Grove, California.
- 2-Düzgüneş, O., T.Kesici, F.Gürbüz., 1983.** İstatistik Metotları I. Ankara Üniv. Ziraat Fak. Yayınları. Yay. No: 861, 229s, Ankara.
- 3-Ergün, M., 1995.** Bilimsel Araştırmalarda Bilgisayarla İstatistik Uygulamaları. Ocak yayınları. 292 s, Ankara.
- 4-Everitt, B.S., 1992:** The Analysis of Contingency Tables. Second Edition. London. Chapman & Hall  
5-<http://perso.club-internet.fr/hdelboy/Nonparam.htm>  
6-<http://www.eeescience.utoledo.edu/Faculty/Neher/biostatistics/McNemar.lec.htm>
- 7-Kesici, T., Z. Kocabaş, 1998.** Biyoistatistik. Ankara Üniv. Eczacılık Fak. Yay. No: 79, 359 s., Ankara.
- 8. Keskin, S., (2001),** İki Yönlü (Contingency) Tablolarda Kappa (K) İstatistiğinin Kullanımı, Biyoteknoloji (KÜKEM) Dergisi, 25(1); 53-57.
- 9-Mood, A.M., F.A.Graybill ve D.C.Boes, 1974:** Introduction to The Theory of Statistics. Third Edition. Tokyo.
- 10-Özdemir, T. 2001.** Van İlinde Birleşmiş Milletler Mülteciler Yüksek Komiserliğine İltica Talebinde Bulunan Mültecilerin Sorunları Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Zooteknik Anabilim Dalı Yüksek Lisans Tezi.
- 11-SAS, 1998.** SAS Institute Inc., Cary, NC, USA.
- 12-Sokal, R.R. and F. J. Rohlf, 1995.** Biometry W.H. Freeman and Company New York, 887p
- 13-Sokal, R.R., F.J.Rohlf, 1981.** Biometry. The Principles and Practice of Statistics in Biological Research. W.H. Freeman and Company New York.