**DEÜ MÜHENDİSLİK FAKÜLTESİ
MÜHENDİSLİK BİLİMLERİ DERGİSİ
Cilt: 14 Sayı: 1 sh. 39-47 Ocak 2012**

KARIŞIMLI İKİLİ LOJİSTİK REGRESYON MODELİNE İLİŞKİN BİR UYGULAMA

(AN APPLIACTION FOR MIXTURE BINARY LOGISTIC REGRESSION MODEL)

Yılmaz KAYA*, Abdullah YEŞİLOVA**

ÖZET/*ABSTRACT*

Lojistik regresyonda, gözlenen varyansın, beklenen varyanstan büyük olması aşırı yayılım olarak tanımlanmaktadır. Karışımlı modellemede, aşırı yayılıma gözlenemeyen heterojenliğin neden olduğu varsayılmaktadır. Veri seti kendi içerisinde homojen alt populasyonlara ayrılarak, aşırı yayılım giderilmektedir. Karışımlı lojistik regresyonda parametre tahminlerinin elde edilmesinde EM algoritmasını esas alan en çok olabilirlik yöntemi kullanılmaktadır. Uygun model seçiminde ise AIC ve BIC ölçütleri yaygın olarak kullanılmaktadır. Çalışmada, lojistik regresyon analizi sonucunda meydana gelen aşırı yayılım, veri seti kendi içerisinde homojen iki alt populasyona ayrılarak giderilmiştir. Modele alınan bağımsız değişkenlerin tamamı etkileri istatistiksel olarak önemli bulunmuştur ($p<0.01$).

In logistics regression, when observed variance is more than expected variance it is defined as over dispersion. In mixture modeling, it is assumed that unobserved heterogeneity causes the over dispersion. The data set is divided into homogenous sub proportions in order to overcome the over dispersion. In obtaining the parameter estimations, Maximum likelihood method which takes the EM algorithm, is used. For suitable model selection, AIC and BIC criteria are widely used. In this study, over dispersion caused by logistic regression analysis was solved by separating the data set to two homogenous sub- populations. All independent variables taken to the model were found statistically significant ($p<0.01$).

ANAHTAR KELİMELER/*KEYWORDS*

AIC, BIC, EM algoritması, Lojistik regresyon, Mixture binary logistic regression

AIC, BIC, EM algorithm, Logistic regression, Karışımlı ikili lojistik regresyon

* Yüzüncü Yıl Ün., Van Meslek Yüksekokulu, Bilgisayar Teknolojileri ve Programcılığı Bölümü, VAN
** Yüzüncü Yıl Ün., Ziraat Fak., Zooktekni Böl., Biyometri ve Genetik Anabilim Dalı, VAN

1. GİRİŞ

Lojistik regresyon (LR), bağımlı değişkenin binom dağılım gösterdiği durumlarda kullanılmaktadır. Başka bir ifadeyle, LR, ikili (binary) bağımlı değişken ile bağımsız değişkenler arasındaki neden sonuç ilişkisini belirlemede kullanılan bir yöntemdir (Bonney, 1987; Zhang, 1999; Stokes vd., 2000; SAS, 2008). LR'de, genelleştirilmiş doğrusal modeller kullanılarak bağımsız değişkenlerin doğrusal yapısını, ikili bağımlı değişkeninin beklenen değerine bağlayan bir bağlantı (link) fonksiyonunu kullanmaktadır. LR'de, kullanılan bağlantı fonksiyonu logit dönüşüm ile verilmektedir. (McCullagh ve Nelder, 1989; Dobson, 1990; Stokes vd., 2000). LR'de, gözlenen varyansın, beklenen varyanstan büyük olması aşırı yayılım (overdispersion) olarak tanımlanmaktadır (Cox, 1983; Lambert ve Roeder, 1995; Lindeys, 1998).

Aşırı yayılım, genellikle gözlenemeyen heterojenliğin neden olduğu bir durumdur (Wang vd., 1996; Wang vd., 1998; Jones vd., 2001; Yeşilova, 2003; SAS, 2008). Gözlenemeyen heterojenliğin belirlenmesinde kullanılan yöntemlerden biri karışımlı lojistik regresyon (KLR)'dir. KLR'de, veri setinin farklı alt populasyonlardan oluşan heterojen bir populasyondan elde edilmiş olduğu varsayılmaktadır. KLR'de, veri setinin dahil olacağı homojen alt populasyonların sayısı belirlenerek, gözlenemeyen heterojenlik giderilmeye çalışılmaktır. Daha sonra her bir alt populasyon için ayrı parametre tahminleri elde edilmektedir (Wang vd.,1996; Wang vd., 1998; Wang ve Putterman, 1998; Okut vd., 2001; Leisch, 2004; Kaya, 2007). Bağımlı değişkenin ikili olması durumunda karışımlı ikili lojistik regresyon (İKLR) kullanılmaktadır.

