

## Lojistik Regresyonda Meydana Gelen Aşırı Yayılımın İncelenmesi

Abdullah YEŞİLOVA<sup>(1)</sup>

İsmail KASAP<sup>(2)</sup>

**Öz:** Lojistik regresyonda, gözlenen varyansın beklenen varyansdan büyük olması aşırı yayılım ya da ekstra-binomiyal varyasyon olarak tanımlanmaktadır. Bu durumda Lojistik regresyon yerine, ekstra-binomiyal varyasyonu dikkate alan yöntemlerin kullanılması gerekir. Lojistik regresyonda, Pearson Khi-kare ile sapma uyum istatistikleri aşırı yayılımı saptamada kullanılmaktadır. Bu çalışmada, bitki koruma alanında elde edilmiş olan gerçek bir veri kümesi kullanılarak Williams metodu ile lojistik regresyonda meydana gelen aşırı yayılım kontrol altına alınmıştır. Williams metodu kullanılarak sapma uyum istatistiği bir (1) değerine çok yakın, Pearson uyum istatistiği ise bir (1) değerine eşit olarak elde edilmiştir. Elde edilen sonuçlar bakımından, uyum istatistikleri ile Akaike bilgi ölçütü ve Bayesian bilgi ölçütü benzerlik göstermişlerdir. Avcı akarlar karşı denenen ilaçlar, akarlar üzerine oldukça etkili olmuş ve akarların ölüm oranlarını arttırdığı saptanmıştır ( $p < 0.01$ ).

**Anhtar kelimeler:** Lojistik regresyon, extra-binomiyal değişim, uyum istatistikleri

### Investigation of Overdispersion in Logistic Regression

**Abstract:** In logistic regression, having higher observed variation than expected variation is called as overdispersion or extra-binomial variation. In this case, it is necessary to use the methods dealing with the extra-binomial variation rather than the logistic variation. Pearson chi-square and deviance goodness of statistics are used to determine overdispersion in logistic regression. In the present study using data obtained from plant protection studies, overdispersion in logistic regression was controlled by using Williams method. By using Williams method, a value very close to one (1) for deviance goodness of statistic and a value equal to one (1) for Pearson goodness of statistic were obtained. Based on the obtained results, goodness of statistics and Akaike information criteria and Bayesian information criteria model fit statistics have showed similarities. Tried pesticides against predator spider mites were very effective on spider mites, increasing the death rate of the spider mites ( $p < 0.05$ ).

**Key words:** Logistic regression, extra-binomial variations, goodness of statistics

### Giriş

Lojistik regresyon, binomiyal dağılım gösteren verilerin analizinde kullanılmaktadır. Lojistik regresyon bağımlı değişkenin ikili (binary) gözlemlendiği durumlarda, bağımlı değişken ile bağımsız değişkenlerin neden sonuç ilişkisini belirlemede kullanılan bir yöntemdir. Lojistik regresyonda, oluşturulan lojistik modellere göre parametre tahminleri elde edilir (Bonney, 1987; Lawles, 1987; Wang ve Putterman, 1998).

Lojistik regresyonda, geliştirilmiş doğrusal modeller (Generalized Linear Model) kullanılarak en çok olasılırlık yöntemi (maximum likelihood) ile parametre tahmini elde edilmektedir (McCullagh ve Nelder, 1989). Geliştirilmiş doğrusal modellerde, hatalar için normal dağılım varsayımına gerek yoktur. Bu yöntem ile binomiyal dağılımın olasılık fonksiyonu üssel formda tanımlanarak verilerin doğrudan beklenen değeri yerine, beklenen değerin doğrusal kombinasyonu kullanılmaktadır. Başka bir ifadeyle geliştirilmiş doğrusal modeller, bağımsız değişkenlerin doğrusal yapısını bağımlı değişkeninin beklenen değerine bağlayan bir bağlantı fonksiyonunu (link function) kullanmaktadır. Lojistik regresyonda, söz konusu bağlantı fonksiyonu logit dönüşüm ile verilmektedir.

