

POISSON REGRESYONU

Mehmet TAMAR

**Yüksek Lisans Tezi
İstatistik Anabilim Dalı
Danışman: Yrd. Doç. Dr. Nurhan HALİSDEMİR
AĞUSTOS-2013**

T.C
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

POISSON REGRESYONU

YÜKSEK LİSANS TEZİ

Mehmet TAMAR
(Enstitü No: .091133110.)

Tezin Enstitüye Verildiği Tarih:

Tezin Savunulduğu Tarih:

Tez Danışmanı: Yrd. Doç. Dr. Nurhan HALİSDEMİR (F.Ü.)

Diğer Jüri Üyeleri: Prof.Dr.Mehmet BEKTAŞ

Doç.Dr.Sinan ÇALIK

Yrd.DoçDr.Nurhan HALİSDEMİR

AĞUSTOS-2013

ÖNSÖZ

Bana araştırma olanağı sağlayan ve bu konuda çalışmaya yönlendiren, çalışmalarımın her aşamasında bana yakın ilgi ve yardımlarının esirgemeyen ve verdiği önerilerle beni yönlendiren değerli hocam sayın Yrd. Doç. Dr. Nurhan HALİSDEMİR'e ve yine çalışmalarım esnasında her türlü yardımı esirgemeyen Yrd. Doç. Dr. MehmetGÜRÇAN, Doç. Dr. Mahmut IŞIK ve Doç. Dr. Sinan ÇALIK'a en içten teşekkürlerimi sunarım.

Bulduğum süre içerisinde bana her türlü katkıyı sağlayan Fırat Üniversitesi Fen Fakültesi İstatistik bölümü yüksek lisan ve doktora arkadaşlarıma en içten teşekkürlerimi sunarım.

Özellikle her zaman yanımda olup, bana maddi ve manevi desteklerini esirgemeyen aileme sonsuz teşekkürlerimi sunarım.

MehmetTAMAR

ELAZIĞ-2013

İÇİNDEKİLER

ÖNSÖZ.....	I
İÇİNDEKİLER.....	II
ÖZET.....	III
SUMMARY.....	IV
TABLolar LİSTESİ.....	V
1. GİRİŞ.....	1
2. MATERYAL VE METOT.....	3
2.1 Regresyon Analizi.....	3
2.2 Çoklu Lineer Regresyon Modelinde Katsayıların Hesaplanması.....	4
2.3 En Küçük Kareler Yöntemi	6
2.4 En Çok Olabilirlik Yöntemi.....	7
2.5 Regresyon Katsayı Tahminlerinin Özellikleri	8
2.6 Genel Regresyon Süreci.....	11
3. BULGULAR.....	13
3.1 Doğrusal Olmayan Regresyon Modelleri	13
3.2 Link Fonksiyonları.....	13
3.3 Poisson Regresyonu	14
3.4 Uygulama.....	19
4. BULGULAR VE TARTIŞMA.....	22
EKLER	24
KAYNAKLAR	31
ÖZGEÇMİŞ	33

ÖZET

Çalışmamızda Poisson regresyon analizinin araştırılması ve bu analiz yardımıyla Aksaray ilinde meydana gelen trafik kazalarının belirleyici önemli nedenlerinin incelenmesi amaçlanmaktadır.

Tez çalışması temel olarak giriş, materyal ve metot, bulgular, sonuç ve tartışma kısımlarından oluşmaktadır. Giriş kısmında tez çalışması hakkında genel bilgiler verilerek uygulamada seçilen deneyin önemi vurgulanmıştır. Materyal ve metot kısmında tezde kullanılan teorik bilgiler tekrar edilmiş ve kullanılan bazı istatistiksel özellikler vurgulanmıştır. Bulgular kısmında Poisson regresyon analizinin uygulandığı ve incelenen verinin genel karakteristik yapısı anlatılarak gözlemler hakkında yeterli bilgi sunulmuştur.

Çalışmanın sonuç ve tartışma kısmında analiz edilen verilerin sonuçları açıklanmış, analizde önemli bulunan faktörler belirtilmiştir. İncelenen trafik kaza sayısı verisi üç farklı model yardımıyla açıklanmış ve bu modellerin AIC kriterlerine göre kıyaslaması yapılmıştır.

Anahtar kelimeler:Poisson regresyonu, Kaza sayısı, Link fonksiyonu, Lineer olmayan regresyon modeli.

SUMMARY

POISSON REGRESSION

In this study is aimed to investigate the causes of the accidents that occur the significant level with the help of Poisson regression analysis in Aksaray city, Turkey

As the basis of this thesis consist of introduction, materials and methods, findings, results and discussion parts. First of all, the introduction section is to try to emphasize the importance of providing general information about its application. Secondly, materials and methods given both the theoretical part of the thesis and some of the statistical properties. Thirdly the findings part is given application of Poisson regression analysis. It is provided sufficient information about the general characteristics of the structure of the data set. Lastly, some important factors in this study are given by results and discussion parts.

In conclusion, the numbers of traffic accident are analyzed with the help of three different models and these models were compared according to the criteria AIC.

Key words: Poisson regression, number of accident, link function, Nonlinear regression models.

TABLULAR LİSTESİ

Tablo 2.1: Bağımlı ve bağımsız değişkenlerin aldığı değerler.	4
Tablo 3.1: Regresyon katsayıları ve standart hataları.....	21
Tablo 4.1: Bayan sürücü sayısı ve buzlanma değişkeni dikkate alındığında elde edilen model katsayıları	22
Tablo 4.2: Birinci modelden ehliyet yılı çıkartıldıktan sonra elde edilen modelin katsayıları ve standart hataları	23

1. GİRİŞ

Tesadüfi deęişken kavramı modern bilimin ortaya koyabildięi en önemli kavramların hiç şüphe yok ki başında gelmektedir. Tesadüfi deęişken kavramı deęişkenler arasında var olan ilişkinin ifadesini vurgulamak için oldukça önemlidir. Matematiksel bir denklem birbiriyle bağlantılı deęişkenlerin bağıntı şeklini ifade etmekle sınırlıdır. Ancak ilişki kavramı tesadüfi deęişkenin ortaya konmasından sonra matematiksel bir ifade de anlamlı olmuş ve matematiksel bağıntıdan daha fazla bilgi içermiştir. Bunun bir sonucu olarak regresyon analizi önemli bir inceleme alanı olarak bilime kazandırılmıştır.

Regresyon incelemesine sadece deęişkenlerin arasında var olan bir matematiksel model gözüyle bakmak oldukça yetersiz olur. Bu incelemede verinin dağılımının bilinmesi bağımlı deęişkenin şartlı dağılımının ortaya konulabilmesi gibi analizi destekleyici bir sürü önemli unsur bulunmaktadır.

Bağımsız deęişkenlerin lineer kombinasyonları seçilebildięi gibi lineer olmayan çeşitli şekillerinin ortaya konması da mümkündür. Örneğin bağımsız deęişkendeki artım miktarı bağımlı deęişkenin artımını çok fazla etkileyebiliyorsa üstel bir fonksiyon veya lineerden daha az etkileyebiliyorsa logaritmik bir fonksiyon modele eklenebilir.

Bazı incelemelerde bağımlı ve bağımsız deęişkenlerin deęişim aralıklarını çeşitli fonksiyonlarla deęiştirmek ihtiyacı duyulabilir. Bu gibi durumlarda da bağımlı deęişken üzerinde bir dönüşüm yapmak mümkündür.