İKLR'de parametre tahminleri, EM (Expectation-Maximization) algoritması kullanılarak en çok olabilirlik yöntemi ile elde edilir (Dempster vd., 1977). Veri setini en iyi açıklayan modelin seçiminde, Akaiki bilgi kriteri (AIC) ve Bayesian bilgi kriteri (BIC) en çok kullanılan model uyum kriterleridir (Wang ve Putterman, 1998; Dalrymple vd., 2002; SAS, 2008).

Bu çalışmada, İKLR modelinin teorik özellikleri incelenerek, Beden Eğitimi ve Spor Öğretmenliği alanında elde edilen gerçek bir veri setine uygulaması yapılmıştır. İlk olarak, veri setinin tek bir populasyondan elde edilmiş olduğu varsayılarak LR analizi yapılmıştır. Daha sonra LR analizi sonucunda oluşan aşırı yayılımı gidermek için veri seti İKLR analizine tabi tutulmuştur. BMLR'de her bir alt populasyon için ayrı parametre tahminleri ve alt populasyonlara düşen bireylerin sayıları tahmin edilmiştir.

2. VERİ SETİ

Bu çalışmada kullanılan veri seti, 2005–2006 öğretim yılı için Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Beden Eğitimi ve Spor Öğretmenliği Bölümü için açılan yetenek sınavına başvuran toplam 467 erkek adaydan oluşmuştur. Verilerin bir bölümü (ÖSS puanı, Ağırlıklı Orta Öğretim Başarı Puanı=AOÖBP) Öğrenci Seçme ve Yerleştirme Merkezi'nin (ÖSYM) internet sayfasından elde edilmiştir. Veri setini oluşturan diğer değişkenler ise sınav esnasında adaylardan yüz yüse alınmıştır. Adayların performans değişkenleri ise sınav esnasında adaylar izlenerek elde edilmiştir.

3. YÖNTEM

3.1. İkili Lojistik Regresyon

Karışımlı lojistik model için kesikli karışımlı dağılım (Wang vd., 1996; Leisch, 2004),

$$p(y) = \sum_{k=1}^{K} Binom(y / v_k \exp(\beta' x))\pi_k \qquad (1)$$

biçiminde yazılabilir. Burada π_k, k'ıncı alt populasyonun karışma olasılığını; y, bağımlı değişkeni; x, bağımsız değişken vektörünü; β, bilinmeyen parametre vektörünü; v, gamma dağılımına sahip rassal bir etki veya değişkeni göstermektedir. y_i, binom dağılımı gösterir ve,

$$P(Y_i = y_i) = \binom{n_i}{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i} \qquad (2)$$

biçiminde yazılır. Burada p_i, istenen olayın gerçekleşme olasılığı, n toplam deneme sayısı, y_i istenen başarılı olay sayısını belirtir. Lojistik regresyonda kullanılan logit bağlantı fonksiyonu,

$$Logit(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = \beta' x$$

olarak yazılabilir (Zhang, 1999). Bu durumda y değerlerine ilişkin marjinal olasılık yoğunluk fonksiyonu,

$$f(y) = \sum_{k=1}^{K} P(C = k) P(Y = y \mid C = k) = \sum_{k=1}^{K} \pi_k \, f(y, p_k) \qquad (3)$$

şeklinde yazılabilir (Yeşilova, 2004; Leisch, 2004). Binom dağılımlı veri setinin, K kadar alt populasyona ait heterojen bir örnek olması durumunda k'ıncı alt populasyona giren i'inci şans değişkeninin olasılığı (Okut vd., 2002),

$$\pi_{ik} = P(c_i = k)$$

biçiminde verilebilir. Bütün veriler için log-olabilirlik fonksiyonu,

$$\mathrm{L}(\mathrm{Y}, \mathrm{X}, \beta, \pi) = \sum_{i=1}^{n} \sum_{k=1}^{K} c_{ik} \log \pi_{ik} + \sum_{i=1}^{n} \sum_{k=1}^{K} c_{ik} \log(\mathrm{binom}(y_i \mid \beta_k, x)) \qquad (4)$$

biçiminde yazılabilir. Eşitlik 4'te, c_{ik} gözlenemeyen gözlemler olup,

$$C = \{c_{ik}, i = 1,2,...n; k = 1,2,...K\}$$

$$\begin{Bmatrix} c_{ik} = 1, c_{ik} \in K \\ c_{ik} = 0, diğer\ durumlar \end{Bmatrix}$$

biçiminde yazılabilir (Okut vd., 2002; Yeşilova, 2003).

