

## Logistik Regresyon Analizi ve Genetik Çalışmalara Ait Bir Uygulama

Muhip ÖZKAN<sup>1</sup>

Handan ÇAMDEVİREN<sup>1</sup>

Geliş Tarihi : 21.02.2000

**Özet :** Bu çalışmada, polychotomous (ikiden çok seviye içeren bağımlı değişken) logistik regresyon analizinin teorik açıklamaları, hesaplama adımları, genetik verilere uygulanması ve elde edilen sonuçların yorumlanması yapılmıştır. Dört farklı bıldırcın hattı (S, K, B, W) ve üç farklı Pgi genotipinin ( $Pg^{AA}$ ,  $Pg^{AB}$ ,  $Pg^{BB}$ ) oluşturduğu çapraz tabloda yer alan frekanslardan yararlanarak, Pgi genotipleri bakımından bıldırcın hatlarının birbirinden olan genetik farklılıkları tespit edilmiş ve bunun sonucu olarak populasyonlar arasında olası genetik farklılıkların tespit edilmesinde logistik regresyon analizinden etkin bir şekilde yararlanılabileceği ifade edilmiştir.

**Anahtar Kelimeler :** Logistik regresyon, odds oranı, regresyon modeli, çapraz tablolar, Japon bıldırcını, fosfoglikoz izomeraz

### Logistic Regression Analysis and an Application to Genetic Studies

**Abstract :** In this study, technical expressions, calculating steps and application of polychotomous logistic regression analysis to genetic data are illustrated. Interpretation of results are further discussed. Genetic differences among different quail lines were determined using frequencies obtained from a cross table formed by four different quail lines (S, K, B, W) and three different Pgi genotypes ( $Pg^{AA}$ ,  $Pg^{AB}$ ,  $Pg^{BB}$ ). It was also confirmed that logistic regression analysis can be used in determining the possible genetic differences among populations.

**Key Words :** Logistic regression, odds ratio, regression model, contingency tables, Japanese quail, phosphoglucose isomerase

#### Giriş

Üzerinde durulan özellikler arasında var olan ilişkilerin ifade edilmesinde kullanılan matematiksel modellerde yer alan bağımlı (varyasyonu açıklanan) ve bağımsız (varyasyona etkisi olduğu düşünülen) değişkenler kesikli veya sürekli olabilmektedir. Kesikli değişkenler elde edilme şekillerine bağlı olarak; sınıflayıcı (nominal), sıralayıcı (ordinal), aralık (interval) ve oranlı (ratio) ölçek olmak üzere sınıflandırılabilir (Kleinbaum ve ark., 1982). Kesikli değişkenler, bu grupta sistemlerinden biri ile sınıflandırıldığında, söz konusu sınıflara farklı kodlar verilebilmektedir. Değişkenlerin seviyelerini tanımlamak amacıyla verilen bu keyfi kodlar, sadece tanımlayıcıdır ve her hangi bir sayısal değeri yoktur. Söz konusu veri tiplerine çoğunlukla sosyal ve davranış bilimleri ile biyolojinin çeşitli dallarında rastlanılmaktadır. Sınıflandırılmış kesikli değişkenler kategorik değişkenler olarak da tanımlanmaktadır. Modellerde, bağımlı ve bağımsız değişkenlerin kategorik olması durumunda, elde edilen veriler çapraz tablolar halinde düzenlenebilir. Bu şekilde oluşturulan tablolarda yer alan frekansların analiz edilmesinde çoğunlukla, Ki-kare, Exact ve Oran testleri kullanılmakla birlikte son

yıllarda log-doğrusal (log-linear) ve logistic regresyon modellerinden de yararlanılmaktadır (Anderson, 1994).

Ki-kare ve Fisher'in kesin (Exact) ki-kare testi çapraz sınıflandırılan iki veya daha fazla kategorik değişkenin birbirine olan bağımlılığının ölçülmesinde kullanılmaktadır. Log-doğrusal modeller, ki-kare değerini, varyans bileşenlerine ayırarak inceler. Böylece bu tip modeller yardımıyla ana etkiler ve etkileşim etkileri hakkında daha açıklayıcı bilgi elde edilebilir (Yıldız ve ark., 1999). Oranlar arası farka ait testler ise, söz konusu hücrelerdeki frekansların birbirinden bağımsız olarak ikili ikili karşılaştırılmasını yapmakta ve bu nedenle ki-kare ve benzeri testlerden daha detaylı bilgi vermektedir. Ancak oranlar arası farkların testinde, mevcut oranların karşılaştırılmaları birbirlerinden bağımsız gibi düşünülerek yapıldığı için, test sonucunda gerçekleşen I. tip hata yapma olasılığı başlangıçta belirlenen düzeyinden daha yüksek gerçekleşmektedir.

Logistik regresyon modellerinde bağımlı değişken daima kategorik yapıda olup iki seviye içerdiği zaman ikili

<sup>1</sup> Ankara Üniv. Ziraat Fak. Zootekni Bölümü. Biyometri ve Genetik A. nabilim Dalı-Ankara

(binary) veya dichotomous, 3 ve daha fazla seviye içerdiği zaman ise polychotomous değişken olarak adlandırılmaktadır. Modelde yer alan bağımsız değişken veya değişkenler ise kesikli veya sürekli tipte olabilmektedir (Hosmer ve ark., 1989; Agresti, 1990; Christensen, 1995). Logistik regresyon modeli, normal dağılım ön şartının yerine gelmediği durumlarda çapraz tabloların analizinde kullanılmayan ki-kare ve oranlar arası farka ait önem testlerine alternatif bir yaklaşım getirmektedir.