Poisson regresyon analizi bu bakımdan oldukça önemli bir inceleme alanıdır. Bağımlı deęişkenin sayı deęeri aldığı durumlarda kullanılır. Sıklıkla kullanılan bir regresyon türü olması gerçek hayatta uygulanabilir olmasından kaynaklanmaktadır. Dikkat edilecek olursa gerek sosyal hayatta gerekse ticari veya ekonomik hayatta incelenen olayların ilgilenilen deęişkenleri sıklıkla kesikli deęişkenler olarak karşımıza çıkmaktadır. Gözlemlenen birçok deneyde bağımlı deęişken tamsayı veya tane miktar sayısı olarak incelenmektedir. Bu ise bu regresyon türünün reel hayatta ne kadar kullanılabilir olduğunu vurgulamaktadır.

Poisson regresyon analizi kesikli deęer alan bağımlı deęişkenin incelenmesi şeklinde de deęerlendirilebilir. Bu bakımdan çoklu grup lojistik ayrımsama da olduğu gibi bağımlı deęişkenin katagorik deęer aldığı durumlarda karıştırılmamalıdır. Önemli olan bağımlı

değişkenin Poisson dağılımına sahip olabilmesidir. Bu konu hakkında oldukça detaylı bir bilgi Winkelmann, R. tarafından [17]'de bulunmaktadır. Bunun yanı sıra Tarone, R.E. kaynak [7]'de kategorik değişkenler için önemli olabilecek bir takım sonuçlar sunmaktadır.

Çalışmamızda Poisson regresyon analizi kullanılarak Aksaray ilinde meydana gelen trafik kazalarının bazı nedenleri incelemeye çalışılmıştır. Alınan örnekleme kaza yapan kadın sürücü sayısı, karayolunda meydana gelen buzlanma, yağış miktarı ve ehliyetli araç kullanım yılı kaza sayısı dikkate alınmıştır. Trafik kazalarında ölçülemeyen çevresel faktörlerin etkisi oldukça önemlidir. İncelemede hatayı azaltabilmek için birçok çevresel faktörü birlikte düşünmek gerekmektedir. Ancak bu gerçek gözlemlerde pek mümkün olmamaktadır. Bundan dolayı açıklayıcı değişken olarak ölçülebilen faktörlerin alınmasıyla inceleme yapılabilir.

2. MATERYAL VE METOT

2.1 Regresyon Analizi

İstatistiksel analizin önemli araştırma konularından biri olan regresyon analizi günümüzde birçok bilim dalının içerisinde yardımcı bir materyal olarak yer almaktadır. Regresyon analizi; uygulamalı istatistiksel analiz, sayısal analiz ve optimizasyon tekniklerinin ortaklaşa kullandıkları bir araç ve inceleme alanıdır.

Regresyon analizi bağımlı ve bağımsız değişkenler arasında var olan ilişkinin matematiksel modelini oluşturmak ve bu model üzerinden bağımlı ve bağımsız değişkenler hakkında istatistiksel yorum yapmayı amaçlayan bir inceleme dalıdır. Bağımlı ve bağımsız değişkenlere bazen sırasıyla açıklanan ve açıklayıcı değişkenler adı da verilmektedir. Yapılan bir deneyin sonucunda elde edilen sayısal değerler bağımlı değişkenin değerleri, deney yapılırken kullanılan sayısal değerler ise bağımsız değişken ya da değişkenlerin sayısal değerleri olarak alınmaktadır. Örneğin topraktan alınan ürünün verimi deneyin sonuç değişkeni olarak alınacak olursa yapılan deneyde bağımlı değişken ürün verimi olur. Aynı şekilde ürün verimini artırmak için toprağa atılan gübre miktarı, tohum cinsi, toprağa düşen yağış veya sulama miktarı gibi değişkenlerde deneyde bağımsız değişkenler olarak alınmalıdır.

Regresyon analizinde en önemli koşullardan biri seçilen bağımsız değişkenlerin kendi içlerinde birbirlerinden ilişkisiz olmalarıdır. Yapılan bir deneyde deneye etki eden faktörler tespit edilirken bu faktörlerin minimum sayıda olması incelemenin daha sağlıklı olması açısından önemlidir. Bununla birlikte seçilen bağımsız değişkenlerin deney sonucunu da yeteri miktarda açıklaması gerekmektedir. Her deneyde kontrol edilemeyen veya ölçülemeyen birtakım değişkenler de bulunabilir. Bu değişkenler regresyon modeline hata terimi olarak katılmaktadır. İyi bir regresyon modelinde açıklayıcı değişkenler bağımlı değişkeni büyük bir miktarda açıklayabilmeli ve hata terimine düşen açıklama oranı düşük olmalıdır.

Genel olarak regresyon analizinde bağımsız değişkenler, X_1, X_2, \dots, X_p şeklinde gösterilmektedir. Regresyonda bağımlı değişken Y olmak üzere buradaki p tane açıklayıcı değişkenin lineer bir fonksiyonu şekilde ifade edilecek olursa,

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

matematiksel denklemi kurulmuş olur. Ancak en ideal şekilde eşitliğin sağlanabilmesi için denkleme kontrol edilemeyen veya ölçülemeyen fakat deneyde etkisi olan bir takım değişkenlerin değerlerini ifade eden bir hata teriminin de eklenmesi sonucunda model denklemi aşağıdaki şekilde yazılır,

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + e$$

Burada e ile modelin hata terimi gösterilmektedir. Ayrıca β_0, \dots, β_p regresyon katsayıları olarak adlandırılmakta olup reel sabitlerdir. Bu şekilde oluşturulan model çoklu lineer regresyon modelidir.

2.2 Çoklu Lineer Regresyon Modelinde Katsayıların Hesaplanması

Çoklu lineer regresyon modelinin kurulabilmesi için regresyon katsayılarının hesabının yapılması gerekmektedir. Bu hesabın yapılabilmesi deneyden elde edilen sayısal değerlerin kullanılmasına bağlıdır. Bu durumda bağımlı ve her bir bağımsız değişkenin değerlerini aşağıdaki tablo ile gösterecek olursak,

Tablo 2.1: Bağımlı ve bağımsız değişkenlerin aldığı değerler.

	Y	X_1	X_2	\dots	X_p
1-inci gözlem değeri:	y_1	x_{11}	x_{21}	\dots	x_{p1}
2-inci gözlem değeri:	y_2	x_{12}	x_{22}	\dots	x_{p2}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n -inci gözlem değeri:	y_n	x_{1n}	x_{2n}	\dots	x_{pn}

her bir gözlem için çoklu lineer regresyonun aldığı değerler aşağıdaki şekilde olacaktır,

$$y_1 = \beta_0 + \beta_1 x_{11} + \dots + \beta_p x_{p1}$$

$$y_2 = \beta_0 + \beta_1 x_{12} + \dots + \beta_p x_{p2}$$

\vdots

$$y_n = \beta_0 + \beta_1 x_{1n} + \dots + \beta_p x_{pn}$$

Yukarıdaki bu denklem sistemi matris formunda aşağıdaki şekilde ifade edilebilmektedir,

$$Y = X\beta$$

Burada Y bağımlı değişken değerlerinden oluşan $n \times 1$ boyutlu vektör, X ilk sütunu birlerden oluşan ve diğer sütunları tablo 1'deki bağımsız değişken değerlerinden oluşan $n \times (p + 1)$ boyutlu veri matrisi ve β ise regresyon katsayılarından oluşan $(p + 1)$ boyutlu katsayılar vektörüdür.

Bu eşitlikten katsayılar vektörü elde edilmek istendiğinde eşitliğin her iki tarafına veri matrisinin tersi soldan çarpılmalıdır. Ancak veri matrisi tam ranklı kare bir matris olmadığından tersi alınamaz. Bu problem eşitliğin her iki tarafını veri matrisinin transpozu ile çarpma suretiyle giderilebilir. Eşitliğin her iki tarafı X^t ile çarpıldığında eşitlik bozulmayacak ve eşitliğin sağ tarafında oluşan X^tX matrisi artık kare bir matris olduğundan tersi de rahatlıkla alınabilecektir. Bu durumda

$$X^tY = X^tX\beta$$

denklem sistemi çözüldüğünde katsayılar matrisi,

$$\beta = (X^tX)^{-1}X^tY$$

şeklinde elde edilir.