(McCullagh ve Nelder, 1972; Stokes ve ark., 2000; SAS, 2007).

Lojistik regresyonda, gözlenen varyansın, beklenen varyansın büyük olması aşırı yayılım olarak tanımlanmaktadır (Cox, 1983; Dean, 1992). Sapma (deviance) ve Pearson Khi-kare uyum istatistikleri lojistik regresyonda yayılım olup olmadığını belirlemede kullanılan uyum istatistikleridir (Czado, 2003). Eğer uyum istatistikleri sonucunda aşırı yayılım söz konusu ise, aşırı yayılımı açıklayan modellerin kullanılması gerekmektedir. Lojistik regresyonunda, aşırı yayılımı açıklayan metotlardan biri de Williams metodudur (SAS, 2007). Williams metodunda, meydana aşırı yayılım modele dahil edilerek, giderilmektedir. Veri kümesini en iyi açıklayan modelin seçiminde, Akaike'nin bilgi ölçütü (Akaike Information Criteria=AIC) ve Bayesian bilgi ölçütü (Bayesian Information Criteria=BIC) en çok kullanılan model uyum ölçütleridir (Lawles, 1987; Wang ve Putterman, 1998). Bu çalışmada, bitki koruma alanında elde edilen gerçek bir veri kümesi kullanılarak aşırı yayılımın önce tespiti, daha sonrada nasıl modellendiği incelenmiştir.

<sup>(1)</sup> **Yazışma Adresi:** Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi, Zootehni Bölümü 65080-VAN, yesilova@yyu.edu.tr

<sup>(2)</sup> Çanakkale Onsekiz Mart Üniversitesi, Ziraat Fakültesi, Bitki Koruma Bölümü, 17020-ÇANAKKALE

## Materyal ve Yöntem

### Materyal

Çalışmanın materyalini oluşturan avcı akar *Kampimodromus aberrans* (Oudemans Acarina: Phytoseiidae) elma bahçelerinden toplanarak laboratuara getirilmiştir. Avcı akar *K. aberrans*, Van ve çevresinde elma bahçelerinde en yaygın olarak görülen akarlardan biridir (Kasap ve ark. 2004). Laboratuara getirilen avcı akarlar  $25 \pm 2$  °C sıcaklık, %  $60 \pm 10$  nem ve 16 saatlik aydınlanma süresine sahip iklim odasında üretimleri yapılmıştır. Avcıların bulunduğu küvetlere, bir huni ve fırça yardımı ile günlük olarak besin fırçalanmıştır. Besin olarak fırçalanan akarlara ek olarak avcı akarın gelişme ve üremesi için ihtiyaç duyduğu huş ağacının (*Betula pendula* Roth (Betulaceae)) polenleri ek besin olarak verilmiştir. Denemelerde Kazak ve Şekeroğlu (1996) ile Kılınçer ve ark. (1990)'ın bildirdiği şekilde mikroskop camı (lam) üzerine yapıştırılmış çift taraflı yapışkan bantlar (klişe bant) kullanılmış ve bu bantlar üzerine çiftleşmiş genç dişiler dorsal kısımlarından ince uçlu bir fırça yardımı ile yapıştırılmıştır. Her mikroskop camı üzerine 5 adet avcı akar olmak üzere her doz için 10 adet lam kullanılmıştır. Üzerinde dorsalden yapıştırılmış avcı akar bulunan lamlar, verilen ilaçların (Mavrik, malathion, neuron, trazam, zolone, folidol, ddvp, talstar, omite, supracide, cupravit, round-up) prospektüs bilgilerine göre önceden hazırlanan konsantrasyonlarına 5 sn süre ile daldırılmış ve avcı akarlar üzerinde biriken ilaç konsantrasyonu bir filtre kağıdına emdirilerek alınmıştır. Üzerinde biriken ilaç konsantrasyonları iyice kurutulduktan sonra, avcı akarlar  $25 \pm 2$  °C sıcaklık, % $60 \pm 10$  nem oranlarına ve 16:8 saat ışıklandırma süresine ayarlı iklim odasına yerleştirilmiştir. Gözlemler, laboratuvarında stereo binoküler mikroskop altında 24 saat sonra yapılmış, akarlara ince uçlu bir fırça yardımı ile dokunularak tepki verenler canlı, diğerleri ölü olarak kaydedilmiştir.