Bıldırcın populasyonları arasındaki var olan genetik farklılıkların ortaya konmasında çeşitli genetik markerlerden yararlanılmaktadır. Bu markerlerden biri olan fosfoglukoz izomeraz (*Pgi*) polimorfizminin japon bıldırcınlarında kodominant iki allel ( $Pgi^A$  ve  $Pgi^B$ ) tarafından kontrol edildiği ve enzimin dimerik yapıya sahip olduğu bildirilmektedir. Buna göre, bir populasyonda söz konusu lokus bakımından  $Pgi^{AA}$ ,  $Pgi^{BB}$  ve  $Pgi^{AB}$  genotipleri meydana gelmekte ve bıldırcın hatlarının *Pgi* lokusu bakımından karşılaştırılması mümkün olabilmektedir (Juneja ve Wilhelmsson 1975; Kimura ve ark., 1978b).

Üzerinde durulan biyokimyasal fonksiyonu olan lokuslar bakımından elde edilen genotip frekansları ile bıldırcın hatları arasındaki olası ilişkilerin tespit edilmesi amacıyla ki-kare, exact ve oranlar arası farka ait önem testleri ile genetik uzaklık değerlerinden yararlanılmaktadır.

Bu çalışmada, kesikli ve sıralayıcı tipte bağımlı (hatlar) ve bağımsız (*Pgi* genotipleri) değişkenler arasında olası ilişkilerin belirlenmesi amacıyla yukarıda sayılan istatistiksel testlere alternatif olarak polychotomous logistic regresyon analizinden yararlanılmıştır. Bu tekniğin tercih edilmesindeki temel noktalar ise aşağıdaki gibi özetlenebilir.

1. Logistik regresyon analizinin uygulanabilmesi için herhangi bir ön şartın yerine gelmesi gerekmemektedir.
2. İlişki veya bağımlılık ölçümünde log-olabilirlik oran istatistiği (G) kullandığı için, beklenen frekansın 1'in altına düşmesi durumunda bile başlangıçta belirlenen hata sınırını koruyucu sonuçlar vermektedir.
3. Hesaplama adımları ve elde edilen sonuçların yorumlanmaları bakımından oldukça esnek ve kolay olması testin uygulanabilirliğini artırmaktadır.
4. Elde edilen sonuçların daha ayrıntılı bilgi vermesidir.

Bu çalışmada da logistik regresyon analiziyle, *Pgi* genotipleri bakımından bıldırcın hatları arasındaki olası farklılıkların ortaya çıkarılması planlanmıştır.

### Materyal ve Yöntem

Araştırma, Ankara Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Bıldırcın Yetiştiriciliği Ünitesi'ndeki Avrupa orijinli siyah (S), kahverengi (K) ve kırmızı gözlü beyaz (B) bıldırcın hatları ile Japon orijinli beyaz (W) tüy rengine sahip bıldırcın hatları ile yürütülmüştür. Avrupa

orijinli hatların (S, K, B) tüy rengini belirleyen genler bakımından genetik analizleri yapılmamakla birlikte, bu hatlar 8 generasyon kendi içlerinde akrabalı yetiştirilmiş ve tüy rengi bakımından her hangi bir fenotipik açılma meydana gelmemiştir. Japon orijinli beyaz (W) bıldırcın hattı, 1994 yılında Gifu Üniversitesi'nden getirilmiştir. Bu genotip grubunda tüy renginin beyaz oluşunu sağlayan otozomal gen (W), yabancı alleleline ( $w^+$ ) eksik dominant etkilidir (Kimura et al. 1978b).

Kan örneklerinin alınması, nişasta jel elektroforez işlemi sırasında kullanılan tampon sistemlerinin içeriği, elektroforez koşulları, eritrosit fosfoglukoz izomeraz (E.C. 5.3.1.9) tiplerinin belirlenmesinde kullanılan aktivite boyasının içeriği ve uygulanması Yıldız (1998)' a göre yapılmıştır.

S, K, W ve B bıldırcın hatlarına sırasıyla 0, 1, 2 ve 3 kodları verilerek sıralanmış ölçekli bağımlı değişken (Y) oluşturulmuştur. *Pgi* genotipleri ise sınıflandırılmış (nominal) ölçekli bağımsız bir değişken (X) olup, 3 seviyesi vardır. Bu seviyelerden,  $Pgi^{AB}$ ,  $Pgi^{BB}$  ve  $Pgi^{AA}$  genotipleri sırasıyla 1, 2 ve 3 yapay değişkenleri ile kodlanmıştır. Sıralanmış ölçekli kesikli değişkenleri sürekli değişkenler gibi modele dahil etmek doğru olmayacağından, kategorik değişkenleri temsil eden yapay (dummy veya dizayn) değişkenlerin kullanılması gerekir (Hosmer ve Lemeshow 1989). Böylece, genotiplerden  $Pgi^{AB}$ , 1, referans grup olarak düşünülmüş ve Çizelge 1' de verilen yapay değişkenler oluşturulmuştur.

Benzer şekilde bağımlı değişken olarak bıldırcın hatlarından S hattı yani "0" seviyesi referans kategori olarak seçilmiş ve hatlara ilişkin yapay değişkenler Çizelge 2' de verilmiştir.

Yukarıda tanımlanan sıralanmış ölçekli bağımlı ve bağımsız değişkenler için oluşturulan yapay değişkenler, karşılaştırılacak hat ve genotipleri belirlemektedir. Yani, karşılaştırmalar, kontrol grubu gibi düşünülerek, seçilen referans kategorilere göre yapılmıştır. Bu karşılaştırmaların dışında diğer mümkün olabilecek bütün karşılaştırmalar da farklı grup ve hatlar referans alınarak yapılabilir.