Model denklemi e hata terimini ihtiva edecek şekilde yazılmak istendiğinde ϵ hata terimlerinden oluşan hata vektörü olmak üzere aşağıdaki formda yazılmaktadır,

$$Y = X\beta + \epsilon$$

Model denkleminde ϵ hata vektörü değişkenler cinsinden yazılmak istendiğinde aşağıdaki şekilde elde edilir,

$$\epsilon = Y - X\beta$$

İdeal bir arařtırmada hata teriminin minimum olması istenir. Bu bakımdan hata vektörü üzerinde yapılan bir optimizasyon işlemi yardımıyla hangi katsayılar vektörü için hata vektörünün minimum olması gerektiđi hesaplanır. Bu sayede bađımlı ve bađımsız deđişken deđerleri belli iken hata vektörünü minimum yapan regresyon katsayılar vektörü hesaplanmış olur. Bu hesaplama yöntemi en küçük kareler yöntemi olarak adlandırılmaktadır.

2.3 En Küçük Kareler Yöntemi

Bu yöntemde minimum yapılmak istenen fonksiyon hata kareler toplamı fonksiyonudur. Hata terimleri pozitif ya da negatif işaretli olabileceklerinden hataların toplamı gerçek hatayı yansıtmamaktadır. Bundan dolayı negatif terimli hataları ortadan kaldırmak için hata deđerlerinin kareleri alınmaktadır. Hata karelerinin toplamı hatanın miktarını daha gerçekçi olarak ifade eder. Dolayısıyla hata kareler toplamı fonksiyonunu minimum yapmak hata terimlerinin toplamda minimum olmasını sağlayacaktır. Hata kareler toplamı fonksiyonunu S ile gösterelim,

$$S = \epsilon^t \epsilon$$

Hata kareler toplamı fonksiyonundan regresyon katsayılarına göre türevler alınarak sıfıra eşitlendiđinde elde edilen denklem sistemi regresyon denkleminin normal denklemleri olarak adlandırılmaktadır. türevin sıfıra eşitlenmesi klasik optimizasyon yöntemi olmakla birlikte ikinci türevler daima pozitif olacađından çözüm minimum probleminin çözümü olacaktır. Bu durumda ařađıdaki eşitliđi dikkate alarak yazabiliriz,

$$S = \sum_{j=1}^n e_j^2$$
$$S = \sum_{j=1}^n (y_j - (\beta_0 + \beta_1 x_{1j} + \dots + \beta_p x_{pj}))^2$$

Bu eşitlikten regresyon katsayılarına göre her bir $j = 0, 1, \dots, n$ için β_j regresyon katsayısına göre türev alacak olursak ařađıdaki denklem sistemini elde ederiz,

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{j=1}^n (y_j - (\beta_0 + \beta_1 x_{1j} + \dots + \beta_p x_{pj})) = 0$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_{j=1}^n (y_j - (\beta_0 + \beta_1 x_{1j} + \dots + \beta_p x_{pj})) x_{1j} = 0$$

⋮

$$\frac{\partial S}{\partial \beta_p} = -2 \sum_{j=1}^n (y_j - (\beta_0 + \beta_1 x_{1j} + \dots + \beta_p x_{pj})) x_{pj} = 0$$

Yukarıdaki denklem sisteminin çözümünden regresyon katsayıları elde edilecektir.

2.4 En Çok Olabilirlik Yöntemi

Bu yöntemde E-hata vektörünü oluşturan her bir bileşenin sıfır ortalamalı normal dağılıma sahip olduğu varsayılmaktadır. Normal dağılımın yoğunluk fonksiyonu maksimum değerini ortalamada aldığından yoğunluk fonksiyonunu maksimum yapan değer bu durumda hatayı minimum yapmaktadır.

E-hata vektörünü oluşturan her bir bileşen sıfır ortalamalı normal dağılımlı varsayıldığından bileşenlerin birbirlerinden bağımsız olmasından dolayı yoğunluk fonksiyonlarının çarpımı ortak yoğunluk fonksiyonu olmaktadır. Bu durumda elde edilen ortak yoğunluk fonksiyonuna olabilirlik fonksiyonu denilmektedir,

$$\mathcal{L} = \prod_{j=1}^n f_j(e_j)$$

Burada $f_j(e_j)$, $N(0, \sigma^2)$ normal dağılımın yoğunluk fonksiyonu olup aşağıdaki şekilde yazılmaktadır,

$$f_j(e_j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}(e_j/\sigma)^2\right\}$$

Olabilirlik fonksiyonu birbirinden bağımsız hata terimlerinin yoğunluk fonksiyonlarının çarpımından oluştuğundan regresyon katsayılarına göre ve türevlerinin alınması ve sıfıra eşitlenerek elde edilen denklem sisteminin çözülmesi oldukça zordur. Bu problemi ortadan kaldırabilmek için olabilirlik fonksiyonunun logaritması alınmaktadır. Logaritma fonksiyonu bire bir ve artan bir fonksiyon olduğundan olabilirlik fonksiyonun maksimum değerini aldığı değişken değeri ile logaritması alınmış log olabilirlik fonksiyonunun maksimum değerini aldığı değişken değeri aynı olacaktır. Bu özellikten dolayı incelemede olabilirlik fonksiyonu yerine aşağıdaki log olabilirlik fonksiyonu kullanılmaktadır,

$$\ln \mathcal{L} = \sum_{j=1}^n \ln f_j(e_j)$$

Log olabilirlik fonksiyonunun regresyon katsayılarına göre türevleri alınarak sırayla eşitlenmek suretiyle elde edilen denklem sistemi çözüldüğünde regresyon katsayıları elde edilebilmektedir. Bu çözüm en küçük kareler yöntemiyle elde edilen çözümle aynı çıkmaktadır.

2.5 Regresyon Katsayı Tahminlerinin Özellikleri

- a) Regresyon katsayılarının tahmin vektörü $\hat{\beta}$ olmak üzere $\hat{\beta}$ tahmin edicisi β için yansız bir tahmin edicidir,

$$\begin{aligned} E\hat{\beta} &= E((X^t X)^{-1} X^t Y) \\ &= (X^t X)^{-1} X^t EY \\ &= (X^t X)^{-1} X^t E(X\beta + e) \\ &= (X^t X)^{-1} X^t \{EX\beta + Ee\} \\ &= (X^t X)^{-1} X^t \{X\beta\} \\ &= (X^t X)^{-1} (X^t X)\beta \\ &= \beta \end{aligned}$$

b) $\hat{\beta}$ parametre tahmin vektörünün kovaryansı aşağıdaki şekilde hesaplanır,

$$\begin{aligned}
Cov(\hat{\beta}) &= E\{(\hat{\beta} - E\hat{\beta})(\hat{\beta} - E\hat{\beta})^t\} \\
&= E\{(\hat{\beta} - \beta)(\hat{\beta} - \beta)^t\} \\
&= E\{((X^tX)^{-1}X^t(X\beta + e) - \beta)((X^tX)^{-1}X^t(X\beta + e) - \beta)^t\} \\
&= E\{(\beta + (X^tX)^{-1}X^te - \beta)(\beta + (X^tX)^{-1}X^te - \beta)^t\} \\
&= E\{((X^tX)^{-1}X^te)((X^tX)^{-1}X^te)^t\} \\
&= E\{(X^tX)^{-1}X^te e^t X(X^tX)^{-1}\} \\
&= (X^tX)^{-1}X^t E\{e e^t\} X(X^tX)^{-1} \\
&= (X^tX)^{-1}X^t [Var(e) + E(e)E(e)^t] X(X^tX)^{-1} \\
&= (X^tX)^{-1}X^t [\sigma^2 I + 0] X(X^tX)^{-1} \\
&= (X^tX)^{-1} \sigma^2 I
\end{aligned}$$

