İki Durumlu Karışımlı Lojistik Regresyona İlişkin Bir Uygulama

Yılmaz KAYA[1], Abdullah YEŞİLOVA[2]

[1] Bilgisayar Mühendisliği Bölümü, Siirt Üniversitesi, Siirt, Türkiye
[2] Biyometri &Genetik Bölümü, Yüzüncü Yil Üniversitesi, Van, Türkiye
ykaya@yyu.edu.tr, yesilova@yyu.edu.tr
(Geliş/Received: 01.07.2011, Kabul/Accepted: 15.09.2011)

***Özet*─** Lojistik regresyon, binary (iki durumlu) bağımlı değişken ile bağımsız değişkenler arasındaki neden sonuç ilişkisini incelemektedir. Lojistik regresyonda, gözlenen varyansın, beklenen varyanstan büyük olması aşırı yayılım olarak tanımlanmaktadır. Veri setinde meydana gelen aşırı yayılımın modellenmesinde kullanılan alternatif yöntemlerden biri de karışımlı lojistik regresyondur. Karışımlı modellemede, aşırı yayılıma gözlenemeyen heterojenliğin neden olduğu varsayılmaktadır. Veri seti kendi içerisinde homojen alt veri setlerine ayrılarak, aşırı yayılım giderilmektedir. Bu çalışmada 2005–2006 öğretim yılı için YYÜ Eğitim Fakültesi Beden Eğitimi ve Spor Öğretmenliği Bölümü için açılan yetenek sınavına katılan 467 erkek adayın başarı durumları karışımlı modeller ile incelenmiştir. Adaylardan elde edilen veri seti aşırı yayılım göstermiştir. Aşırı yayılım, veri setinin homojen iki alt veri setine indirgenmesi ile giderilmiştir. Her alt veri seti için elde edilen parametre tahminlerine göre bireylerin sınavı kazanmalarında mekik sayıları, ÖSS ve AOÖBP değişkenleri etkili oldukları saptanmıştır (p<0.05).

***Anahtar Kelimeler*─** AIC (Akaike Information Criteria), BIC (Bayesian Information Criteria), EM (Expection-Maximization) algoritması, Lojistik regresyon, Karışımlı model

An Application for Binary Mixture Logistic Regression

***Abstract*─** The logistics regression, examines the reason–result relation between the binary dependent variable and independent variables. In logistics regression, observed variance being more than expected variance is called overdispersion. In cases where there is overdispersion in the data set, mixture logistics regression is an alternative method. In mixture modeling, it is assumed that the data set shows a heterogeneous structure. This heterogeneity is defined as unobservable heterogeneous. The data set is separated to homogenous sub data sets within itself in order to overcome the overdispersion. In this study, the status exam success of 467 male candidates that participated to the skill exam made in 2005–2006 education year in YYÜ Faculty of Education Gymnastics and Sports Teaching Department were examined with mixture models. The data set obtained from candidates was showed overdispersion. The overdispersion in data set was obviated by the separating data to homogenous two sub data sets. According to the parameter estimates for each sub data set, the number flies, ÖSS points and AOÖBP variables effect the result of the exam directly (p<0.05).

***Keywords*─** AIC (Akaike Information Criteria), BIC (Bayesian Information Criteria), EM (Expection-Maximization) Algorithm, Logistic Regression, Mixture model

1. GİRİŞ [*]

Lojistik regresyon (LR), bağımlı değişkenin binom dağılımı gösterdiği durumlarda kullanılmaktadır. Başka bir ifadeyle LR, bağımlı değişkenin ikili veya sıralı olması durumunda bağımlı değişken ile bağımsız değişkenler arasındaki neden sonuç ilişkisini belirlemede kullanılan bir yöntemdir [1,2,3,4]. LR'de, bağımsız değişkenlerin doğrusal yapısı, iki durumlu bağımlı değişkeninin beklenen değerine bağlayan bir link fonksiyonunu kullanmaktadır. Genellikle logit, probit veya couchit gibi bağlantı fonksiyonları kullanılmaktadır [3,5,6]. Lojistik regresyonda, gözlenen varyansın, beklenen varyanstan büyük olması aşırı yayılım (overdispersion) olarak tanımlanmaktadır [7,8,9].