Çizelge 1. *Pgi* genotipleri için yapay değişkenler

<i>Pgi</i> genotipleri	Yapay değişkenler	
	$D_1$	$D_2$
$Pgi^{AB}$	0	0
$Pgi^{BB}$	1	0
$Pgi^{AA}$	0	1

Çizelge 2. Bıldırcın hatları için yapay değişkenler

Bıldırcın Hatları	Yapay değişkenler		
	$E_1$	$E_2$	$E_3$
S	0	0	0
K	1	0	0
W	0	1	0
B	0	0	1

Bağımlı değişkenin her hangi bir seviyesinin gerçekleşme olasılığının, bir diğer seviyesinin gerçekleşme olasılığına oranı teklik (odds) oranı ( $\varphi$ ) olarak adlandırılır. Sadece iki seviye içeren bir bağımlı değişken için teklik oranı;

$\varphi = p / (1-p)$  eşitliğinden hesaplanır. Buradaki,

$p$  : Bağımlı değişkenin 1. seviyesinin gerçekleşme olasılığı

$1-p$  : Bağımlı değişkenin 2. seviyesinin gerçekleşme olasılığını göstermektedir.

Bağımlı değişkenin kategorik yapıda olması durumunda hataların dağılımı binomiyal veya multinomiyal olur. Logistik regresyon analizinde, doğrusal olasılık modelinin uygulanabilmesi için hataların dağılımının şeklinin normal dağılım olması istenir. Hataların dağılımının şeklini normale dönüştürmek ve parametreleriyle doğrusal olan bir model kurabilmek için, teklik oranının doğal logaritması ( $\ln$ ) alınır. Logaritması alınmış teklik oranına logit (log-odds) fonksiyonu adı verilir ve  $g(x)$  ile gösterilir.

Bu çalışmada seçilen referans kategoriler dikkate alındığında logit fonksiyonlar aşağıdaki gibi tanımlanabilir.

1. S hattı ile K hattının karşılaştırılmasında kullanılabilir logit fonksiyon ( $Y=0$ 'a karşı  $Y=1$ );

$$g_1(x) = \ln \left[ \frac{P(Y=1|D)}{P(Y=0|D)} \right] = \hat{\beta}_0 + \hat{\beta}_{1(1,0)}(2,1)D_1 + \hat{\beta}_{2(1,0)}(3,1)D_2$$

Buradaki,

$\hat{\beta}_0$  : Kurulan logistik fonksiyondaki sabit terim,

$\hat{\beta}_{1(1,0)}(2,1)$  : K (1) hattı ile S (0) hattının  $P_{gi}^{BB}$  (2) genotipinin  $P_{gi}^{AB}$  (1) genotipine nazaran bulunma sıklığı bakımından karşılaştırılmasında kullanılacak regresyon katsayısını,

$\hat{\beta}_{2(1,0)}(3,1)$  : K (1) hattı ile S (0) hattının  $P_{gi}^{AA}$  (3) genotipinin  $P_{gi}^{AB}$  (1) genotipine nazaran bulunma sıklığı bakımından karşılaştırılmasında kullanılacak regresyon katsayısını göstermektedir.

$D_1$  ve  $D_2$  ise  $P_{gi}$  genotipleri için oluşturulan dizayn değişkenleri olup referans genotip olan  $P_{gi}^{AB}$  (1) genotipinin sırasıyla  $P_{gi}^{BB}$  (2) ve  $P_{gi}^{AA}$  (3) genotipleri ile karşılaştırılmasında kullanılan yapay (dummy) değişkenlerdir.

2. S hattı ile W hattının karşılaştırılmasında kullanılabilir logit fonksiyon ( $Y=0$ 'a karşı  $Y=2$ );

$$g_2(x) = \ln \left[ \frac{P(Y=2|D)}{P(Y=0|D)} \right] = \hat{\beta}_0 + \hat{\beta}_{1(2,0)}(2,1)D_1 + \hat{\beta}_{2(2,0)}(3,1)D_2$$

ve

3. S hattı ile B hattının karşılaştırılmasında kullanılabilir logit fonksiyon ( $Y=0$ 'a karşı  $Y=3$ );

$$g_3(x) = \ln \left[ \frac{P(Y=3|D)}{P(Y=0|D)} \right] = \hat{\beta}_0 + \hat{\beta}_{1(3,0)}(2,1)D_1 + \hat{\beta}_{2(3,0)}(3,1)D_2$$

dir.

Referans kategorilerle olan karşılaştırmaların dışında, olası diğer karşılaştırmalar için kullanılabilir logit fonksiyonlar ise aşağıdaki gibidir.

4. K hattı ile W hattının karşılaştırılmasında kullanılacak ( $Y=1$ 'e karşı  $Y=2$ ) logit fonksiyon. Bu fonksiyon,  $Y=0$ 'a karşı  $Y=2$ 'nin logit fonksiyon değerinden,  $Y=0$ 'a karşı  $Y=1$ 'in logit fonksiyon değerinin çıkarılmasıyla elde edilir.

5. K hattı ile B hattının karşılaştırılmasında kullanılacak ( $Y=1$ 'e karşı  $Y=3$ ) logit fonksiyon ise,  $Y=0$ 'e karşı  $Y=3$ 'ün logit fonksiyonundan,  $Y=0$ 'a karşı  $Y=1$ 'in logit fonksiyonunun çıkarılmasıyla elde edilir.

6. W hattı ile B hattının karşılaştırılmasında kullanılacak ( $Y=2$ 'e karşı  $Y=3$ ) logit fonksiyon olup,  $Y=0$ 'a karşı  $Y=3$ 'ün logit fonksiyonundan,  $Y=0$ 'a karşı  $Y=2$ 'nin logit fonksiyonu çıkarılarak elde edilir.