c) Kabul edelim ki $Var(e) = \sigma^2 V$ olsun. Bu durumda $E(e) = 0$ varsayımı altında hata terimi varyansı $\sigma^2 I$ olacak şekilde bir dönüşüm yapmak mümkündür. Burada V bilinmeyen bir matris olup,

$$V = P^t P, P \text{ simetrik}$$

şeklinde çarpanlarına ayrılabilir. Buna göre hata terimine,

$$U = P^{-1}e$$

Dönüşümü uygulanırsa yeni bir hata terimi elde edilmiş olur,

$$CovU = E(U - EU)(U - EU)^t$$

Burada,

$$EU = E(P^{-1}e) = P^{-1}E(e) = 0$$

olduğu dikkate alınacak olursa aşağıdaki eşitlik yazılabilir,

$$\begin{aligned}CovU &= E(U - 0)(U - 0)^t \\ &= E(P^t e e^t (P^{-1})^t) \\ &= E(P^{-1} e e^t (P^t)^{-1})\end{aligned}$$

d) Regresyon denkleminin her iki tarafı P^{-1} matrisi ile çarpılacak olursa aşağıdaki denklem elde edilir,

$$P^{-1}Y = P^{-1}X\beta + P^{-1}e$$

$$P^{-1}Y = P^{-1}X\beta + U$$

Elde edilen bu denklem için $EU = 0$ ve $VarU = \sigma^2 I$ varsayımları sağlanmaktadır. Dolayısıyla regresyon katsayıları en küçük kareler metodu ile tahmin edilebilir. Denklemden aşağıdaki dönüşümler yapılsın,

$$Z = P^{-1}Y$$

$$Q = P^{-1}X$$

Bu durumda dönüşüm sonucunda elde edilen regresyon denklemi aşağıdaki şekilde olur,

$$Z = Q\beta + U$$

Bu modelde regresyon katsayılarının en küçük kareler tahmin edicisi aşağıdaki şekilde bulunacaktır,

$$\begin{aligned}\hat{\beta} &= (Q^t Q)^{-1} Q^t Z \\ &= (X^t (P^{-1})^t P^{-1} X)^{-1} X^t (P^{-1})^t P^{-1} Y \\ &= (X^t (P^t)^{-1} P^{-1} X)^{-1} X^t (P^t)^{-1} P^{-1} Y \\ &= (X^t P^{-1} P^{-1} X)^{-1} X^t (P^{-1} P^{-1}) Y \\ &= (X^t (PP)^{-1} X)^{-1} X^t (PP)^{-1} Y \\ &= (X^t V^{-1} X)^{-1} X^t V^{-1} Y\end{aligned}$$

Elde edilen bu tahmin ediciye genelleştirilmiş en küçük kareler tahmin edicisi denir.

2.6 Genel Regresyon Süreci

Regresyon herhangi bir inceleme konusu üzerinde bağımlı değişkenleri bağımsız değişkenler yardımıyla matematiksel modeller kurarak inceleyen bilim dalıdır. Kurulacak modeller bağımlı değişkenleri bağımsız değişkenlerin bir fonksiyonu olarak ifade eder. Matematiksel modeller kesinlik ifade etmesine karşılık, bir matematiksel model gözüyle bakabileceğimiz regresyon modelleri kesinlikten daha ziyade tesadüfîlik kurallarına bağlıdır. Regresyon modelleri aşağıdaki genel formül ile gösterilebilir,

$$y = f(\beta, x) + e$$

Burada hata terimine tesadüfî bir değişken gözüyle bakılmaktadır.

Bir regresyon modelinin kullanım amaçları şu şekilde verilebilir,

- a) Salt Tanımlama: bağımlı değişkenin iyi tanımlanması istendiğinde olabildiğince çok ve etkin bağımsız değişkenin modele alınması.
- b) Katsayı Tahmini: regresyon incelemesinde araştırmacının amacı uygun modelini tanımladıktan sonra model katsayılarının tahmin edilmesidir. Bu durumda tahmin edicinin yansızlık ve minimum varyanslı olma gibi iki önemli özelliği sağlaması gerekmektedir.
- c) Denetim: bu amaçla bağımsız değişkenlere ilişkin girdi düzeylerinin değiştirilerek çıktı düzeylerinin kontrolü istenmektedir.
- d) Uyum Tahmini: veri kümesindeki bir gözlem için bağımlı değişkenin ortalama değerinin bulunmasıdır.
- e) Ön Tahmin: ön tahmin için kullanılan modellere ön tahmin modelleri denilmektedir. Bu modeller bağımlı değişken değerlerinin önceden tahmin edilmesini amaçlamaktadır. Dolayısıyla böyle modellerde parametre tahminleri de ön tahmin amacıyla yapılır. Ön tahmin modelleri iki farklı amaç için yapılmaktadır. Bunlardan ilki iç değer bulma ikincisi ise dış değer bulmadır. İç değer bulma veri kümesi içerisindeki bağımsız değişken değerleri yardımıyla bağımlı değişken değerlerini tahmin etmektir. Dış değer bulma ise verilerin bulunduğu bölgenin dışındaki bir gözlem için ön tahminde bulunmaktır.

3. BULGULAR

3.1 Doğrusal Olmayan Regresyon Modelleri

Bağımlı ve bağımsız değişkenler arasındaki ilişkinin denkleminin doğrusal olmadığı durumda yapılan regresyon analizine lineer olmayan regresyon analizi veya kısaca eğrisel regresyon analizi denilmektedir. En çok kullanılan eğrisel regresyon modelleri polinomsal, üstel ve logaritmik regresyon modelleridir. Bunlarla birlikte sıkça olmasa da trigonometrik regresyon modelleri de bazı deneylerde kullanılabilir.

Eğrisel bir regresyon modelinde regresyon katsayılarının tahminleri için en küçük kareler yöntemi kullanılırken hata kareler toplamından katsayılara göre türevler alınarak oluşturulan normal denklemleri genelde çözülememektedir. Bundan dolayı eğrisel model hata kareler toplamına yerleştirilirken eğrisel modelin sadece iki terimli Taylor açılımı kullanılmaktadır. Bu sayede eğrisel model birinci dereceli doğrusal bir denklemlerle temsil edilebilir ve en küçük kareler yönteminin çözümü bulunabilmektedir. Taylor açılımı yapılırken başta eğrisel model tahmini katsayı değerleri civarında seriye açılarak işlem yapılır. En küçük kareler çözümü bulunduktan sonra işlemler bir kere daha tekrar edilirken Taylor açılımı elde edilen bu çözüm civarında yapılmalıdır. Çözümler birbirine yaklaştıkça optimal regresyon katsayı değerlerine ulaşılmaktadır.

3.2 Link Fonksiyonları

Bazı araştırmalarda bağımlı ve bağımsız değişken değerlerinin cinsleri ve değişim aralıkları birbirlerine uymamaktadırlar. Bu gibi durumlarda bağımlı ve bağımsız değişken değerlerinin değişim aralıklarını birbirlerine eşitleyebilmek için bağımlı değişken değeri üzerinde dönüşüm yapmak gerekmektedir. Bu gibi dönüşümlere link fonksiyonları ismi verilmektedir.

En çok kullanılan link fonksiyonlarının başında lojistik link fonksiyonları gelmektedir. Lojistik link fonksiyonları bağımlı değişkenin sıfır ve bir gibi kategorik değerler aldığı durumlarda kullanılmaktadır. Lojistik link fonksiyonu bağımlı değişken değeri Y olmak üzere aşağıdaki şekilde kullanılır,

$$\ln \frac{Y}{1-Y} = f(X, \beta)$$

Lojistik regresyon modellerinde genellikle bağımsız değişkenlerin kombinasyonu lineer olarak tercih edilmektedir. Regresyon katsayılarının tahminleri en küçük karelerde lineer yaklaşım yöntemi kullanılarak ardışık olarak yapılabilmektedir.