Aşırı yayılım, genellikle gözlenemeyen heterojenliğin neden olduğu bir durumdur [4,10,11,13,14]. Gözlenemeyen heterojenliğin belirlenmesinde kullanılan yöntemlerden biri karışımlı lojistik regresyondur (KLR;Mixture Logistic Regression=MLR). Karışımlı

[*] Bu çalışma 2007 yılında bitirilen Yılmaz KAYA'nın yüksek lisans tezinden alınmıştır.

lojistik regresyonda, veri setinin farklı alt veri setlerinden oluşan heterojen bir veri seti olduğu varsayılmaktadır. KLR'de amaç, veri setinin dağılacağı homojen alt veri setlerin sayısı belirleyerek, gözlemlerin alt veri setlerine dağılmasını sağlamaktır. Daha sonra her bir alt veri seti için ayrı parametre tahminleri elde edilir [10,11,12,15,16,17]. Aşırı yayılım gösteren veri setleri için bağımlı değişkenin ikili olması durumunda iki durumlu karışımlı lojistik regresyon (IKLR: Binary Mixture Logistic Regression=BMLR) kullanılmaktadır [17].

IKLR'de parametre tahminleri, beklenti maksimizasyonu (Expectation-Maximization=EM) algoritması kullanılarak en çok olabilirlik (Maksimum Likelihood=ML) yöntemi ile elde edilir [18]. Veri setini en iyi açıklayan modelin seçiminde, Akaiki bilgi ölçütü (Akaike Information Criteria=AIC) ve Bayesian bilgi ölçütü (Bayesian Information Critera=BIC) en çok kullanılan model uyum ölçütleridir [4,12,13,19].

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

2. VERİ SETİ

Bu çalışmada kullanılan veri seti, 2005–2006 öğretim yılı için YYÜ Eğitim Fakültesi Beden Eğitimi ve Spor Öğretmenliği Bölümü için açılan yetenek sınavına katılan 467 erkek adaydan oluşmuştur. Verilerin bir bölümü (ÖSS puanı, AOÖBP) Öğrenci Seçme ve Yerleştirme Merkezi'nin (ÖSYM) web sayfasından elde edilmiştir. Veri setini oluşturan diğer değişkenler ise sınav esnasında adaylardan bire bir alınmıştır. Adayların performans değişkenleri ise sınav esnasında adaylar izlenerek elde edilmiştir. Adayların sınavı kazanıp kazanmaması (sınav sonucu) bağımlı değişken, Mekik Sayısı, Öğrenci Seçme Sınavı(ÖSS) puanı ve Ağırlıklı Orta Öğretim Başarı Puanları (AOÖBP) bağımsız değişkenler olarak modele alınmıştır. Veri setine ait tanıtıcı istatistikler Tablo 1'de verilmiştir.

Tablo 1. Bağımsız değişkenlere ilişkin tanıtıcı istatistikler

Değişken	N	Ortalama±Standart 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

3. YÖNTEM

3.1. İki Durumlu Karışımlı Lojistik Regresyon

Karışımlı lojistik model için kesikli karışımlı dağılım [10,16],

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

biçiminde yazılabilir. Burada π_k, k'ıncı alt veri setine ait olma olasılığını; y, bağımlı değişkeni; x, bağımsız değişken vektörünü; β, bilinmeyen parametre vektörünü; v, gamma dağılışına sahip tesadüfi bir etki veya değişkeni göstermektedir. y_i , binomial dağılış gösterir ve,

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

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,

$$Logit(\pi) = \beta' x$$

olarak yazılabilir [2]. 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 [14,16]. Binom dağılışlı veri setinin, K kadar alt popülasyona ait heterojen bir örnek olması durumunda k'ıncı alt popülasyona (kategoriye) giren i'inci şans değişkeninin olasılığı [15] ,

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

biçiminde verilebilir. Bütün veriler için log-olabilirlik fonksiyonu,

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

biçiminde yazılabilir. 4 numaralı eşitlikte, c_{ik} gözlenemeyen gözlemler olup,

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

$$\begin{Bmatrix} c_{ik} = 1, c_{ik} \in K \\ c_{ik} = 0, diger\ durumlarda \end{Bmatrix}$$

biçiminde yazılabilir [14,15].