Logit fonksiyonlarda verilen,  $P(Y=j|D)$  ( $j=0, 1, 2, 3$ ) olasılıklarına şartlı olasılıklar adı verilir ve aşağıdaki eşitlikler yardımıyla hesaplanmıştır.

$$P(Y=0|D) = \frac{1}{1 + \exp(g_1(x)) + \exp(g_2(x)) + \exp(g_3(x))}$$

$$P(Y=1|D) = \frac{\exp(g_1(x))}{1 + \exp(g_1(x)) + \exp(g_2(x)) + \exp(g_3(x))}$$

$$P(Y=2|D) = \frac{\exp(g_2(x))}{1 + \exp(g_1(x)) + \exp(g_2(x)) + \exp(g_3(x))}$$

$$P(Y=3|D) = \frac{\exp(g_3(x))}{1 + \exp(g_1(x)) + \exp(g_2(x)) + \exp(g_3(x))}$$

Bu olasılıklar için yazılacak genel formül ise;

$$P(Y=j|D) = \frac{\exp(g_j(x))}{\sum_{k=0}^3 \exp(g_k(x))} \quad (\text{Hosmer ve Lemeshow 1989})$$

olarak verilebilir.

Çalışmada kullanılan materyale ilişkin şartlı olasılıklar Çizelge 3' de verilmiştir.

Yukarıda görüldüğü üzere modelde bağımlı değişken konumunda bulunan olasılıklar, logit fonksiyonunda yer alan regresyon veya eğim katsayılarıyla doğrusaldır.

Çizelge 3. Y' nin 4, X' in 3 seviyesi söz konusu olduğunda şartlı olasılıklar

Seviyeler	Bağımsız değişken (X)		
	1	2	3
0	$P(Y=0   X=1)$	$P(Y=0   X=2)$	$P(Y=0   X=3)$
1	$P(Y=1   X=1)$	$P(Y=1   X=2)$	$P(Y=1   X=3)$
2	$P(Y=2   X=1)$	$P(Y=2   X=2)$	$P(Y=2   X=3)$
3	$P(Y=3   X=1)$	$P(Y=3   X=2)$	$P(Y=3   X=3)$

Logit modelde yer alan regresyon katsayılarının standart hataları ise,

$$S(\hat{\beta}) = \sqrt{\sum_i \sum_j \frac{1}{f_{ij}}}$$

eşitliği yardımıyla hesaplanmıştır. Bu eşitlikte yer alan  $f_{ij}$ ,  $i$ . satır  $j$ . sütundaki frekansı göstermektedir.

Regresyon katsayılarının önem kontrolünde Wald istatistiği (WS) kullanılmış ve  $WS = \frac{\hat{\beta}}{S(\hat{\beta})}$  eşitliği

yardımıyla hesaplanmıştır. Bu istatistiğin karesi, 1 serbestlik dereceli ki-kare dağılımı gösterir ve dolayısıyla Wald istatistiği aslında standart normal dağılımı gösteren bir rastgele değişken olur. Dolayısıyla, WS istatistiğinin hesaplanan değeri standardize edilmiş normal dağılımdan elde edilen değere eşittir. Bu durumda standart normal dağılımı gösteren değişkenin karesi,  $\alpha$  yanılma olasılığında 1 serbestlik dereceli ki-kare dağılımı gösterir ve  $Z^2 = (WS)^2 \sim \chi^2_{\alpha, 1}$  olarak yazılabilir.

Regresyon katsayılarına ilişkin  $1 - \alpha / 2$  ihtimalli güven aralığı;  $P(\hat{\beta} - Z_{\alpha/2} S(\hat{\beta}) \leq \beta \leq \hat{\beta} + Z_{\alpha/2} S(\hat{\beta})) = 1 - \alpha / 2$  dir.

Logistik modeldeki regresyon katsayılarının  $e$  üssü değeri ise teklik oranını verir ki bu oran sonuçların yorumlanmasında çok etkili bir istatistiktir. X değişkeninin  $a$  ve  $b$  gibi iki değeri dikkate alındığında, bağımlı değişkenin (Y' nin)  $j$ . seviyesi için hesaplanacak teklik oranı aşağıdaki eşitlikte tanımlanmıştır.

$$\varphi_{(i,l)}(a,b) = \frac{P(Y = i | x = a) / P(Y = 0 | x = a)}{P(Y = i | x = b) / P(Y = 0 | x = b)} \quad (i=1,2,3; i'=0)$$

Teklik oranına ait  $1 - \alpha / 2$  ihtimalli güven aralığı ise,  $P[\exp(\hat{\beta} - Z_{\alpha/2} S(\hat{\beta})) \leq \varphi \leq \exp(\hat{\beta} + Z_{\alpha/2} S(\hat{\beta}))] = 1 - \alpha / 2$  dir.

Teklik oranına ait güven aralığında 1 değerinin bulunması durumunda tahmin edilen teklik oranının istatistik olarak önemli olmadığına karar verilir. Çünkü teklik oranının 1'e eşit olması durumunda bu oranın pay ve paydasında yer alan olasılıkların aynı olduğu sonucu ortaya çıkar.

Modelde yer alan bağımsız değişkenin modele katkısının önemli olup olmadığını belirlemede kullanılan bir diğer istatistik ise olabilirlik oran istatistiğidir (G). Bu istatistik aşağıdaki eşitlik yardımıyla hesaplanır.

$$G = -2 [LL_{\beta_0, \beta_1} - LL_{\beta_0}] \quad \text{Bu eşitlikte,}$$

$LL_{\beta_0, \beta_1}$ : Bağımsız değişken içeren modelin log-olabilirliği  $LL_{\beta_0}$ : Sadece sabit terim içeren modelin log-olabilirliğidir.