Lojistik regresyonda bağımlı değişken 0 ve 1 gibi iki kategori olabildiği gibi 0, 1, ..., n gibi ikiden fazla kategori de olabilmektedir. Bu durumda link fonksiyonu tüm ikili karşılaştırmalar üzerinden hesaplanır.

Lojistik link fonksiyonlarının yanı sıra önemli bir başka link fonksiyonu da Gompertz modelinin link fonksiyonudur,

$$\ln(\ln Y) = f(X, \beta)$$

Gompertz modeli çifte üstel model olarak da adlandırılmaktadır.

İncelenmek istenen veri yapısına bağlı olarak link fonksiyonları bu şekilde farklılık gösterebilmektedir. Bağımlı değişkenin değişim aralığına ve cinsine bağlı olarak farklılık gösteren link fonksiyonları bağımlı değişken değerlerine bir manada olasılık anlamı katabilmektedir. Olasılık dağılım fonksiyonları reel düzlemin $y = 0$ ve $y = 1$ doğruları ile sınırlı şeridinde sürekli ve artan bir şekil oluşturduğundan bu tipli modellere genellikle büyüme modelleri de denilmektedir.

3.3 Poisson Regresyonu

Araştırılan deneyde bağımlı değişken değerleri 0, 1, ... gibi tamsayı değerlerini alıyorsa incelenen model poisson regresyon modeli olarak adlandırılmaktadır.

Herhangi bir olayın belirlenen bir süreç içerisinde yapılan denemeler sonucunda meydana gelme sayısı, sayma verileri olarak ifade edilebilir. Sayma veri modelleri özel bir regresyon türüdür. Bilindiği gibi, verilerin sürekli olduğu durumlarda doğrusal regresyon analizi kullanılabilir. Ancak analizlerde kullanılacak veriler her zaman sürekli halde bulunmayabilir. Bu gibi durumlarda yani, verilerin kesikli olması durumunda da doğrusal regresyon modelleri kullanılarak yapılacak analizler etkisiz, tutarsız ve çelişkili sonuçlar

verecektir. Bu sebepten dolayı kesikli veriler için tüm koşullar sağlandığında kullanılabilen en etkin modellerden biri dePoisson regresyon modelleridir.

Poisson regresyon lojistik regresyondan sonra en genel olan ikinci genelleştirilmiş doğrusal modeldir. Bağımlı değişken oluş sayısı ile belirtilen bir veri olduğunda yani belirli bir zaman ya da yerde olan olayların sayısı olduğunda Poisson regresyon kullanılmaktadır.

Poisson regresyon analizi, açıklayıcı değişkenler ile sayıya dayalı olarak elde edilen cevap değişkeni arasındaki ilişkiyi açıklamaktadır. Poisson regresyonunda açıklayıcı değişkenlerin doğrusal yapısını cevap değişkeninin beklenen değerine bağlayan bağlantı fonksiyonu, logaritmik dönüşüm ile verilmektedir.

Bağımlı değişkenin 0, 1, 2, 3, ... gibi kesikli değer aldığı fakat kategorik olmadığı durumlar vardır. Bu tür değişkenlere, doğalgaz boruları üzerinde kazaların sayısı, verilen patentlerin sayısı, yazlıklarda çıkan yangınların sayısı gibi örnekler gösterilebilir. Kesikli ve kategorik olmayan, nadir olaylarla ilişkili bağımlı değişkenli model, bazı varsayımlar altında Poisson regresyon modeli olarak adlandırılır. Poisson regresyon modeli daha çok sayma verilerini analiz etmek için kullanılmaktadır.

Sayma verilerinin analizi için ilk sorulan soru “özel” yöntemlerin gerekliliği veya doğrusal regresyon modelinin yeterli olup olmadığıdır. Sayma verilerinden oluşan değişkenler için sürekli ve doğrusal regresyon modelinin uygulanabileceği düşünülür. Ancak bu verilere doğrusal regresyon modeli uygulanması halinde sonuçlar, etkisiz ve tutarsız olduğu gibi çelişkili tahminleri yapılabilir. Sayma sonuçlarının özelliklerini kesin olarak veren birçok model vardır. Ancak Poisson regresyon birçok analizin başlangıç noktası olarak düşünülür. Poisson regresyon modeli sayma verileri için en sık kullanılan ve en basit olan yöntemdir. Bu model ile sayımın olasılığı, Poisson dağılımı ile belirlenir. Bu modelin belirgin özelliği, sonucun koşullu ortalamasının koşullu varyansına eşit olmasıdır. Ancak uygulamada bazen koşullu varyans, koşullu ortalama değerini asabilir. İşte bu tür durumlarda, negatif binom regresyon modelleri kullanılır

Poisson regresyon modelinde regresyon sürecindeki genel kestirimler en çok olabilirlik yöntemi ile gerçekleştirilmektedir.

Θ parametrelili bir kitleden tesadüfi olarak seçilen n birimlik örnek birimleri, x_1, x_2, \dots, x_n olsun. Bu şekilde Θ için,

$$\mathcal{L}(\Theta) = f(x_1, x_2, \dots, x_n, \Theta)$$

şeklinde tanımlanan fonksiyona olabilirlik fonksiyonu denir.

Buna göre en çok olabilirlik metodu, olabilirlik fonksiyonunu maksimum yapan Θ değerlerinin bulunmasından ibarettir. Bunun için,

i) $\mathcal{L}(\Theta)$ olabilirlik fonksiyonu yazılmalıdır.

ii) $\ln\mathcal{L}(\Theta)$ oluşturulmalıdır.

iii) $\frac{\partial \ln\mathcal{L}(\Theta)}{\partial \Theta} = 0$ denklemi çözülmelidir.

Poisson en çok olabilirlik kestirimi için,

- 1) Koşullu ortalamanın doğru tanımlanmasında bağımlılık şartı sağlanmalıdır. Ayrıca bağımlı değişken y 'nin Poisson dağılması gereklidir.
- 2) En çok olabilirlik standart hataları ve t istatistikleri kullanarak hesaplanan istatistiksel sonuçlar, hem koşullu ortalama, hem varyansın doğru tanımlanmasını gerektirmektedir. Burada istenen koşul, koşullu varyans ve ortalamanın eşit olmasıdır.
- 3) Veriler için koşullu varyans ve koşullu ortalamanın eşit olmaması durumunda, en çok olabilirlik yönteminin uygulanması ile elde edilmiş istatistiksel sonuçlar, koşullu ortalamanın doğru tanımlandığının ispat edildiği durumlarda geçerli ve doğrudur.
- 4) Veriler için koşullu varyans ve ortalamanın eşit olmaması durumunda, Poisson en çok olabilirlik tahmin edicisinden daha etkin tahmin ediciler kullanılabilir.