3.2. İkili Lojistik Regresyon Modeli İçin EM Algoritması ve En Çok Olabilirlik Yöntemi

BMLR modeli için EM algoritmasının aşamaları aşağıdaki gibi verilebilir (Wang vd., 1996; Wang ve Putterman, 1998).

Birinci aşamada, $\beta^{(0)}$ ve $\pi_k^{(0)}$ başlangıç değerleri belirlenir.

E aşamasında, $\beta^{(0)}$ ve $\pi^{(0)}$ başlangıç değerleri verildiğinde gözlenmiş veriler (X, Y) ve parametrelerin başlangıç değerleri üzerinden, C eksik gözlemleri elde edilir. $\hat{C}_{ik}(\beta^{(0)}, \pi^{(0)})$ kullanılarak c_i 'nin k'ıncı unsurunun koşullu olasılığı,

$$\hat{c}_{i,k} = \left(\beta^{(0)}, \pi_k^{(0)}\right) = \frac{\pi_k \text{binom}\left(y_i \mid x_i, \beta_k^{(0)}\right)}{\sum_{k=1}^{K} \pi_k \text{binom}\left(y_i \mid x_i, \pi_k^{(0)}\right)}, \quad k = 1, 2, ..., K \tag{5}$$

biçiminde verilebilir.

M aşamasında ise z_i indikatör değişken olup parametre tahminleri, Eşitlik 4'te verilen log olabilirlik fonksiyonun β ve π 'ye göre maksimize edilmesi ile,

$$Q = \left(\beta^{(0)}, p \mid \beta^{(0)}, \pi_k^{(0)}\right) = E\left\{ (L(Y, C, \beta, p, X)) / Y, X, \beta^{(0)}, \pi_k^{(0)} \right\} \tag{6}$$

$$Q = Q_1 + Q_2$$

biçiminde elde edilir.Burada, Q_1 ve Q_2,

$$Q_1 = \sum_{i=1}^{n} \sum_{k=1}^{K} c_{i,k} \left(\beta^{(0)}, \pi_k^{(0)}\right) \log\left(\pi_k\right) \tag{7}$$

$$Q_2 = \sum_{i=1}^{n} \sum_{k=1}^{K} c_{i,k} \left(\beta^{(0)}, \pi_k^{(0)}\right) \log bi(y_i / n_i, \pi_{ik}) \tag{8}$$

biçiminde elde edilir. Eşitlik 7 ve 8'de verilen $\hat{\beta}$ ve $\hat{\pi}$ tahmin edicileri, Q_1 ve Q_2 eşitliklerinin π ve β 'ya göre türevlerinin alınması ile,

$$\frac{\partial Q_1}{\partial \pi_K} = 0\text{ , } k = 1,\ldots,K-1 \tag{9}$$

$$\frac{\partial Q_2}{\partial \beta} = 0 \tag{10}$$

biçiminde elde edilir. Eşitlik 9 kullanılarak $\hat{\pi}_k$,

$$\hat{\pi}_k = \frac{1}{n}\sum_{i=1}^{n}\hat{c}_{i,k}\ , \qquad k = 1,\ldots,K-1 \tag{11}$$

biçiminde elde edilmektedir (Wang vd., 1996; Wang vd., 1998; Wang ve Putterman., 1998). Yukarıda verilen Eşitlik 9 ve 10'da kapalı formunun çözümünün zor olmasından dolayı, parametre tahminleri için Quasi-Newton yaklaşımı kullanılarak E ve M aşamaları,

1. aşamada, $\beta^{(0)} = \left(\beta_1^{(0)},\ldots,\beta_k^{(0)}\right)$ ve $\pi^{(0)} = \left(\pi_1^{(0)},\ldots,\pi_k^{(0)}\right)$ başlangıç değerlerinin ε ve ε_0 tolerans değerlerine göre belirlenmesi,

2. aşamada (E-aşaması), 5.eşitlik kullanılarak

$$\hat{c}_i = \left(\hat{c}_{i,1},\ldots,\hat{c}_{i,K}\right)^{-1} \quad i = 1,2,\ldots,n \tag{12}$$

değerleri hesaplanılır. $\hat{c}_{i,k}$'nın hesaplanmasında aşırı yayılımı engellemek için Eşitlik 5'de verilen fonksiyonun payı , payda toplamı içinde yer alan en büyük değere bölünür.