Bu istatistik (G),  $p$  ( $p$ : modelde tahmin edilen parametre sayısı) serbestlik dereceli ki-kare dağılımı gösterir.

En küçük kareler regresyon modellerinde, ele alınan bağımsız değişken veya değişkenler tarafından bağımlı değişkendir varyansın açıklanan oranı veya kısaca belirleme katsayısı olan  $R^2$  ye benzer ölçüler logistik regresyon için de geliştirilmiştir. Bunlardan birisi McFadden'in  $R^2$  istatistiğidir ve aşağıdaki eşitlik yardımıyla hesaplanır. Bu istatistik olabilirlik oran indeksi (LRI) olarak ta bilinmektedir.

$$MF - R^2 = 1 - \left[ \frac{L(\hat{\beta}_0, \hat{\beta}_1)}{L(\hat{\beta}_0)} \right] = 1 - \left[ \frac{-2L(\hat{\beta}_0, \hat{\beta}_1)}{-2L(\hat{\beta}_0)} \right]$$

Ancak bu değer, doğrusal regresyon modelindeki  $R^2$  den oldukça düşük sonuçlar vermektedir. Bir başka deyişle, logistik regresyonda maksimum  $R^2$  yi elde etmek oldukça zordur. Bu istatistiğin logistik regresyonda en uygun kullanımı uygun modelin seçimindedir (Whitehead 1998).

Sonuçlara ilişkin hesaplamalar için, Minitab istatistik paket programı kullanılmıştır. Regresyon katsayıları, regresyon katsayılarının standart hataları ve teklik oranlarının, en çok olabilirlik tahminleri, iteratif yeniden ağırlıklandırılmış en küçük kareler (iterative reweighted least square) algoritması kullanılarak elde edilmiştir.

## Bulgular ve Tartışma

Çizelge 4' de verilen frekansların bağımlılık durumunu araştırmak için oluşturulan 4x3 lük çapraz tablo için hesaplanan ki-kare ( $\chi^2$ ) istatistiği 53,974' tür. Bu değer 6 serbestlik dereceli ki-kare tablo değeri ile karşılaştırıldığında istatistik olarak önemlidir ( $P < 0,01$ ).

Buna göre bildircin hatlarının  $Pgi$  genotiplerine bağımlı olduğu söylenebilir. Ancak Çizelge 4' de görülebileceği gibi tablonun bazı gözlerinde beklenen frekansların 5' den küçük olması nedeniyle bağımlılık durumunun araştırılmasında G-testi veya Fisher exact testlerinin uygulanması, sonuçlarda hata yapma olasılığının, başlangıçta belirlenen düzeyde kalmasını sağlayacaktır. Logistik regresyon analizinde bağımlı ve bağımsız değişkenler arasındaki ilişkinin ölçülmesinde log-olabilirlik oran istatistiği kullanılır ki bu istatistik genel olarak G-testi olarak bilinmektedir. Bu test çapraz tablolarda gözlerdeki beklenen frekansların 1' in altında olması durumunda bile hata yapma olasılığını sınır değerinde koruyan sonuçlar vermektedir. Çizelge 5' de görüleceği üzere G değeri (61,225; SD=6) istatistik olarak önemli bulunmuştur. Bu sonuç, kurulan logistik regresyon modeline bağımsız değişken olarak dahil edilen  $Pgi$  genotiplerinin, modele katkısının önemli olduğunu, diğer bir ifade ile, bildircin hatları ile  $Pgi$  genotipleri arasındaki ilişkinin istatistik olarak önemli olduğunu göstermektedir ( $P < 0,01$ ).

Bıldircin hatları ile  $Pgi$  genotipleri arasında tespit edilen ilişkinin daha ayrıntılı bir şekilde incelenmesi amacıyla, farklı koşullar için teklik oranları, regresyon katsayıları ve güven aralıkları tahmin edilmiştir. Bu değerlere ilişkin bulunan sonuçlar ise aşağıdaki gibi özetlenebilir.

B bıldircin hattı S bıldircin hattı ile,  $Pgi^{BB}$  homozigot genotipin,  $Pgi^{AB}$  heterozigot genotipe nazaran bulunma sıklığı bakımından karşılaştırıldığında, teklik oranı,  $\hat{\varphi}_{(3,0)}(2,1) = \frac{11 \times 5}{8 \times 42} = 0,16$  olarak tahmin edilmiştir (Çizelge 4). Bulunan bu değere ait % 95 ihtimalli güven

aralığı incelendiği zaman, aralıkta 1 değeri bulunmamaktadır. Bu durum, tahmin edilen teklik oranının istatistik olarak önemli olduğunu göstermektedir ( $P < 0,05$ ). Yani, homozigot resesif  $Pgi^{BB}$  genotipi, heterozigot  $Pgi^{AB}$  genotipine nazaran, B hattında S hattından önemli düzeyde düşük bulunmuştur. Bir başka ifadeyle,  $Pgi^{BB}$  genotipinin  $Pgi^{AB}$  genotipine nazaran, S ve B hattı arasında bulunma sıklığı oranlandığı zaman S / B=100 / 16 olduğu söylenebilir.

Benzer şekilde  $Pgi^{BB}$  genotipi  $Pgi^{AB}$  genotipine nazaran K hattında S hattından önemli düzeyde düşük ( $P < 0,05$ ) bulunmuştur (Çizelge 4). Bu koşula ilişkin tahmin edilen teklik oranı  $\hat{\varphi}_{(1,0)}(2,1) = 0,03$  tür.  $Pgi^{BB}$  genotipi  $Pgi^{AB}$  genotipine nazaran bulunma sıklığı bakımından, K hattı ile S hattı arasında oranlandığı zaman K / S=3 / 100 olduğu görülür.