Bağımlı değişken y_i 'i sayılabilir olsun. Bu durumda gözlemler $y_i = 0,1,2, \dots$, olmak üzere sayılabilir veri seti için en uygun olasılık modeli sıklıkla poisson modelidir. Poisson dağılımı,

$$f(y_i) = \frac{e^{-\mu} \mu^{y_i}}{y_i!}, \quad y_i = 0,1,2, \dots,$$

parametresi $\mu_i > 0$ olmaktadır. Poisson dağılımının ortalaması ve varyansı μ parametresine eşittir.

$$E(y) = \mu \text{ ve } Var(y) = \mu$$

Poisson regresyonda lineer belirleyici olarak bağımlı değişkenin ortalamasıyla ilgili bir g fonksiyonun var olduğu kabul edilirse, Bu durumda

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k = x_i' \beta$$

g fonksiyonu link fonksiyonu olarak adlandırılır. Ortalama ve lineer belirleyici arasındaki ilişki ise aşağıdaki gibi tanımlanır,

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(x_i' \beta)$$

Poisson dağılımıyla ilgili yaygın olarak kullanılan bir kaç tane link fonksiyonu vardır. Bunlardan biri özdeş link fonksiyonudur,

$$g(\mu_i) = \mu_i = x_i' \beta$$

Bu link fonksiyonu kullanıldığı zaman $E(y_i) = \mu_i = x_i'\beta$ ve $\mu_i = g^{-1}(x_i'\beta)$. Poisson dağılımı için bir başka popüler link fonksiyonu ise log link fonksiyonudur. Log link fonksiyonu için lineer belirleyici ve bağımlı değişkenin ortalaması arasındaki ilişki,

$$\mu_i = g^{-1}(x_i'\beta) = e^{x_i'\beta}$$

Log-link fonksiyonu poisson regresyonda, bağımlı değişken için tahmin edilen değerler negatif olmadığı durumlarda daha çok kullanılır.

Poisson regresyonda parametre tahmini için en çok olabilirlik yöntemi kullanılır. Bağımlı değişken y ve bağımsız değişkenler x olmak üzere n tane tesadüfi gözleme sahip olunursa, Poisson regresyonda olabilirlik fonksiyonu aşağıdaki gibi yazılır.

$$\mathcal{L}(\beta; y) = \prod_{i=1}^n f_i(y_i) = \prod_{i=1}^n \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} = \frac{\prod_{i=1}^n \mu_i^{y_i} \exp(-\sum_{i=1}^n \mu_i)}{\prod_{i=1}^n y_i!}$$

$$\mu_i = g^{-1}(x_i'\beta)$$

Link fonksiyonu seçileceği zaman, log-olabilirlik maksimum yapılır.

$$\ln \mathcal{L}(\beta; y) = \sum_{i=1}^n y_i \ln(\mu_i) - \sum_{i=1}^n \mu_i - \sum_{i=1}^n \ln(y_i!)$$

Poisson regresyonda parametrelerin tahminlerini en çok olabilirlik yöntemiyle bulmak için iterativ yöntemler kullanılabilir. β parametresinin tahmini elde edildiği zaman, uygun Poisson regresyon modeli aşağıdaki gibi olur.

$$\hat{y}_i = g^{-1}(x_i'\beta)$$

Özdeş link fonksiyonu kullanıldığı zaman tahmin denklemi

$$\hat{y}_i = g^{-1}(x_i'\beta) = x_i'\beta$$

olur.

Log link fonksiyonu seçilirse, o zaman tahmin denklemi aşağıdaki gibi olmaktadır.

$$\hat{y}_i = g^{-1}(x_i'\beta) = \exp(x_i'\beta)$$

Bununla birlikte model ve parametreleri üzerinde istatistiksel çıkarımlar yapmak için, full model ile indirgenmiş model arasındaki sapmalar kullanılarak modellerin performansları testlerle değerlendirilebilir. Bu testler Likelihood ve Wald test istatistiklerine bakılabilir, aynı zamanda modelin sapmasına, uyum iliği test sonuçlarına bakılarak en uygun modelin seçimi sağlanabilir.

3.4 Uygulama

Çalışmamızda Poisson regresyon analizine uygun olarak bağımlı değişken değerlerinin tamsayı değeri aldığı bir trafik kaza sayısı üzerinde inceleme yapacağız. Trafik kazalarının incelendiği bölge olarak Aksaray ili seçilmiştir.

Modern ulaşım ve taşıma sistemlerinin negatif sonuçları, karayollarında meydana gelen trafik kazaları ve yaralanmalar olmaktadır. Son yıllarda Dünya Sağlık Örgütü (DSÖ) ve Dünya Bankası ortak olarak hazırlanan Trafik kazalarına ilişkin Dünya raporunda da bu durum açıkça belirtilmiştir. Karayolu trafik kazaları ortalama olarak her yıl 1,2 milyon kişinin ya da her gün 3242 kişinin ölümüne yol açmaktadır. Karayolu trafik kazaları her yıl 20 – 50 milyon kişinin yaralanmasına ya da sakat kalmasına yol açmaktadır. Karayolu trafik kazaları

tek başına tüm dünyada meydana gelen ölümlerin %2,1'ine yol açmakta ve ölüm nedenleri arasında 11. sırada gelmektedir. Karayolu trafik sistemini daha az tehlikeli hale getirmek, sistemin bir bütün olarak ele almakla mümkündür. Sistemi oluşturan öğeler; araç, yol, yol kullanıcısı ve bunların içinde bulunduğu fiziksel, sosyal ve ekonomik çevre arasındaki etkileşimin anlaşılmasını ve müdahale edilecek yerlerin belirlenmesini gerektirir. Bu nedenle karayolu trafik sistemi ve güvenliği ulaşım ve taşıma stratejilerinde önemli bir yer tutmaktadır.

Bu çalışmada Aksaray Emniyet Müdürlüğü Trafik Şube Müdürlüğünden sayısal verilerek temin edilerek Poisson regresyon modelini 2011-2013 döneminde Aksaray'da yapılan trafik kaza sayılarına uygulamaya çalıştık. Trafik kaza sayılarının belirleyicileri olarak, cinsiyet, hava şartları, sürücünün ehliyet kullanma yılı seçilmiştir. Bu sayede Poisson regresyon yöntemiyle trafik kaza sayılarının tespiti amaçlanmaktadır.

Veriler Aksaray ilinde belirlenen belli kavşaklarda 2011-2013 yıllarının aralık, ocak ve şubat ayları dikkate alınarak toplanmıştır. Bağımsız değişken olarak cinsiyet; kaza yapan kadın sürücü sayısı olarak alınmıştır. Buna ilaveten bağımsız değişken olarak karayolundaki buzlanma durumu, havanın yağış miktarı ve sürücünün ehliyetli olarak araba kullanma yılı alınmıştır. Bu durumda ilgili bağımsız değişkenlerimiz aşağıdaki şekilde seçilmiştir,

X_1 : Kaza yapan kadın sürücü sayısı

X_2 : Karayolunda buzlanma, var(1), yok(0)

X_3 : Yağış miktarı, yağış yok(0), az yağış(1), yoğun yağış(2), sağnak yağış(3)

X_4 : Kaza yapan sürücülerin ehliyetli araba kullanım yılı ortalaması

Elde edilen 120 adet veri Ek 1'deki tabloda verilmiştir.

Uygulamada incelenecek olan Poisson regresyon verileri Poisson olasılıkları üzerinde $-\ln$ link fonksiyonu kullanılarak bağımsız değişkenlerin kombinasyonuna eşitlenecektir. Bu sayede bağımlı değişken değerleri pozitif değerlere eşitlenmiş olacaktır,

$$-Y \ln \mu + \ln(Y!) + \mu = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

Poisson regresyon analizinde Poisson olasılıkları oldukça küçük değerler olarak elde edilmektedir. Bu nedenle link fonksiyonları bu değerleri bağımsız değişkenlerin değer aralığına dönüştürürken değer aralıklarını anlamlı bir şekilde farklılamamaktadır. Dolayısıyla dönüşen değerlerin çoğu bir değer etrafında yığılmaktadır. Bu durum dönüşüm sonucunda elde edilen bağımlı değişken değerlerinin bağımsız değişkenler tarafından sağlıklı bir şekilde bir regresyon modeli yardımıyla tahmin edilebilmesini imkânsız hale getirmektedir. Bu problemin ortadan kaldırılabilmesi için bağımsız değişkenlerin regresyon modeli kurulurken hassas bir modelin tercih edilmesi önemlidir.

Elde edilen veriler R programı uygun kodları kullanılarak incelendi. Analiz sonucunda regresyon katsayıları Tablo 1'deki gibi elde edilmiştir.

