

**T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**



**LOJİSTİK VE POİSSON REGRESYON MODELLERİNDE PARAMETRE
TAHMİNİ VE MODEL SEÇİMİ**

MEHMET MEŞE

YÜKSEK LİSANS TEZİ

**T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

YÜKSEK LİSANS TEZİ

**LOJİSTİK VE POISSON REGRESYON MODELLERİNDE PARAMETRE
TAHMINİ VE MODEL SEÇİMİ**

MEHMET MEŞE

İSTATİSTİK ANABİLİM DALI

**SAMSUN
2019**

Her hakkı saklıdır.

TEZ ONAYI

Mehmet MEŞE tarafından hazırlanan "Lojistik ve Poisson Regresyon Modellerinde Parametre Tahmini ve Model Seçimi" adlı tez çalışması 17/07/2019 tarihinde aşağıdaki jüri tarafından Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı'nda **Yüksek Lisans Tezi** olarak kabul edilmiştir.

Danışman Doç. Dr. Erol Terzi
İstatistik Anabilim Dalı

Jüri Üyeleri

Başkan Doç. Dr. Kamil ALAKUŞ
Ondokuz Mayıs Üniversitesi
İstatistik Anabilim Dalı



Üye Doç. Dr. Erol TERZİ
Ondokuz Mayıs Üniversitesi
İstatistik Anabilim Dalı



Üye Doç. Dr. Muhammet BEKÇİ
Sivas Cumhuriyet Üniversitesi
İstatistik ve Bilgisayar Bilimleri Bölümü



Yukarıdaki sonucu onaylarım 17 /07/2019

Prof. Dr. Bahtiyar Öztürk

Enstitü Müdürü

ETİK BEYAN

Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez içindeki bütün bilgilerin doğru ve tam olduğunu, bilgilerin üretilmesi aşamasında bilimsel etiğe uygun davrandığımı, yararlandığım bütün kaynakları atıf yaparak belirttiğimi beyan ederim.

17/07/2019

MEHMET MEŞE



ÖZET

Yüksek Lisans Tezi

LOJİSTİK VE POİSSON REGRESYON MODELLERİNDE PARAMETRE TAHMİNİ VE MODEL SEÇİMİ

Mehmet MEŞE

Ondokuz Mayıs Üniversitesi

Fen Bilimleri Enstitüsü

İstatistik Anabilim Dalı

Danışman: Doç. Dr. Erol Terzi

Araştırması yapılan konu hakkında elde edilen veriler her zaman sayısal veri olmayabilir. Bu durumda karşımıza kategorik veriler gelmektedir. Kategorik veri analizleri için en çok kullanılan analiz yöntemlerinden biri lojistik regresyon analizidir. Lojistik regresyon analizinde; doğrusal regresyon analizi için oluşturulmuş olan varsayımlar olmadığı için daha çok tercih edilen istatistik türlerinden biridir. Poisson regresyon analizinde bağımlı değişkenler kesikli değerler aldığı için günlük hayatımızda karşımıza çıkan olayların analizinde kullanılmaktadır. Poisson regresyon analizinde araştırmacının amacı elde edilen açıklayıcı değişkenlerin kesikli (sayıya dayalı) cevap değişkeni ile arasındaki ilişkiyi göstermektir. Bu çalışmada, Lojistik regresyon ve Poisson regresyon analizi hakkında teorik bilgiler verilmiştir. Lojistik regresyon ve Poisson regresyon analizi için literatür taraması yapılmıştır. Lojistik regresyon ve Poisson regresyon katsayılarının nasıl hesaplandığı anlatılmıştır. Katsayılar kullanılarak tahminsel model bulunmuştur. Daha verimli bir model bulabilmek için olabilirlik oranı testi ve Wald testi uygulanmıştır. Uygulanan testler sonrası güven aralıkları oluşturulmuştur ve modelin anlamlı olup olmadığı yorumlanmıştır. Ayrıca adımsal analiz yöntemiyle de en sade model bulunmaya çalışılmıştır. Ortaya çıkan modelin elde edilen verilere uyup uymadığına Pearson Ki-Kare yöntemiyle bakılmıştır. Sınıflandırma tablosu oluşturulmuştur. Oluşturulan veri setleri SPSS programı kullanılarak Lojistik regresyon ve Poisson regresyon analizine göre uygulanmıştır.

Temmuz 2019, 43 sayfa

Anahtar Kelimeler: Lojistik Regresyon, Poisson Regresyon, Model Seçimi,
Parametre Tahmini, En Çok Olabilirlik Oranı, Ki-kare Testi

ABSTRACT

Master's Thesis

PARAMETER ESTIMATION AND MODEL SELECTION IN LOGISTIC AND POISSON REGRESSION MODELS

Mehmet MEŞE

Ondokuz Mayıs University
Graduate School of Sciences
Department of Statistic

Supervisor: Assoc. Prof. Dr. Erol TERZİ

The data obtained from the research subject may not always be numerical data. In this case, we come across categorical data. One of the most used analysis methods for categorical data analysis is logistic regression analysis. In logistic regression analysis; it is one of the more preferred types of statistics since there are no assumptions for linear regression analysis. Since the dependent variables take discrete values in the Poisson regression analysis, they are used in the analysis of the events encountered in our daily lives. The aim of the researcher in Poisson regression analysis is to show the relationship between the obtained explanatory variables and the discrete (census-based) variable. In this study, theoretical information about Logistic regression and Poisson regression analysis is given. A literature review was performed for logistic regression and Poisson regression analysis. Logistic regression and Poisson regression coefficients are explained. Predictive model was found by using coefficients. Likelihood ratio test and Wald test were applied to find a more efficient model. Confidence intervals were established after the tests and whether the model was significant or not. In addition, it has been tried to find the simplest model by step analysis method. The Pearson Chi-Square method was used to determine whether the resulting model fit the data obtained. Classification table is created. The data sets were applied according to Logistic regression and Poisson regression analysis using SPSS program.