### Yöntem

Bu bölüm üç aşamada oluşmaktadır. İlk olarak lojistik regresyon ve uyum istatistikleri, sonraki bölümde Williams metodu ve son olarak model uyumuna ilişkin teorik bilgiler verilmiştir. Çalışmada, canlı akar/toplam akar  $r_i/n_i$  oranı bağımlı değişken, ilaçların etkisi de bağımsız değişken olarak modele alınmıştır.

### Lojistik regresyon

Lojistik regresyonda ilgilenilen olayın sayısı olan Y bağımlı değişkeni  $y_1, \dots, y_n$  biçiminde, X bağımsız değişkeni de  $x_1, \dots, x_n$  biçiminde verilmiş olsun. Burada Y değişkeni binom dağılımına sahip olduğu varsayılmaktadır. Her bir  $i=1, \dots, n$  için olasılık değeri,

$$\pi_i = \pi(x_i) = P(Y_i = 1/x_i) \quad (1)$$

Biçiminde verilebilir. Daha sonra lojistik regresyon modeli aşağıdaki biçimde yazılabilir,

$$\log \text{it}(\pi) = X\beta$$

ya da,

$$\log \text{it}(\pi_i) = \log \left( \frac{\pi_i}{1 - \pi_i} \right) = X_i' \beta \quad (2)$$

Eşitlik 2'te  $\beta$  bilinmeyen parametreleri göstermektedir (Bonney, 1987; Buis, 2005; SAS, 2007).

Sapma ve Pearson Khi kare istatistikleri lojistik regresyonda yaygın olarak kullanılan uyum istatistikleridir. Her iki uyum istatistik değerinin bire (1) eşit ya da çok yakın çıkması, lojistik regresyonda aşırı yayılım olmadığını anlamına gelmektedir. Sapma uyum istatistiği,

$$D^* = (y, \hat{\mu}) = \frac{1}{\phi} D(y, \hat{\mu}) \quad (3)$$

biçiminde yazılabilir (SAS, 2007). Eşitlik 3'te,  $D(y, \hat{\mu})$  model için hata sapması olmakta ve bireysel olarak sapmaların toplamını göstermektedir.  $\phi$  bilinmeyen skala parametresini (scale parameter) göstermektedir. Pearson Khi-kare uyum istatistiğine ilişkin teorik bilgiler Williams metodu konusu altında verilmiştir.

### Lojistik regresyonda aşırı yayılım

Lojistik regresyonda aşırı yayılıma başarı olasılıkları arasındaki varyasyon ve ikili cevap değişkenleri arasındaki korelasyon neden olmaktadır.  $Y_i$  bağımlı değişkenin beklenen değer ve varyansı,

$$E(Y_i) = n_i \pi_i$$

$$\text{Var}(Y_i) = n_i \pi_i (1 - \pi_i) \quad (4)$$

biçiminde verilebilir. İkili bağımlı değişkenleri arasında korelasyon olması durumunda,

$$Y_i = \sum_{j=1}^{n_i} R_{ij} \quad \text{ve} \quad R_{ij} = \begin{cases} 1 & \text{başarı} \\ 0 & \text{diğer} \end{cases}$$

$$P(R_{ij} = 1) = \pi_i$$

Biçiminde verilebilir. Bu durumda  $\text{Var}(Y_i)$ ,

$$\text{Var}(Y_i) = \sum_{j=1}^{n_i} \text{Var}(R_{ij}) + \sum_{j=1}^{n_i} \sum_{k \neq j} \text{cov}(R_{ij}, R_{ik})$$

$$\text{Var}(Y_i) = n_i \pi_i (1 - \pi_i) + \sum_{j=1}^{n_i} \sum_{k \neq j} \text{cov}(R_{ij}, R_{ik})$$