Tablo 3.1: Regresyon katsayıları ve standart hataları

AIC:400	Katsayı Tahminleri	Katsayıların Standart Hataları
Regresyon Sabiti (β_0):	-0.153678	0.170311
Bayan Sürücü Sayısının Katsayısı (β_1):	0.299915	0.029052
Buzlanma Katsayısı (β_2):	0.432069	0.170587
Yağış Miktarının Katsayısı (β_3):	0.092923	0.053705
Ehliyet Kullanım Yılıının Katsayısı (β_4):	0.001414	0.008607

Regresyon katsayılarından cinsiyet ve buzlanma değişkenlerinin katsayıları anlamlı diğer yağış miktarı ve ehliyet kullanım yılı değişkenlerinin katsayıları ile regresyon sabiti anlamsız olarak elde edilmiştir. Bu sonuca göre, Aksaray ilinde kış aylarında meydana gelen trafik kaza sayıları üzerinde yolda buzlanmanın olup olmaması ve sürücünün bayan olmasının önemli etkisi bulunmaktadır.

Ayrıca oluşturulan model için AIC bilgi kriteri 400 olarak hesaplanmıştır. Bu kriter modelden yağış miktarı ve ehliyet kullanım yılı çıkartılarak hesaplama yapıldığında 399'a düşmektedir. Buna karşılık modelden sadece ehliyet kullanım yılı değişkeni çıkartıldığında ise 398'e düşmektedir. Bu durumda AIC kriterleri arasında kıyaslama yapıldığında düşük kritere sahip olan model daha anlamlı olmaktadır.

4. BULGULAR VE TARTIŞMA

Çalışmada incelenen Poisson regresyon verilerinde Aksaray ilinde meydana gelen trafik kazaları incelenmeye çalışılmıştır. 2011-2013 yılları arasında aralık, ocak ve şubat kış aylarında meydana gelen trafik kazaları dikkate alınmış ve bu kazaların olmasında etkili olan faktörler incelenmeye çalışılmıştır. Alınan kaza örneklerinden elde edilen veriler incelendiğinde kaza sayısına etki eden önemli faktörlerin sürücünün cinsiyeti, karayolundaki buzlanma durumu, yağış miktarı ve ehliyet kullanım yılı olarak belirlenmesine karar verilmiştir.

Bu bağımsız değişkenler yardımıyla inceleme yapıldığında Poisson regresyon modelinde kaza sayısına etki eden önemli iki faktörün sürücünün bayan olup olmaması ve karayolundaki buzlanma durumu olduğu ortaya çıkmıştır. Bunun yanısıra yağış miktarının ve ehliyetli araç kullanım yılının kaza sayısına daha az etkili iki faktör olduğu saptanmıştır.

Oluşturulan modelden yağış miktarı ve ehliyetli araç kullanım yılı faktörleri çıkartıldığında oluşturulan yeni modelin katsayıları ve standart hataları aşağıdaki tabloda olduğu gibidir,

Tablo 4.1: Bayan sürücü sayısı ve buzlanma değişkeni dikkate alındığında elde edilen model katsayıları

AIC:399	Katsayı Tahminleri	Katsayıların Standart Hataları
Regresyon Sabiti (β_0):	-0.05310	0.14189
Bayan Sürücü Sayısının Katsayısı (β_1):	0.32286	0.02485
Buzlanma Değişkeninin Katsayısı (β_2):	0.40309	0.16352

Bu modelde de regresyon sabiti anlamlı bulunmamıştır. AIC kriteri daha düşük olduğundan model birinci modelden daha uygundur.

Yine oluşturulan birinci modelden sadece ehliyetli araba kullanım yılının çıkarılmış durumunu incelediğimizde oluşturulan yeni modelin regresyon katsayıları tahmini ve tahminlerin standart hataları aşağıdaki tabloda olduğu gibidir,

Tablo 4.2: Birinci modelden ehliyet yılı çıkartıldıktan sonra elde edilen modelin katsayıları ve standart hataları

AIC:398	Katsayı Tahminleri	Katsayıların Standart Hataları
Regresyon Sabiti (β_0):	-0.14091	0.15137
Bayan Sürücü Sayısının Katsayısı (β_1):	0.29882	0.02826
Buzlanma değişkeninin Katsayısı (β_2):	0.43917	0.16498
Yağış Miktarının Katsayısı (β_3):	0.09060	0.05180

Oluşturulan bu üçüncü modelde AIC kriteri 398 olarak bulunmuş olup elde edilen üç model içerisinde AIC kriteri en düşük olan modeldir. Buna göre bu model kaza sayısını açıklamakta diğer ikisine göre daha anlamlıdır. Bu modelde bayan sürücü sayısı değişkeni ile karayolundaki buzlanma durumu anlamlı çıkarken bunlarla birlikte yağış miktarı da anlamlı bir değişken olarak elde edilebilmiştir.

Oluşturulan bu üç modelde de modelde bulunan regresyon sabiti anlamsız olarak elde edilmiştir. Genellikle Poisson regresyonu modellerinde sabit olarak bağımlı değişkenin ortalama değeri bulunduğundan bir daha regresyon sabiti modele eklendiğinde anlamsız olmaktadır. Ancak bu regresyon sabitinin modelden çıkartılabilmesi anlamına gelmez. Modelde elde edilen regresyon sabiti bağımlı değişkenin ortalama değerinin civarında meydana gelen değişimi vurgulamaktadır. İncelediğimiz trafik kazası verisinde de 2011-2013 yıllarının aralık, ocak ve şubat aylarında meydana gelen ortalama trafik kazası sayısı 2.9 olarak hesaplanmıştır.

EKLER

EK 1. Veriler

Tablo 1: 2011-2013 yılları aralık, ocak ve şubat aylarına ait veriler.

Sıra No:	Y: Kaza sayısı	X_1	X_2	X_3	X_4
1	1	1	1	0	8
2	1	1	0	1	11
3	3	2	0	1	6
4	0	0	0	1	0
5	0	0	0	1	0
6	0	0	0	1	0
7	0	0	0	0	0
8	0	0	0	0	0
9	6	2	1	0	7
10	7	5	1	0	8
11	1	0	1	0	8
12	5	4	1	0	7
13	5	4	1	0	7
14	5	3	1	0	7
15	5	4	1	0	6
16	9	5	1	3	6
17	9	5	1	3	6
18	9	7	1	3	6
19	9	5	1	3	3
20	10	6	1	3	3
21	10	7	1	3	3
22	8	7	1	2	3
23	9	7	1	3	3
24	8	3	1	2	3
25	10	2	1	3	3
26	0	0	0	2	0
27	9	6	1	2	3

Sıra No:	Y: Kaza sayısı	X_1	X_2	X_3	X_4
28	9	4	1	2	3
29	7	3	0	2	3
30	7	3	0	2	3
31	5	4	0	1	6
32	0	0	0	1	0
33	0	0	0	1	0
34	0	0	0	1	0
35	0	0	0	1	0
36	2	2	0	1	6
37	2	0	1	0	6
38	4	3	1	0	21
39	1	0	0	0	22
40	1	1	0	0	17
41	1	1	0	0	15
42	1	1	0	0	13
43	1	1	0	0	11
44	1	0	0	0	1
45	1	0	1	0	11
46	1	1	1	0	14
47	2	0	1	0	14
48	2	0	1	0	15
49	3	0	1	1	16
50	2	0	1	1	12
51	1	1	1	1	12
52	1	1	1	1	10
53	1	1	1	1	9
54	2	2	1	1	9
55	3	1	1	1	9
56	4	3	0	1	5
57	1	0	0	1	5
58	1	1	0	1	8