$Pgi^{BB}$  genotipinin  $Pgi^{AB}$  genotipine nazaran bulunma sıklığı bakımından W hattı ile S hattı ile karşılaştırıldığında, tahmin edilen teklik oranına göre istatistik olarak önemli bir fark olmadığı gözlenmiştir (Çizelge 4).

Çizelge 4. Bıldircin hatlarının  $Pgi$  genotiplerine göre dağılımı

	Pgi genotipleri			Toplam
	$Pgi^{AB}$ (1)	$Pgi^{BB}$ (2)	$Pgi^{AA}$ (3)	
S (0)	5 (=f <sub>11</sub> )	42 (=f <sub>12</sub> )	0 (=f <sub>13</sub> )	47 (=f <sub>1</sub> )
K (1)	31 (=f <sub>21</sub> )	8 (=f <sub>22</sub> )	8 (=f <sub>23</sub> )	47 (=f <sub>2</sub> )
W (2)	5 (=f <sub>31</sub> )	12 (=f <sub>32</sub> )	0 (=f <sub>33</sub> )	17 (=f <sub>3</sub> )
B (3)	8 (=f <sub>41</sub> )	11 (=f <sub>42</sub> )	1 (=f <sub>43</sub> )	20 (=f <sub>4</sub> )
Toplam	49 (=f <sub>1</sub> )	73 (=f <sub>2</sub> )	9 (=f <sub>3</sub> )	131 (=f <sub>..</sub> )

Çizelge 5. Polytohomous logistik regresyon analizi sonuçları

Logit 1: (B / S)	Regresyon katsayısı ( $\hat{\beta}$ )	S( $\hat{\beta}$ )	$\chi^2$	p	$\hat{\varphi}$	%95 ihtimalli güven aralığı	
						Alt sınır	Üst sınır
Sabit ( $\hat{\beta}_0$ )	0,4700	0,5701	0,672	0,410			
$Pgi^{BB} / Pgi^{AB}$	-1,8098	0,6631	7,453	0,006**	0,16	0,04	0,60
$Pgi^{AA} / Pgi^{AB}$	---	---	---	---	---	---	---
Logit 2: (W / S)							
Sabit ( $\hat{\beta}_0$ )	0,0000	0,5325	0,000	1,000			
$Pgi^{BB} / Pgi^{AB}$	-1,2528	0,7121	3,098	0,079	0,29	0,07	1,15
$Pgi^{AA} / Pgi^{AB}$	---	---	---	---	---	---	---
Logit 3: (K / S)							
Sabit ( $\hat{\beta}_0$ )	1,8245	0,4819	14,364	0,000**			
$Pgi^{BB} / Pgi^{AB}$	-3,4828	0,6173	31,810	0,000**	0,03	0,01	0,10
$Pgi^{AA} / Pgi^{AB}$	---	---	---	---	---	---	---
$MF - R^2 = \% 18,15$							
$L(\hat{\beta}_0) = -168,658 ; L(\hat{\beta}_0, \hat{\beta}_1) = -138,045$							
Log-olabilirlik oran istatistiği (G) = -2 [-138,045 - (-168,658)] = 61,225							
$G \approx \chi^2_{0,05, 6}$ ve $P < 0,01$							

Bu koşul için hesaplanan teklik oranı,  $\hat{\varphi}_{(2,0)}(2,1) = 0,29$  olup bu değere ilişkin oluşturulan %95 ihtimalli güven aralığında 1 değeri yer almaktadır. Bu sonuç, W ve S hatlarında  $P_{gi}^{BB}$  ve  $P_{gi}^{AB}$  genotiplerinin benzer oranlarda bulunduğunu göstermektedir.

$P_{gi}^{AA}$  genotipinin  $P_{gi}^{AB}$  genotipine nazaran görülme sıklığı bakımından K, W ve B hatları S hattı ile karşılaştırılmamıştır. Çünkü S hattında  $P_{gi}^{AA}$  genotipine hiç rastlanılmamıştır.

S hattı referans hat olarak seçilerek, yapılan hesaplamalar sonucunda elde edilen bu bulgulara ilaveten, eğer istenirse söz konusu genotipler, görülme sıklıkları bakımından da diğer hatlarla karşılaştırılabilir. Diğer karşılaştırmalar için kullanılacak teklik oranları ise sırasıyla şöyledir.

W hattı ile K hattı,  $P_{gi}^{BB}$  genotipinin  $P_{gi}^{AB}$  genotipine nazaran bulunma sıklığı bakımından karşılaştırılmasında kullanılacak teklik oranı,

$$\hat{\varphi}_{(2,1)}(2,1) = \frac{31 \times 12}{8 \times 5} = 9,3 \text{ olarak tahmin edilir.}$$

Bulunan bu değere göre  $P_{gi}^{BB}$  genotipi  $P_{gi}^{AB}$  genotipine nazaran W hattında K hattından 9.3 kat daha fazla bulunmaktadır ve istatistik olarak önemli bulunmuştur ( $p < 0,01$ ). Bu teklik oranı referans hattın S olması durumunda hesaplanan regresyon katsayılarının farklarından da hesaplanabilir. Şöyle ki;  $P_{gi}^{BB}$  genotipinin  $P_{gi}^{AB}$  genotipine nazaran bulunma sıklığı bakımından K hattı ile S hattının karşılaştırılmasında elde edilen regresyon katsayısı; -3.4828 ve W hattı ile S hattının karşılaştırılmasında elde edilen regresyon katsayısı; -1.2528 dir. Bu iki regresyon katsayısı arasındaki fark,

$\hat{\beta}_{(2,1)}(2,1) = -1,2528 - (-3,428) = 2,23$  dür. Bu değer e üssü yani,  $\exp(2,23)=9,3$  olup yukarıda hesaplanan teklik oranını verir. Bu koşul için bulunan regresyon katsayısının

standart hatası,  $S(\hat{\beta}) = \sqrt{\frac{1}{31} + \frac{1}{12} + \frac{1}{8} + \frac{1}{5}} = 0,664$  tür.