$$\text{Var}(Y_i) = n_i \pi_i (1 - \pi_i) [1 + (n_i - 1)\phi] \quad (5)$$

biçiminde olacaktır (Czado, 2005; SAS, 2007). Bu durumda eşitlik 4'te verilen beklenen varyans ile eşitlik 5'de verilen gözlenen varyans birbirine eşit değildir. Gözlenen varyans beklenen varyanstan ya büyük ya da küçük olmaktadır. Gözlenen varyansın beklenen varyanstan büyük çıkması aşırı yayılım, gözlenen varyansın beklenen varyanstan küçük çıkması da az yayılım olarak adlandırılmaktadır.  $\phi = 0$  olduğunda gözlenen varyans beklenen varyansa eşit olduğundan ekstra binomiyal yayılım yoktur.  $\phi > 0$  olduğunda ise veri kümesinde aşırı yayılım söz konusudur. Uygulamada, genellikle aşırı yayılım ile karşılaşılmaktadır (Lawles, 1987; Wang ve Putterman, 1998).

### Williams metodu

Veri kümesinin büyüklüğü  $n$  olmak üzere,  $i$ 'nci gözlem için  $r_i/n_i$  gözlenen oran değeri olsun.  $i$ 'nci gözlem için cevap olasılık değeri,  $p_i$  rasgele değişkeni olsun.  $p_i$ 'nin ortalama ve varyansı,

$$E(p_i) = \pi_i \text{ ve } V(p_i) = \phi \pi_i (1 - \pi_i)$$

biçiminde verilebilir.  $r_i$ 'nin ortalama ve varyansını aşağıdaki biçimde yazılabilir.

$$E(r_i) = n_i \pi_i$$

$$V(r_i) = n_i \pi_i (1 - \pi_i) [1 + (n_i - 1)\phi] \quad (6)$$

Williams metodunda, bilinmeyen  $\phi$  parametresinin tahmini, Pearson Khi-kare istatistik eşitliğinin yaklaşık beklenen değerine eşitlenmesiyle elde edilmektedir.  $w_i^*$ 'nin,  $i$ 'nci gözleme ilişkin ağırlık değeri olduğu varsayalım. Pearson Khi-kare istatistiği,

$$\chi^2 = \sum_{i=1}^n \frac{w_i^* (r_i - n_i \pi_i)^2}{n_i \pi_i (1 - \pi_i)}$$

biçiminde verilebilir.  $g'(\cdot)$ ,  $g(\cdot)$  bağlantı fonksiyonun birinci türevini gösterebilir.  $\chi^2$ 'nin yaklaşık beklenen değeri,

$$E(\chi^2) = \sum_{i=1}^n w_i^* (1 - w_i^* v_i d_i) [1 + \phi(n_i - 1)] \quad (7)$$

biçiminde yazılabilir. Eşitlik 7'de  $v_i = n_i / \left( \pi_i (1 - \pi_i) [g'(\pi_i)]^2 \right)$  ve  $d_i = \alpha_i + x_i \beta$  doğrusal tahminleyicisinin varyansını göstermektedir. Bilinmeyen  $\phi$  parametresi aşağıdaki iteratif işlem kullanılarak tahmin edilmektedir.

Birinci aşamada  $w_i^* = 1$  ve  $\pi_i$  yaklaşık olarak  $r_i/n_i$  değerini alarak başlanır.  $\phi$ 'nin başlangıç tahmini,

$$\phi_0 = \frac{\chi^2 - (n - p)}{\sum_i (n_i - 1)(1 - v_i d_i)}$$

biçiminde yazılır. Burada,  $p$  toplam parametre sayısını gösteriyor. Ağırlıkların başlangıç tahminleri,

$$w_{i0}^* = [1 + (n_i - 1)\phi_0]^{-1}$$

biçiminde olur. Modelin ağırlıklandırılmış uyumu yapıldıktan sonra  $\hat{\beta}$  ve  $\chi^2$  yeniden hesaplanır. Daha sonra tekrar düzeltilen  $\phi$ 'nin tahmini,

$$\phi_1 = \frac{\chi^2 - \sum_i w_i^* (1 - w_i^* v_i d_i)}{\sum_i w_i^* (n_i - 1)(1 - w_i^* v_i d_i)}$$

biçiminde verilebilir. İteratif işlem  $\chi^2$  uyum istatistiğinin kendi serbestlik derecesine yaklaşıncaya kadar devam edilir (SAS, 2007).