Sıra No:	Y: Kaza sayısı	X_1	X_2	X_3	X_4
59	0	0	0	3	0
60	0	0	0	3	0
61	0	0	0	3	0
62	0	0	0	3	0
63	1	1	0	3	5
64	1	1	0	3	9
65	1	1	0	3	9
66	1	1	0	3	10
67	3	2	0	3	10
68	4	2	1	3	9
69	5	3	1	3	9
70	6	2	1	3	9
71	6	3	1	3	9
72	6	2	1	3	10
73	5	3	1	3	11
74	6	4	1	2	11
75	6	3	1	2	11
76	6	4	1	2	12
77	7	5	1	2	10
78	6	4	1	0	10
79	6	4	1	0	25
80	6	5	1	0	22
81	1	1	1	1	35
82	1	1	1	1	32
83	1	1	1	0	44
84	1	0	1	0	23
85	1	0	1	0	29
86	1	0	1	1	2
87	2	2	1	1	2
88	2	2	1	1	5
89	2	2	1	1	32

Sıra No:	Y: Kaza sayısı	X_1	X_2	X_3	X_4
90	1	0	1	1	21
91	1	1	1	1	21
92	1	1	1	1	21
93	1	1	1	1	18
94	1	1	1	1	15
95	1	0	1	0	14
96	1	0	1	0	14
97	1	1	0	0	14
98	1	0	0	0	8
99	1	0	0	0	8
100	1	0	0	0	8
101	2	0	0	0	8
102	2	1	0	0	8
103	2	2	1	0	8
104	6	4	1	0	8
105	5	5	1	0	3
106	4	3	1	1	5
107	2	1	1	1	5
108	2	0	1	1	5
109	2	2	1	1	5
110	2	0	1	1	5
111	1	0	1	1	5
112	1	0	1	1	5
113	1	0	1	1	9
114	0	0	1	1	0
115	0	0	1	1	0
116	0	0	1	1	0
117	1	1	1	1	7
118	1	1	1	1	7
119	1	1	1	1	4
120	3	3	1	1	6

EK 2: R Kodları

```
summary(m1 <- glm(kaza ~ buz+sex+yagmur+yil, family =  
"poisson", data = data1))
```

Call:

```
glm(formula = kaza ~ buz + sex + yagmur + yil, family =  
"poisson",  
     data = data1)
```

DevianceResiduals:

Min	1Q	Median	3Q	Max
-1.7027	-0.7690	-0.2637	0.4007	3.0356

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept)	-0.153678	0.170311	-0.902	0.3669
buz	0.432069	0.170587	2.533	0.0113 *
sex	0.299915	0.029052	10.323	<2e-16 ***
yagmur	0.092923	0.053705	1.730	0.0836 .
yil	0.001414	0.008607	0.164	0.8695

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersionparameterforpoissonfamilytakento be 1)

Nulldeviance: 332.30 on 119 degrees of freedom

Residualdeviance: 102.93 on 115 degrees of freedom

AIC: 400.8

Numberof FisherScoringiterations: 5

```
>summary(m1 <- glm(kaza ~ buz+sex+yagmur, family = "poisson",  
data = data1))
```

Call:

```
glm(formula = kaza ~ buz + sex + yagmur, family = "poisson",  
     data = data1)
```

DevianceResiduals:

Min	1Q	Median	3Q	Max
-1.7177	-0.7764	-0.2592	0.3919	3.0211

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.14091	0.15137	-0.931	0.35189
buz	0.43917	0.16498	2.662	0.00777 **
sex	0.29882	0.02826	10.574	< 2e-16 ***
yagmur	0.09060	0.05180	1.749	0.08031 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersionparameterforpoissonfamilytakento be 1)

Nulldeviance: 332.30 on 119 degrees of freedom

Residualdeviance: 102.96 on 116 degrees of freedom

AIC: 398.83

Numberof FisherScoringiterations: 5

```
>summary(m1 <- glm(kaza ~ buz+sex, family = "poisson", data =  
data1))
```

Call:

```
glm(formula = kaza ~ buz + sex, family = "poisson", data =  
data1)
```

DevianceResiduals:

Min	1Q	Median	3Q	Max
-----	----	--------	----	-----

-1.6847 -0.7576 -0.2825 0.4591 3.3986

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -0.05310 0.14189 -0.374 0.7082

buz 0.40309 0.16352 2.465 0.0137 *

sex 0.32286 0.02485 12.994 <2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 332.30 on 119 degrees of freedom

Residual deviance: 106.03 on 117 degrees of freedom

AIC: 399.9

Number of Fisher Scoring iterations: 5

KAYNAKLAR

- [1] **Abramowitz, M. and Stegun, I.A. (Eds.)** (1964) “Handbook of Mathematical Functions, Applied Mathematics Series”, Vol. 55. National Bureau of Standards, US Department of Commerce, Washington, DC.
- [2] **Barlow, R.E. and Proschan, F.** (1975) “Statistical Theory of Reliability and Life Testing: Probability Models”. New York: Rinehart and Winston
- [3] **Bailey, R. and L. Simon** (1960). Twostudies in automobileinsuranceratemaking. ASTIN Bulletin V.1, N.4, 192{217.
- [4] **Cameron, A. and P. Trivedi** (1998). Regressionanalysis of count data. Cambridge UniversityPress.
- [5] **Cramer, H.** (1946) “Mathematical Methods of Statistics” *Princeton University Press:* New Jersey
- [6] **Olkin I., and Liu R.,** 2002 “A Bivariate Beta Distribution” Department of Statistics, Stanford University, *statistic & probability Letters*, Elsevier
- [7] **Tarone, R.E.** (1979) “Testing the goodness of fit of the binomial distribution”, *Biometrika*
- [8] **Tong, Y.L.,** (1980) “Probability Inequalities in Multivariate Distributions”. *Academic Press*, New York.
- [9] **Aytaç, M.,** (2004) “Matematiksel İstatistik” Ezgi Kitabevi, Bursa
- [10] **Johnson, N.L., Kotz, S. and Balakrishnan, N.** (1995) “Continuous Univariate Distributions” volume 2. New York: *John Wiley and Sons.*
- [11] **Parker, C. S.** (1988) “Thegeneralized beta as a model forthedistribution of earnings”, *economicsletters*, Elsevier
- [12] **Krishnamoorthy, K.** (2006) “Handbook of Statistical DistributionswithApplications”, University of Louisiana at Lafayette, Taylor & Francis Group, LLC, U.S.A.

- [13] **Kotz, S. and Dorp, J.R.** (2004) “Beyond Beta Other Continuous Families of distributions with Bounded Support and Applications”, World Scientific, The George Washington University, USA
- [14] **Shahbazov, A.** (2005) “Olasılık Teorisine Giriş”, Birsen Yayınevi, İstanbul
- [15] **Evans, M., Hastings, N. and Peacock, B.** (1993) “Statistical Distribution Second Edition” A Wiley-Interscience Publication *JOHN WILEY & SONS, INC.*
- [16] **Winkelmann, R.** (1997). Econometric analysis of count data (Second, revised and enlarged ed.). Berlin: Springer-Verlag.

ÖZGEÇMİŞ

1998 yılı Elazığ İli Maden İlçesi doğumluyum. İlköğrenimime 1993 yılında Gazi Osman Paşa İlköğretim Okulunda başlayıp 2001 yılında tamamladım.2001 yılında Korgeneral Hulusi Sayın Lisesi'nde orta öğretim hayatıma başladım ve 2004 yılında mezun oldum.2005 yılında Fırat Üniversitesi İstatistik Bölümü'nde Lisans öğrenimime başladım ve 2009 yılında 76,75 ortalama ile mezun oldum.2009 yılı Eylül ayında Tezli Yüksek Lisans ile lisansüstü eğitimime başladım.2010 yılında Adile Sadullah Mermerci Polis Meslek Eğitim Merkezi'nde eğitim hakkı kazanarak 2011 yılından beri polis olarak meslek hayatıma devam etmekteyim.