Teklik oranı için % 95 ihtimalli güven aralığı ise;

$$\exp(2,23-1,96 \times 0,664) < \varphi_{(2,1)}(2,1) < \exp(2,23+1,96 \times 0,664)$$

$= 2,531 < \varphi_{(2,1)}(2,1) < 34,158$  olup aralıkta 1 değerini içermediği için tahmin edilen teklik oranı istatistik olarak önemli bulunmuştur ( $P < 0,05$ ).

Aynı  $P_{gi}$  genotipleri bakımından B hattı ile K hattının karşılaştırılması için kullanılacak teklik oranı;

$$\hat{\varphi}_{(3,1)}(2,1) = \frac{11 \times 31}{8 \times 8} = 5,328 \text{ olup istatistik olarak önemli}$$

bulunmuştur ( $p < 0,05$ ). Bu istatistiğe ait güven aralığı ise;  $\exp(1,673-$

$$1,96 \times 0,611) < \varphi_{(3,1)}(2,1) < \exp(1,673+1,96 \times 0,611)=$$

$$1,6087 < \varphi_{(3,1)}(2,1) < 17,647$$

olup aralıkta 1 değerini içermediği için B hattında K hattından  $P_{gi}^{BB}$  genotipinin  $P_{gi}^{AB}$  genotipine nazaran önemli derecede yüksek bulunduğu sonucu ortaya çıkar.

$P_{gi}^{BB}$  genotipinin  $P_{gi}^{AB}$  genotipine nazaran görülme sıklığı bakımından, B hattı ile W hattının karşılaştırılmasında kullanılacak teklik oranı;

$$\hat{\varphi}_{(3,2)}(2,1) = \frac{11 \times 5}{8 \times 12} = 0,573 \text{ tür. Bu orana ilişkin \%95}$$

ihтимalli güven aralığı,

$$\exp(-0,557-1,96 \times 0,706) < \varphi_{(3,2)}(2,1) < \exp(-$$

$0,557+1,96 \times 0,706) = 0,143 < \varphi_{(3,2)}(2,1) < 2,286$  olup güven aralığında 1 değerinin bulunmasından dolayı  $P_{gi}^{BB}$  genotipinin  $P_{gi}^{AB}$  genotipine nazaran B hattı ile W hattı arasında istatistik olarak önemli bir farkın olmadığını gösterir.

$P_{gi}^{AA}$  genotipi  $P_{gi}^{AB}$  genotipine nazaran görülme sıklığı bakımından B hattı K hattı ile karşılaştırıldığında,

$$\hat{\varphi}_{(3,1)}(3,1) = \frac{31 \times 1}{8 \times 8} = 0,484 \text{ olarak hesaplanmıştır. Bu}$$

koşula ilişkin regresyon katsayısı

$\hat{\beta}_{(3,1)}(3,1) = \ln(\hat{\varphi}) = -0,726$  ve regresyon katsayısının standart hatası ise,

$$S(\hat{\beta}) = \sqrt{\frac{1}{31} + \frac{1}{1} + \frac{1}{8} + \frac{1}{8}} = 1,132 \text{ dir. Bu durumda tahmin}$$

edilen teklik oranı için % 95 ihtimalli güven aralığı;

$$\exp(-0,726-1,96 \times 1,132) < \varphi_{(3,1)}(3,1) < \exp(-0,726+1,96 \times 1,132)=$$

$0,053 < \varphi_{(3,1)}(3,1) < 4,449$  olup aralıkta 1 değerini içerdiği için B hattında K hattından  $P_{gi}^{AA}$  genotipinin  $P_{gi}^{AB}$  genotipine göre görülme sıklığı bakımından önemli bir fark olmadığı sonucuna varılır.

$P_{gi}^{AA}$  genotipinin  $P_{gi}^{BB}$  genotipine nazaran görülme sıklığı bakımından B hattı ile K hattı arasında önemli bir farklılığın olup olmadığını araştırmak amacıyla

$$\text{hesaplanan teklik oranı; } \hat{\varphi}_{(3,1)}(3,2) = \frac{8 \times 1}{11 \times 8} = 0,091 \text{ olup}$$

regresyon katsayısı ve regresyon katsayısının standart hatası ile sırasıyla,

$\hat{\beta}_{(3,1)}(3,2) = \ln(0,091) = -2,397$  ve  $S(\hat{\beta}) = 1,158$  dir. Bu sonuçlara göre teklik oranı için % 95 ihtimalli güven aralığı;

$0,009 < \varphi_{(3,1)}(3,2) < 0,88$  olarak bulunur. Bulunan aralıkta 1 değeri yer almadığı için B hattında K hattından  $P_{gi}^{AA}$  genotipinin  $P_{gi}^{BB}$  genotipine nazaran önemli düzeyde daha az görüldüğü söylenebilir ( $P < 0,05$ ).

Genotipler bakımından hatların karşılaştırılmasında kullanılacak diğer kombinasyonlara ilişkin teklik oranlarının tahminleri ve güven aralıkları, teklik oranlarının sıfır veya tanımsız çıkmasından dolayı hesaplanmamıştır.