3. aşamada (M-aşaması),

a) Quasi- Newton algortiması kullanılarak Eşitlik 9 'da $\hat{\pi}$ parametresinin hesaplanması.

b) Quasi-Newton algoritması kullanılarak Eşitlik 10'nun çözümünden $\hat{\beta}$ parametresinin hesaplanması.

4. aşamada, aşağıdaki koşullardan en az biri doğru ise,

$\beta^{(0)} = \hat{\beta}$, $\pi^{(0)} = \hat{\pi}$ olur ve 1. asamaya gidilir, aksi durumda c aşamasına gidilir.

1) $\left|\hat{\beta} - \beta^{(0)}\right| \geq \varepsilon$ (13)

2) $\left|\hat{\pi} - \pi^{(0)}\right| \geq \varepsilon$ (14)

3) $\left|L\left(Y,X,\hat{\beta},\hat{\pi}\right) - L\left(Y,X,\beta^{(0)},\pi^{(0)}\right)\right| \geq \varepsilon_0$ (15)

c) Quasi-Newton algoritması kullanılarak gözlenmiş $L(Y, X, \beta, \pi)$ log olabilirlik fonksiyonu maksimize edilerek işlem sonlandırılır (Wang ve ark., 1996; Wang ve ark., 1998).

3.3- Uygun Model Seçimi

Karışımlı modellerde uygun model seçimi için AIC ve BIC bilgi ölçütleri kullanılır. Uyum ölçütleri genel olarak;

$$AIC = -LogL + 2p \tag{16}$$

$$BIC = -LogL + p\ ln(n) \tag{17}$$

biçiminde yazılabilir. Burada, p parametre sayısını göstermektedir (Wang vd., 1996).

4. SONUÇLAR

Çalışmada, gerekli analizler SAS istatistik yazılım programı (PROCEDURE TRAJ) kullanılarak yapılmıştır. Öğrencilerin sınavı kazanıp kazanmaması (sınav sonucu) bağımlı değişken, Mekik Sayısı, ÖSS ve AOÖBP bağımsız değişkenler olarak modele alınmıştır. Veri setine ait tanımlayıcı istatistikler Çizelge 1'de verilmiştir.

Çizelge 1. Bağımsız değişkenlere ilişkin tanıtıcı istatistikler

Değişkenler	N	Ortalama+S.Sapma	Minimum	Maksimum
Mekik Sayısı	467	112.6±16.233	53.000	159.000
ÖSS Puanı	467	217.7±17.020	166.400	257.900
AOÖBP	467	78.19±6.475	64.660	98.4600

Veri setinin ilk önce lojistik regresyona göre analizi yapılmıştır. Lojistik regresyonunda, deviance uyum istatistiğine ilişkin yayılım parametre değeri 6.328, Pearson ki-kare uyum istatistiğine ilişkin değer ise 6.121 olarak bulunmuştur. Elde edilen uyum istatistiği değerlerinin 1 değerinden büyük çıkması veri setinde aşırı yayılım olduğunu göstermektedir (SAS, 2008).

Veri setindeki aşırı yayılım tespit edildikten sonra, karışımlı İKLR uygulanmıştır. İkili karışımlı lojistik regresyona ilişkin elde edilen model uyum ölçütleri Çizelge 2'de verilmiştir. Çizelge 2'ye bakıldığında iki alt populasyondan (alt populasyon–2) sonra AIC ve BIC uyum ölçütlerinin büyüdüğü görülmektedir. Birinci alt populasyon (lojistik regresyon) için uyum istatistikleri AIC=104.293 ve BIC=109.155 olarak bulunmuştur. En küçük AIC ve BIC değerlerine sahip model, veri setini en iyi açıklayan model olarak bilinmektedir. Bundan dolayı, koyu harflerle gösterilen iki alt populasyonlu model en iyi model olarak seçilmiştir.