3.2. İki Durumlu Karışımlı Lojistik Regresyon modeli için EM algoritması ve en çok olabilirlik yöntemi

IKLR modeli için EM algoritmasının aşamaları aşağıdaki gibi verilebilir [10,12]

Birinci aşamada, $\beta^{(0)}$ ve $\pi_k^{(0)}$ başlangıç değerleri ε ve ε_0 tolerans değerlerine göre 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,

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

biçiminde verilebilir.

M aşamasında,

$$\left\{c_i\left(\beta^{(0)},\pi_k^{(0)}\right)=\left(z_{i,1},...,z_{i,K}\right)';\ i=1,2,...,n\right.$$

koşullu olasılıkları verilmişken, parametre tahminleri, eşitlik 4'te verilen log olabilirlik fonksiyonun β ve π 'ye göre maksimize edilmesi ile,

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

$$Q=Q_1+Q_2$$

biçiminde elde edilir.Burada, Q_1 ve Q_2,

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

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

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,

$$\frac{\partial Q_1}{\partial \pi_K}=0,\ k=1,...,K-1 \quad (9)$$

$$\frac{\partial Q_2}{\partial \beta}=0 \quad (10)$$

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

$$\hat{\pi}_k=\frac{1}{n}\sum_{i=1}^{n}\hat{c}_{i,k},\ k=1,...,K-1 \quad (11)$$

biçiminde elde edilmektedir [10,11,12]. Yukarıda verilen eşitlik 10'da kapalı formunun çözümünün zor olmasından dolayı, parametre tahminleri için Quasi-Newton yaklaşımı kullanılarak E ve M aşamaları,

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

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

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

değerleri hesaplanılır. $\hat{c}_{i,k}$ 'nın hesaplanmasında aşırı taşmayı engellemek için eşitlik 5'de verilen fonksiyonun pay paydası, payda toplamının en büyük değerine bölünür.

3. aşamada (M- aşaması),

a) Eşitlik 9 kullanılarak $\hat{\pi}$ parametresinin hesaplanması

b) Quasi-Newton algoritması kullanılarak 10 numaralı eşitliğin 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.

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

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

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

c) β parametreleri için, $\hat{\pi}$ 'ler başlangıç değerleri olarak alınır ve Quasi-Newton algoritması kullanılarak gözlenmiş $L(Y,X,\beta,\pi)$ log olabilirlik fonksiyonu maksimize edilip, işlem sonlandırılır [10,11].

3.3. *Uygun Model Seçimi*

Karışımlı modellerde uygun model seçiminde kullanılan uyum ölçütleri Akaiki Bilgi Ölçütü (Akaike Information Criteria=AIC) ve Bayesian bilgi ölçütü (Bayesian Information Criteria=BIC). AIC ve BIC uygun model seçimi için kullanılır. Uyum ölçütleri genel olarak;

$$AIC=-LogL+2p \quad (16)$$

$$BIC=-LogL+p\ln(n) \quad (17)$$

biçiminde yazılabilir. Burada, p parametre sayısını göstermektedir [10].

4. UYGULAMALAR

Çalışmada, analizler için R yazılımı kullanılmıştır. R istatistik, matematik, veri madenciliği gibi çok farklı amaçlar için kullanılabilecek bir programlama dilidir [20]. Açık kaynaklı bir program olup GNU lisansı altında dağıtılmaktadır.

Veri setinin ilk önce lojistik regresyona göre analizi yapılmıştır. Lojistik regresyonda, devians (deviance) uyum istatistiğine ilişkin yayılım parametre değeri 6.328, Pearson Khi-kare uyum istatistiğine ilişkin değer ise 6.121 olarak bulunmuştur. Elde edilen uyum istatistiği değerlerinin bir (1) değerinden büyük çıkması veri setinde aşırı yayılım olduğunu göstermektedir [4].

Veri setindeki aşırı yayılım tespit edildikten sonra, iki durumlu karışımlı lojistik regresyon uygulanmıştır. İki durumlu karışımlı lojistik regresyona ilişkin elde edilen model uyum ölçütleri Tablo 2'de verilmiştir. Tablo 2'ye bakıldığında iki alt veri setinden (alt veri seti–2 adet) sonra AIC ve BIC uyum ölçütlerinin büyüdüğü görülmektedir. İkinci alt veri seti için uyum istatistikleri AIC=52.672 ve BIC=67.260 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 veri setli model en iyi model olarak seçilmiştir. Alt veri seti seçme işlemine ait kaba kod aşağıda verilmiştir.