### Model uyumu

AIC ve BIC karışımli model için yaygın olarak kullanılan uyum ölçütleridir En küçük uyum ölçütüne sahip model en iyi model olarak kabul edilmektedir (Wang ve Putterman, 1998). Genel olarak uyum ölçütleri;

$$AIC = -2LL + 2p$$

$$BIC = -2LL + p \log(n)$$

biçiminde yazılabilir. Burada, LL: Lojistik regresyon modelinde iterasyon bittikten sonra elde edilen log-olabilirlik değeri,

$p$ : parametre sayısı

$n$ : örnek büyüklüğü olmaktadır.

### Bulgular

Veri kümesine ilk olarak lojistik regresyon uygulandı. Çizelge 1'de lojistik regresyona ilişkin uyum istatistikleri verilmiştir. Uyum istatistiklerinin yayılım parametre değerleri, hesap değerinin kendi serbestlik derecesine bölünmesi ile elde edilmektedir. Lojistik regresyonunda, sapma uyum istatistiğine ilişkin yayılım parametre değeri (Hesap değeri/sd) 3.491, Pearson Khi-kare uyum istatistiğine ilişkin değer (Hesap değeri/sd) ise 3.033 olarak bulunmuştur. Elde edilen uyum istatistiği değerlerinin bir (1) değerinden büyük çıktığı saptanmıştır. Sapma ve Pearson Khi-kare uyum istatistiği değerlerinin 1'den büyük çıkması veri kümesinde aşırı yayılım olduğunu göstermektedir.

Çizelge 1. Sapma ve Pearson Khi-kare uyum istatistikleri  
Table 1. Goodness of statistics for deviance and Pearson Chi-square

| Uyum istatistikleri<br>Goodness of statistics | sd/df | Hesap değeri<br>value | Hesap değeri<br>sd/value/df |
|---|-------|-----------------------|-----------------------------|
| Sapma/deviance                                | 63    | 219.953               | 3.491                       |
| Pearson Khi-kare                              | 63    | 191.063               | 3.033                       |

sd serbestlik derecesi, df: Degrees of freedom

Çizelge 2’de modele alınan değişkenlere ilişkin lojistik regresyon parametre tahminleri verilmiştir. Çizelge 2’de hem intercept’in hem de ilaçların etkisi önemli bulunmuştur ( $p<0.01$ ). Bundan dolayı avcı akarlar karşı denenen ilaçlar, akarlar üzerine oldukça etkili olmuş ve akarların ölüm oranlarını arttırdığı saptanmıştır.

Çizelge 2. Lojistik regresyonu sonucunda elde edilen parametre tahminleri  
Table 2. Parameter estimates for logistic regression

| Parametreler /Parameters        | Sd /df | Tahmin (Std. Hata<br>Estimation (std. Error)) |
|---------------------------------|--------|---|
| Intercept                       | 1      | -2.065(0.373)**                               |
| İlaçlama/insectifit application | 1      | 0.161(0.042)**                                |

sd serbestlik derecesi, df: Degrees of freedom, \*\*  $p<0.01$

Çizelge 1’de verilen uyum istatistikleri veri kümesinde aşırı yayılım olduğunu gösterdiğinden dolayı, aşırı yayılımı modellemek için Williams metodu uygulanmıştır. Williams Metodu kullanılarak lojistik regresyona ilişkin elde edilen uyum istatistikleri Çizelge 3’te verilmiştir.