Çizelge 2. Farklı alt populasyonlar ilişkin uyum ölçütleri

Alt Populasyonlar	AIC	BIC
Alt Populasyon 1	104.293	109.155
Alt Populasyon 2	**52.672**	**67.260**
Alt Populasyon 3	60.533	84.847
Alt Populasyon 4	92.287	126.326

En iyi model olarak seçilen iki alt populasyonlu modelde, her bir alt populasyon için elde edilen parametrelerin ortalama değerleri Çizelge 3'te verilmiştir. Birinci alt populasyonda ortalama mekik sayısı 100.133, ortalama ÖSS puanı 209.249 ve ortalama AOÖBP puanı 78.362 olarak elde edilmişken, ikinci alt populasyonda ortalama mekik sayısı 115.892, ortalama ÖSS puanı 219.960 ve ortalama AOÖBP puanı 78.142 olarak elde edilmiştir.

Çizelge 3. İki alt populasyonlu modele ait ortalama parametre değerleri

Alt Populasyon	Mekik Sayısı	ÖSS	AOÖBP
1	100.133	209.249	78.362
2	115.892	219.960	78.142

İkili karışımlı lojistik regresyon için elde edilen parametre tahmin değerleri ve standart hataları Çizelge 4'te verilmiştir. Mekik sayısı, OSS ve AOÖBP bağımsız değişkenlerin tamamının sınavı kazanma üzerine olan etkileri her iki alt populasyonda önemli bulunmuştur ($p<0.01$).

Çizelge 4. Karışımlı lojistik regresyon analiz sonuçları

Alt Populasyon	Karışma Olasılıkları (%)	β_0	Mekik Sayısı	OSS	AOÖBP
Alt Populasyon 1	20.9	1642.460	21.054 (0.044)**	-15.766 (0.027)**	0.037 (0.014)**
Alt Populasyon 2	79.1	4010.235	30.726 (0.031)**	0.010 (0.020)**	0.044 (0.045)**

** $p<0.01$

İki alt populasyonlu modelde, bireylerin her bir alt populasyona dağılma olasılıkları ve sayıları Çizelge 5'de verilmiştir. Birinci alt populasyona bireylerin 58 (% 20.9), ikinci alt populasyona ise 409 (% 79.1) dahil olmuştur.

Çizelge 5. Adayların alt populasyonlara dağılımı

Alt Populasyon	Sayı	Oran %
Alt Populasyon 1	58	20,9
Alt Populasyon 2	409	79,1

5. TARTIŞMA

Binom dağılımda, gözlenen varyansın beklenen varyanstan büyük olduğu durum, aşırı yayılım veya extra-binomial varyasyon olarak adlandırılmaktadır (Cox, 1983; Lindsey, 1998; Kaya, 2007). Bu durumda lojistik regresyonun kullanılması doğru ve tutarlı olmayan parametre tahminlerin ve standart hataların elde edilmesine neden olmaktadır. Çizelge 2'de farklı alt populasyonlu modeller için hesaplanan AIC ve BIC uyum ölçütleri iki alt populasyonlu modelden sonra giderek büyüdüklerinden dolayı, dört alt populasyonlu modelden sonraki alt populasyonlu modellere yer verilmiştir.

Çalışmada, lojistik regresyonda meydana gelen aşırı yayılım, veri seti kendi içerisinde iki alt populasyona ayrılarak giderilmiştir. Böylece her bir alt populasyon içi homojenlik sağlanırken, alt populasyonlar arası heterojenlikte otaya konmaya çalışılmıştır. Çizelge 3'de birinci alt populasyonda mekik sayıları ve ÖSS puanlarının ortalama değerlerinin ikinci alt populasyondan elde edilen ortalama değerlere göre daha küçük olduğu saptanmıştır. Mekik sayısı, ÖSS ve AÖOBP puanları öğrencilerin sınavı kazanmalarında doğrudan etkili olan faktörlerdir (Kaya, 2007). Çalışmada, özellikle mekik sayısı ve ÖSS puanın, öğrencilerin sınavı kazanmalarında doğrudan etkili olduğu belirlenmiştir. Bununla birlikte, AÖOBP ortalama değeri her iki alt populasyonda benzerlik göstermiştir. Bu bakımdan, Çizelge 4'de verilen bağımsız değişkenlerin tamamının, sınavı kazanma üzerindeki etkilerinin önemli olması, Çizelge 3'de verilen ortalama değerleri ile birbirlerini desteklemektedir. Özellikle mekik sayısı ve ÖSS puanının her iki alt populasyonda farklılık göstermeleri, bu her iki özelliğin sınav sonucunu doğrudan etkiledikleri söylenebilir. Bu bağlamda, ikinci alt populasyondaki öğrencilerin sınavı kazanma şanslarının daha yüksek olduğu saptanmıştır. Başka bir ifadeyle, birinci alt populasyona dahil edilen adayların 58 (% 20.9)'inin sınavı kazanma şanlarının ikinci alt populasyondaki 409 (% 79.1) adaydan daha az olduğu saptanmıştır. Bunun yanı sıra, AÖOBP her iki alt populasyonda da yakın değerler almıştır.