Bağımlı ve bağımsız değişkenlerin kategorik olduğu durumlarda uygulanan logistik regresyon analizi sonucunda hesaplanan regresyon katsayılarının yorumlanması fazla tercih edilmez. Genellikle bu katsayıların yerine teklik oranları ile sonuçlar açıklanmaktadır.

### Sonuç

Bulunan sonuçlar Çizelge 6' da özetlenmiştir.

Söz konusu *Pgi* genotipleri bakımından, Çizelge 6' da yer alan farklı harfler bildircin hatları arasındaki farklılıkları, aynı veya ortak harfler ise benzer bildircin hatlarını göstermektedir. Ayrıca harflerin alfabetik sırası ise karşılaştırılan genotiplerin görülme sıklığı bakımından yüksekten düşüğe doğru olmasına göre seçilmiştir. Mesela, S hattında K hattından, *Pgi<sup>BB</sup>* genotipi *Pgi<sup>AB</sup>* genotipine nazaran daha yüksek bulunmaktadır.

Çizelge 6' da *Pgi<sup>AA</sup>* genotipinin *Pgi<sup>AB</sup>* genotipine nazaran (*Pgi<sup>AA</sup> - Pgi<sup>AB</sup>*) ve *Pgi<sup>AA</sup>* genotipinin *Pgi<sup>BB</sup>* genotipine nazaran (*Pgi<sup>AA</sup> - Pgi<sup>BB</sup>*) görülme sıklıkları bazı hatlarda karşılaştırılmamıştır. Çünkü S ve W hatlarında *Pgi<sup>AA</sup>* genotipine hiç rastlanılmamıştır.

Özet olarak dört hat içinde S ve W hatlarında *Pgi<sup>BB</sup>* genotipi *Pgi<sup>AB</sup>* genotipinden daha yüksek bulunmuş, B hattında ise S hattına nazaran daha düşük olmasına karşın, W hattı ile benzer oranlarda rastlanmıştır.

K hattında ise *Pgi<sup>BB</sup>* genotipi *Pgi<sup>AB</sup>* genotipine nazaran görülme sıklığı bakımından diğer hatlardan önemli düzeyde düşük olduğu görülmüştür. Ayrıca *Pgi<sup>AA</sup>* genotipi *Pgi<sup>AB</sup>* genotipine nazaran görülme sıklıkları bakımından K ve B hatlarının benzer olmasına karşın, *Pgi<sup>AA</sup>* genotipinin *Pgi<sup>BB</sup>* genotipine nazaran görülme sıklığı bakımından K hattında daha yüksek olduğu tespit edilmiştir. Yani K hattında B hattından *Pgi<sup>AA</sup>* genotipi *Pgi<sup>BB</sup>* genotipine göre daha yüksektir.

Buraya kadar yapılan açıklamalar doğrultusunda, denemeler sonunda elde edilen verilerin çapraz tablolar halinde sınıflandırılması durumunda, alışılmış olarak kullanılan ki-kare, Fisher'in kesin ki-kare testi ve log-doğrusal modellerin yanında logistik regresyon analiziden de etkin bir şekilde yararlanılabileceği söylenebilir.

Çizelge 6. *Pgi* genotipleri bakımından bildircin hatlarının karşılaştırılması

Hatlar	<i>Pgi</i> genotipleri		
	<i>Pgi<sup>BB</sup> - Pgi<sup>AB</sup></i>	<i>Pgi<sup>AA</sup> - Pgi<sup>AB</sup></i>	<i>Pgi<sup>AA</sup> - Pgi<sup>BB</sup></i>
S	a	---	---
K	c	a	a
W	ab	---	---
B	b	a	b

### Kaynaklar

- Agresti, A. 1990. *Categorical Data Analysis*.; John Wiley & Sons, 558 p., New York-USA.
- Anderson, E.B. 1994. *The Statistical Analysis of Categorical Data*. Third Edition. Springer-Verlag, Germany.
- Christensen, R. 1995. *Log-Linear Models*. Springer-Verlag, 407 p., New York-USA.
- Hosmer, D.W. and S. Lemeshow, 1989. *Applied Logistic Regression*. John Wiley & Sons, 307 p., New York-USA.
- Juneja, R.K. and M. Wilhelmsson, 1975. Genetic Control of Phosphoglucose Isomerase Isozymes in the Japanese Quail Erythrocytes. *Anim. Blood Grps. Biochem. Genet.* (6), 3-7.
- Kleinbaum, D.G., L.L. Kupper and L.E. 1982. *Chambless Logistic Regression Analysis of Epidemiologic Data: Theory and Practice*. *Commun. Statist.-Theor. Meth*2. 11(5), 485-547.
- Kimura, M., Ishiguro, M. Ito, and S. Isogai, 1978. Glucosephosphate Isomerase Isozymes of the Japanese Quail. *Japan. Poulit. Sci.* 1978b:15: 316-318.
- Whitehead, J. 1998. *Logistic Regression*.; Prepared for the ECU Faculty Development and academic Computing's Statlab Summer Workshop. Erişim tarihi: 15-11-1999, Erişim adresi : <http://www1.ecu.edu/~ecwhiteh/logit/home.htm>.
- Yıldız, M.A. 1998. Çeşitli Japon Bildircini (*Coturnix japonica*) Hatlarında Protein Polimorfizmi ile Bunların Verimle Olan İlişkileri. A.Ü. Fen Bilm. Enst. Doktora Tezi.
- Yıldız, M.A., H. Çamdeviren ve T. Kesici, 1999. Bildircin Hatlarının Tanımlanmasında Arilesteraz Tiplerinin Kullanılabilirliğinin Belirlenmesi. *Tarım Bilimleri Dergisi*. 5 (2), 65-68.