Çizelge 3. Williams metodu kullanılarak elde edilen sapma ve Pearson Khi-kare uyum istatistikleri

Table 3. Goodness of statistics for deviance and Pearson Chi-square using Williams method

| Uyum istatistikleri<br>Goodness of statistics | Sd/df | Hesap değeri<br>value | Hesap değeri/sd<br>value/df |
|---|-------|-----------------------|-----------------------------|
| Sapma/Deviance                                | 63    | 65.915                | 1.046                       |
| Pearson Khi-kare                              | 63    | 62.998                | 1.000                       |

sd serbestlik derecesi, df: Degrees of freedom

Çizelge 3’e bakıldığında, Williams metodu kullanılarak sapma uyum istatistiğinin (Hesap değeri/sd) bir (1) değerine çok yakın bir değer, Pearson uyum istatistiğine ilişkin değer (Hesap değeri/sd) bir (1) değerine eşit olarak elde edildiği görülmektedir. Böylece Williams Metodunun lojistik regresyondaki yayılım parametresini kontrol etmede oldukça etkili olduğu görülmektedir. Bu durumun parametre tahminlerine nasıl yansıdığını gösteren sonuçlar Çizelge 4’te verilmiştir.

Çizelge 4. Williams metodu kullanılarak elde edilen lojistik regresyona ilişkin parametre tahminleri

Table 4. Parameter estimates for logistic regression using Williams method

| Parametreler/parameters         | Sd/df | Tahmin (Std. Hata)/estimation<br>(std. Error) |
|---------------------------------|-------|---|
| Intercept                       | 1     | -2.499(0.693)**                               |
| İlaçlama/insectifit application | 1     | 0.191(0.078)**                                |

sd serbestlik derecesi, df: Degrees of freedom, \*\*  $p<0.01$

Çizelge 4’te Williams metodu kullanılarak aşırı yayılım kontrol altına alındıktan sonra, lojistik regresyona ilişkin parametre tahmin değerleri verilmiştir. Çizelge 2’de olduğu gibi, Çizelge 4’te de hem intercept’in hem de ilaçların etkisi önemli bulunmuştur ( $p<0.01$ ). Çizelge 2 ve Çizelge 4’e bakıldığında parametre değerleri ve standart hatalarının farklılık gösterdiği görülmektedir. Özellikle ilaç etkisinin parametre tahmini Çizelge 4’te daha yüksektir. Avcı akarlar karşı denenen ilaçlar, akarlar üzerine oldukça etkili olmuş ve akarların ölüm oranlarını arttırdığı saptanmıştır.

Uygun model uyum ölçütleri Çizelge 5’te verilmiştir. Genel olarak en küçük AIC ve BIC değerine sahip model, en iyi model olarak tanımlanmaktadır. Model uyum ölçütlerine bakıldığında Williams metodu kullanılarak elde edilen AIC ve BIC uyum ölçütlerinin daha küçük olduğu saptanmıştır. Bu durumda Williams metodu kullanılarak yapılan lojistik regresyon analizinde, aşırı yayılımın ortadan kalktığı hem uyum istatistikleri sonuçlarında hem de model uyumu sonuçlarında görülmektedir.

Çizelge 5. Uygun model için uyum ölçütleri  
Table 5. Model selection criteria

| Regresyon yöntemi<br>Regression method   | -2Log L | AIC     | BIC     |
|--|---------|---------|---------|
| Lojistik regresyon/<br>Logistic regression   | 327.961 | 329.961 | 331.589 |
| Williams Metodu kullanılarak<br>lojistik regresyon<br>Logistic regression using Williams<br>method | 97.405  | 99.405  | 101.033 |