1. *BAŞLA;*
2. *P=1; //Alt veri seti sayısı*
3. *I=1;*
4. *AICi, BICi hesapla*
5. *P=P+1; I=I+1;*
6. *EM algoritmasına göre gözlemleri gruplara dağıt.*
7. *Her alt veri seti için parametre tahminlerini hesapla*
8. *AICi, BICi hesapla ve sakla;*
9. *Eğer AICi-1<AICi veya BICi-1<BICi GİT 11*
10. *GİT 5*
11. *DUR*
12. *YAZ ("Alt Veri Sayısı="+P);*
13. *BİTİŞ*

Tablo 2. Farklı alt veri setleri için uyum ölçütleri

# Alt Veri Setleri	AIC	BIC
1 adet	104.293	109.16
2 adet	**52.672**	**67.260**
3 adet	60.533	84.847
4 adet	92.287	126.33

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

Tablo 3. İki alt veri setli modele ait ortalama parametre değerleri

Alt Veri Setleri	Mekik Sayısı	ÖSS	AOÖBP
1.alt veri seti	100.133	209.249	78.362
2.alt veri seti	115.892	219.960	78.142

İki durumlu karışımlı lojistik regresyon için elde edilen parametre tahmin değerleri ve standart hataları Tablo 4'te verilmiştir. Mekik sayısı ve ÖSS puanının alt veri setleri arasındaki önemli farkı, bu iki değişkenin sınavı kazanma üzerine direk bir etkisinin olduğu görülmektedir. Tablo 4'te verilen parametre tahminlerine göre AOÖBP değişkenin alt veri seti içi sınav başarısı üzerinde çok küçük bir etkiye sahip olduğu görülmektedir.

Tablo 4. Karışımlı lojistik regresyon analiz sonuçları

Alt Veri Setleri	Dağılım Oranları (%)	Inter cept	Mekik Sayısı (Standart Hatalar)	OSS	AOÖBP
1	20.9	1643	21.1 (0.044)*	-15.8 (0.027)*	0.04 (0.014)*
2	79.1	4010	30.7 (0.031)*	0.01 (0.020)*	0.04 (0.045)*

* $p<0.05$

İki alt veri setli modelde, bireylerin her bir alt veri setine dağılma oranları Tablo 5'te verilmiştir. Birinci alt veri setinde bireylerin 58'i (20.9%), ikinci alt veri setine ise 409'u (79.1%) dahil olmuştur. Her iki alt veri setindeki ($p<0.05$) değerlerine göre sınavı kazanmada Mekik Sayısı, ÖSS ve AÖOBP değişkenlerinin önemli olduğu saptanmıştır.

Tablo 5. Adayların alt veri setlerine dağılımı.

Alt Veri Setleri	Sayı	Oran %
Alt Veri Seti	58	20,9
Alt Veri Seti	409	79,1

5. TARTIŞMA

Binom dağılımda, gözlenen varyansın beklenen varyanstan büyük olduğu durum, aşırı yayılım (overdispersion) veya extra-binomial varyasyon olarak adlandırılmaktadır [7,9,17]. Bu durumda lojistik regresyonun kullanılması doğru ve tutarlı olmayan parametre tahminlerinin ve standart hataların elde edilmesine neden olmaktadır. Çalışmada, lojistik regresyonda meydana gelen aşırı yayılım, veri seti kendi içerisinde iki alt veri setine ayrılarak giderilmiştir. Böylece her bir alt veri seti içi homojenlik sağlanırken, alt veri setleri arası heterojenlikte ortaya konmaya çalışılmıştır.

Tablo 2'de farklı alt veri setli modeller için hesaplanan AIC ve BIC uyum ölçütleri iki alt veri setli modelden sonra giderek büyüdüklerinden dolayı, dört alt veri setli modelden sonraki alt veri setli modellere yer verilmemiştir.