Bu çalışmada, bağımlı değişkenin ikili olduğu durumlarda veri setinde meydana gelene aşırı yayılımı modellemek için karışımlı ikili lojistik model kullanılmıştır. Elde edilen sonuçlar doğrultusunda, karışımlı ikili lojistik modelin, logistic regresyonda meydana gelen aşırı yayılım modellemede oldukça etkin olduğu belirlenmiştir.

KAYNAKLAR

G. E. Bonney (1987): "Logistic Regression for Dependent Binary Observations: Biometrics", Cilt 43, No. 4, s. 951-973.

R. Cox (1983): "Some Remarks on Overdispersion: Biometrika", Cilt 70, s. 269-274.

M. L. Dalrymple, I. L. Hudson, R. P. K. Ford (2003): "Finite Mixture, Zero-Inflated Poisson and Hurdle Models with Application to SIDS", Computational Statistics and Data Analysis, Cilt 41, s. 491-504.

A. P. Dempster, N. M. Laird, D. B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algrithm", Journal of Royal Statisticial Society, Cilt 39, s. 1-18.

J. A. Dobson (1990): "An Introduction to Generalized Linear Models", New York: Chapman and Hall.

B. Jones, S. D. Nagin, K. Roeder (2001): "A SAS Procedure Based on Mixture for Estimating Developmental Trajectories", Sociological Methods and Research, Cilt 29, No. 3, s. 374-393.

Y. Kaya (2007): "Binary Karışımlı Lojistik Regresayon (Yüksek Lisans Tezi, Basılmamış), Yüzünyü Yıl Üniversitesi, Fen Bilimleri Enstitüsü.

D. Lambert, K. Roeder (1995): "Overdispersion Diagnostics for Generalized Linear Models", Journal of the Amarican Statistical Association, Cilt 90, No. 432, s. 1225-1236.

F. Leisch (2004): "FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R", Journal of Statistical Software, Cilt 11, No. 8.

J. K. Lindsey (1998): "On the Use of Corrections for Overdispersion" Appl. Statist, Cilt 48, No. 4, s. 553 561.

McCullagh, P., Nelder, JA. (1989). Generalized Linear Models: Second Edition, , London, Chapmann and Hall, 486.

H. Okut, T. E. Duncan, S. C. Duncan, L. A. Strycker (2002): "Growth Mixture Modeling of Zero-Inflated Count Data", J. of Psychopathology and Behavioral Assessment.

SAS (2008): "SAS/STAT Software:Hangen and Enhanced. SAS", Inst. Inc., USA.

M. E. Stokes, C. S. Davis, G. G. Koch (2000): "Categorical Data Analysis Using the SAS System", John Wiley and Sons, USA.

P. Wang, I. M. Cockburn, M. L. Puterman (1998): "Analysis of Patent Data-Mixed Poisson Regression Model Approach", Journal of Business and Economic Statistics, Cilt 16, No. 1, s. 27-41.

P. Wang, M. L. Puterman, I. M. Cockburn, N. Le (1996): "Mixed Poisson Regression Models with Covariate Dependent Rates", Biometrics, Cilt 52, s. 381-400.

P. Wang, M. L. Putterman (1998): "Mixed Logistic Regression Models", Journal of Agriculture, Biological and Environmental Statistics, Cilt 3, No. 2, s. 175-200.

A. Yeşilova (2003): "The Use of Mixed Poisson Regression Models for Categorical Data in Biology", Doktora Tezi, Yüzüncü Yıl Üniversitesi, Van.

B. Zhang (1999): "A Chi-Squared Goodness-of-Fit-Test for Logistic Regression Models Based on Case-Control Data", Biometrika, Cilt 86, s. 531-539.