## Tartışma ve Sonuç

Kategorik veri kümelerine uygulanan Poisson ve lojistik gibi regresyon yöntemlerinin uygulanabilir olup olmadığını gösteren uyum istatistikleri (sapma ve Pearson Khi-kare) büyük önem taşımaktadır (Cox, 1992; Dean, 1992; Czado, 2003). Genel olarak uygulanan regresyon yönteminin doğruluğu bakımından, her iki uyum istatistiği değerlerinin 1’e eşit ya da çok yakın olması istenir. Poisson ve lojistik gibi regresyon analizlerinde, söz konusu uyum istatistikleri ortalama ile varyans arasındaki ilişkiye bağlı olarak değişmektedir (Cameron ve Trivedi, 1998; Lawles, 1987). Örneğin Poisson dağılımında ortalama ile varyans aynı parametre değerine eşittir. Uygulamada, bu eşitliğin sağlanamaması durumunda elde edilecek uyum istatistikleri sonuçları istenen değerden (1 değerinden) daha farklı çıkacaktır (Cox, 1983; Cameron ve Trivedi, 1998; Agresti, 1997; Stokes ve ark., 2000). Benzer olarak, 2. bölümde de bahsedildiği gibi, binomiyal dağılımda da ortalama ile varyans formüllerinde aynı terimler bulunmaktadır. Ancak lojistik regresyon aşırı yayılım, gözlenen varyansın beklenen varyanstan büyük çıkması durumunda gerçekleşmektedir. Aşırı yayılıma birçok neden gösterilebilir. Lojistik regresyonda, söz konusu nedenlerden en önemli olanları, başarı olasılıkları arasındaki varyasyon, ikili cevaplar arasındaki korelasyon, yanlış bağlantı

fonksiyonun kullanılması, modelde olması gereken önemli terimlerin olmaması ve örnek büyüklüğünün yetersiz olması şeklinde verilebilir (Lawles, 1987; Bonney, 1987; Czado, 2005; SAS, 2007).

Bu çalışmada, bitki koruma alanından alınan gerçek bir veri kümesi kullanılarak lojistik regresyonda meydana gelen aşırı yayılım incelenmiştir. İlk olarak veri kümesine lojistik regresyon uygulanmıştır. Daha sonra lojistik regresyon analizi sonucunda elde edilen uyum istatistik değerleri 1'den çok büyük çıktığı ve veri kümesinde aşırı yayılım olduğu saptanmıştır. Bu durumda, lojistik regresyon analizi sonuçlarının kullanılması sapmalı parametre tahminlerine neden olacaktır. Veri kümesindeki aşırı yayılımı (extra-binomiyal) modele dahil eden metodların kullanılması gerekmektedir. Bu bağlamda, Williams metodu kullanılarak aşırı yayılım kontrol altına alınmıştır (bkz., Çizelge 3). Williams metodu ile hem sapma hem de Pearson Khi-kare uyum istatistiklerinin aşırı yayılım değerleri 1 (bir) değerine oldukça yakın bulunmuştur. Çizelge 5'te verilen model uyumu ölçütleri, Williams metodu ile elde edilen uyum istatistikleri sonuçları birbirlerini desteklemektedir. En küçük AIC ve BIC model uyumu ölçütlerine sahip model en iyi model olarak kabul edilmektedir (Wang ve Putterman, 1998). Çizelge 5'e bakıldığında, Williams metodu kullanılarak elde edilen AIC ve BIC sonuçları daha küçük bulunmuştur. Bununla birlikte özellikle son yıllarda, veri kümelerindeki aşırı yayılımı gidermede karışımli model (Mixture model) yaklaşımı da yaygın olarak kullanılmaktadır (Wang ve Putterman, 1998).

Veri kümesinde aşırı yayılım olduğundan dolayı çizelge 2 ile Çizelge 4'te, lojistik regresyonu ile Williams metodu kullanılarak elde edilen lojistik regresyonundaki parametre tahmin değerleri birbirlerinden farklı bulunmuştur. Bu farklılık tamamen aşırı yayılımdan kaynaklanmaktadır. Örneğin Çizelge 2'de ilaçların etkisi 0.161 iken Williams metodu uygulandıktan sonraki ilaçların etkisi 0.191 olmuştur. Avcı akar *K. aberrans*'a karşı denenen, tarımsal savaş ilaçları, 24 saat sonra yapılan gözlemlerde avcı akarın canlılığı üzerinde oldukça etkili olmuş ve akarların çoğunluğunun öldüğü saptanmıştır. Bu ilaçların doğada zararlılara karşı kullanılırken yararlı türler üzerine de etkileri göz önünde bulundurularak rasgele kullanılmamalı ve mutlaka hedef alınan canlı grubuna etkili ilaçlar tercih edilerek zamanlaması göz önünde bulundurularak kullanılmalıdır (Kazak ve Şekeroğlu, 1996; Kılınçer ve ark., 1996; Kasap ve ark., 2004).