Tablo 3'te birinci alt veri seti için mekik sayıları ve ÖSS puanlarının ortalama değerlerinin ikinci alt veri setinden elde edilen ortalama değerlere göre daha küçük olduğu saptanmıştır. Çalışmada, özellikle mekik sayısı ve ÖSS puanın, öğrencilerin sınavı kazanmalarında doğrudan etkili olduğu belirlenmiştir. Bununla birlikte, AÖOBP ortalama değeri her iki alt veri seti için benzerlik göstermiştir. Bu bakımdan, Tablo 4'te verilen bağımsız değişkenlerin tamamının, sınavı kazanma üzerindeki etkilerinin önemli olması, Tablo 3'de verilen ortalama değerleri ile birbirlerini desteklemektedir. Tablo 3'e bakıldığında özellikle mekik sayısı ve ÖSS puanının her iki alt veri seti için farklılık göstermeleri, bu her iki özelliğin sınav sonucunu doğrudan etkiledikleri söylenebilir.

Bu çalışmada, bağımlı değişkenin iki durumlu olduğu durumlarda veri setinde meydana gelene aşırı yayılımı modellemek için iki durumlu karışımlı lojistik model kullanılmıştır. Elde edilen sonuçlar doğrultusunda, iki durumlu karışımlı lojistik modelin, lojistik regresyonda meydana gelen aşırı yayılım modellemede önemli bir yöntem olduğu gözlenmiştir.

KAYNAKLAR

[1]. G. E. Bonney, Logistic Regression for Dependent Binary Observations, *Biometrics*, 43(4), 951-973, 1987.
[2]. B. Zhang, A Chi-Squared Goodness-of-Fit- Test for Logistic Regression Models Based on Case-Control data, *Biometrika*, 86, 531-539, 1999.
[3]. M. E. Stokes, C. S. Davis, G. G. Koch, **Categorical Data Analysis Using the SAS System**, John and Wiley & Sons Incorporation, USA, 2000.
[4]. SAS, **SAS/STAT Software:Hangen and Enhanced**. SAS, Inst. Inc., USA, 2008
[5]. P. McCullagh, J. A. Nelder, **Generalized Linear Models**: Second Edition, Chapmann and Hall, London, 1989.
[6]. J. A. Dobson, **An introduction to generalized linear models**: New York: Chapman and Hall, 174, 1990.
[7]. R. Cox, Some Remarks on Overdispersion, *Biometrika*, 70, 269-274, 1983.
[8]. D. Lambert, K. Roeder, Overdispersion Diagnostics for Generalized Linear Models, *Journal of the American Statistical Association*, 90(432), 1225-1236, 1995.
[9]. J. K. Lindsey, On the Use of Corrections for Overdispersion, *Appl. Statist*, 48(4), 553 561, 1998.
[10]. P. Wang, M. L. Puterman, I. M. Cockburn, N. Le, Mixed Poisson Regression Models with Covariate Dependent Rates, *Biometrics*, 52, 381-400, 1996
[11]. P. Wang, I. M. Cockburn, M. L. Puterman, Analysis of Patent Data- Mixed Poisson Regression Model Approach, *Journal of Business and Economic Statistics*, 16(1), 27-41, 1998.
[12]. P. Wang, M. L. Putterman, Mixed Logistic Regression Models: Journal of Agriculture, *Biological and Environmental Statistics*, 3(2), 175-200, 1998.
[13]. B. Jones, S. D. Nagin, K. Roeder, A SAS Procedure Based on Mixture for Estimating Developmental Trajectories, *Sociological Methods and Research*, 29(3), 374-393, 2001.
[14]. A. Yeşilova, **The Use of Mixed Poisson Regression Models for Categorical Data in Biology**, Ph. D. Dissertation, University of Yüzüncü Yıl, Van, 2003.
[15]. H. Okut, T. E. Duncan, S. C. Duncan, L. A. Strycker, Growth Mixture Modeling of Zero-Inflated Count Data:, *J. of Psychopathology and Behavioral Assessment*, 2002.
[16]. F. Leisch, FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R, *Journal of Statistical Software*, 11(8), 2004.
[17]. Y. Kaya, **Binary Karışımlı Lojistik Regresayon** (Yüksek Lisans Tezi,Basılmamış). Yüzünyü Yıl Üniversitesi Fen Bilimleri, 2007.
[18]. M. L. Dalrymple, I. L. Hudson, R. P. K. Ford, Finite Mixture, Zero-Inflated Poisson and Hurdle Models with Application to SIDS, *Computational Statistics & Data Analysis*, 41, 491-504, 2003.
[19]. A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of Royal Statistical Society*, 39, 1-18, 1977.
[20]. Internet: R Project for statistical computing. http://www.r-project.org, 2011.