Sonuç olarak, elde edilen uyum istatistikleri, kullanılan regresyon modelinin uygun olup olmadığı bakımından oldukça önemlidir. Uyum istatistikleri sonucunda veri kümesinde yayılım olup olmadığı saptanmaktadır. Veri kümesinde aşırı yayılım olması durumunda, aşırı yayılımı dikkate alan yöntemlerin kullanılması daha doğru olacaktır.

## Kaynaklar

- Agresti A., 1997. Categorical Data Analysis. John and Wiley & Sons, Inc., 710.
- Bonney, G. E., 1987. Logistic Regression for dependent binary observations. *Biometrics*, 43(4): 951-973.
- Buis, M., L., 2005. Unobserved heterogeneity in logistic regression.
- Cameron, A. C., Trivedi, P. K. 1998. Regression Analysis of Count Data. Cambridge University Pres. 411.
- Cox, R., 1983. Some Remarks on Overdispersion. *Biometrika*, 70: 269-274.
- Czado, C., 2003. Overdispersion in Logistic Regression. [http://odin.stochastik.math.unigoettingen.de/zfs/past\\_courses/GLM-course/lec5.pdf](http://odin.stochastik.math.unigoettingen.de/zfs/past_courses/GLM-course/lec5.pdf) (Erişim tarihi: 13.04.2007)
- Dean, C. B., 1992. Testing for Overdispersion in Poisson and Binomial Regression Models. *Journal of American Statistical Association*, 87(418): 451-457.
- Kasap, İ., Çobanoğlu, S., Aktuğ, Y., Denizhan, E., 2004. Van Gölü çevresi elma bahçelerinde saptanan zararlı ve yararlı akar türleri. *Türkiye 1. Bitki Koruma Kongresi Bildirileri*, 8-10 Eylül Samsun, 104.
- Kazak, C., Şekeroğlu, E., 1996. Bazı tarımsal savaş ilaçlarının daldırma yöntemi ile avcı akar *Phytoseiulus persimilis* Athias Henriot (Acarina: Phytoseiidae)' e etkilerinin belirlenmesi. Türkiye 3. Entomoloji Kongresi Bildirileri, 24-28 Eylül Ankara, 639-647.
- Kılınçer, N., Çobanoğlu, S., Gürkan, M. O., 1990. Bazı pestisitlerin doğal düşmanlardan *Trichogramma turkeiensis* Kostadinov ve *Phytoseiulus persimilis* A. H.'e laboratuvar koşullarında yan etkileri. *Türkiye II. Biyolojik Mücadele Kongresi Bildirileri*, 26-29 Eylül Ankara, Entomoloji Derneği Yayınları, No: 4, 273-281.
- Lawles, J. F., 1987. Negative binomial and mixed poisson regression. *The Canadian Journal of Statistics*, 15(3): 209-225.
- McCullagh, P., Nelder, J. A., 1989. *Generalized Linear Models*. Second Edition, Chapman and Hall, London, 486.
- SAS, 2007. *SAS/STAT Software: Changen and Enhanced*. SAS, Inst. Carry Inc., MC USA
- Stokes, M. E., Davis, C. S., Koch, G. G., 2000. Categorical data analysis using the SAS system. John and Wiley & Sons, Inc., 626.
- Wang, P., Putterman, M. L., 1998. Mixed logistic regression models. *Journal of Agriculture, Biological and Environmental Statistics*, 3(2): 175-200.
- Zhang, B., 1999. A Chi-Squared goodness-of-fit- test for logistic regression models based on case-control data. *Biometrika*, (86): 531-539.