July 2019, 43 pages

Key Words: Logistic Regression, Poisson Regression, Model Selection, Parameter Estimation, Maximum Likelihood Ratio, Chi-Square Test

ÖNSÖZ VE TEŞEKKÜR

Akademik eğitim sürecimin bir üst noktası olan yüksek lisans tez çalışmam süresince yardım ve desteğini benden esirgemeyen danışman hocam Sayın Doç.Dr. Erol TERZİ'ye teşekkürü bir borç bilirim.

Beni bugünlere getirmek için hiçbir fedakârlıktan kaçınmayan annem, babam ve amcam Necati Meşe'ye emeklerinden dolayı sonsuz teşekkür ederim.

Ondokuz Mayıs Üniversitesi Arş. Gör. Dr. Mehmet Kuru'ya tez dönemimde yapmış olduğu görüş ve katkılarından dolayı teşekkür ederim.

Yıllardır aynı evi paylaştığım dost bildiğim Mustafa Kundak'a tez döneminde göstermiş olduğu hoşgörü ve sabırlarından dolayı teşekkür ederim.

Yakın zamanda tanıdığım çok değerli Ayhan Demirel'e tez döneminde yapmış olduğu yardım ve desteklerinden ötürü çok teşekkür ederim.

Son olarak bu tez çalışmam boyunca emeği geçen tüm hocalarıma ve arkadaşlarıma teşekkür ederim.

Temmuz 2019, Samsun
Mehmet MEŞE

İÇİNDEKİLER DİZİNİ

ÖZET.....	i
ABSTRACT.....	ii
ÖNSÖZ VE TEŞEKKÜR.....	iii
İÇİNDEKİLER DİZİNİ.....	iv
SİMGELER.....	vi
ÇİZELGELER DİZİNİ.....	vii
1. GİRİŞ.....	1
2. LOJİSTİK REGRESYON.....	3
2.1. Lojistik Regresyon Analizi.....	3
2.1.1. Lojistik regresyon tanımı ve tercih nedenleri.....	3
2.1.2. Basit Lojistik regresyon modeli.....	4
2.1.3. Basit Lojistik regresyon tahminsel modeli.....	5
2.1.4. Lojistik regresyon modeli için önemlilik testleri.....	6
Olabilirlik oranı testi.....	6
Wald test istatistiği.....	9
Score testi.....	9
2.1.5. Tahmin edicilerin güven aralığı.....	10
2.2. Çoklu Lojistik Regresyon.....	11
2.2.1. Çoklu Lojistik regresyon modeli.....	11
2.2.2. Çoklu Lojistik regresyon tahminsel modeli.....	12
2.2.3. Çoklu Lojistik regresyon modeli için önemlilik testleri.....	14
Olabilirlik oranı testi.....	14
Wald testi.....	14
Score testi.....	14
2.2.4. Çoklu Lojistik regresyonda tahmin edicilerin güven aralığı.....	15
2.2.5. Model geliştirme yöntemleri.....	15
Tam regresyon modeli.....	15
İleri doğru değişken seçim yöntemi.....	16
Geriye doğru değişken seçim yöntemi.....	16
Adımsal değişken seçim yöntemi.....	16
2.2.6. Model uygunluğunun değerlendirilmesi.....	17
Standart olmayan hatalar.....	17
Standart hatalar.....	17
Sapma değerleri.....	17
Gözlem uzaklık değerleri.....	18
Cook uzaklığı.....	18
Lojistik regresyon analizinde ilişki ölçüsü.....	18
2.2.7. Uyum iyiliği.....	19
2.3. Lojistik Regresyon Analizi Kullanılarak Yapılan Çalışmalar: Literatür Taraması.....	19
3. POISSON REGRESYON.....	24
3.1. Poisson Regresyon Analizi.....	24
3.2. Poisson Regresyon Modeli.....	24
3.3. Poisson Regresyon Analizinde Katsayıların Tahmini.....	25
3.4. Poisson Regresyonda Uyum İyiliği Testi.....	25
3.4.1. Pearson İstatistiği.....	25
3.4.2. Yapay R^2 İstatistiği.....	26
3.4.3. Sapma İstatistiği.....	27

3.4.4. Akaike bilgi ölçütü.....	27
3.4.5. Bayes bilgi ölçütü	27
3.5. Poisson Regresyon Analizi Kullanılarak Yapılan Çalışmalar: Literatür Taraması.....	27
4. BULGULAR ve TARTIŞMA.....	30
4.1. Lojistik Regresyon Analizi Uygulaması	30
4.2. Poisson Regresyon Analizi Uygulaması	36
5. SONUÇ ve ÖNERİLER.....	38
KAYNAKLAR	40
ÖZGEÇMİŞ	43



SİMGELER

x_i	i inci gözlem için bağımsız değişken
Y_i	i inci gözlem için bağımlı değişken
β_0	Sabit
β_p	p inci bağımsız değişkenin katsayısı
$\hat{\beta}_p$	p inci bağımsız değişkenin katsayısının tahmin edicisi
Cov	Kovaryans
D	Sapma
$E(Y_i)$	i inci gözlem için bağımlı değişkenin beklenen değeri
$E(Y \setminus x)$	x' e bağlı olarak Y' nin beklenen değeri
e_i	i inci gözlemin hata değeri
G	Olabilirlik oranı testi değeri
OR	Odds Oranı
$\widehat{OR}$	Odds Oranının tahmin edicisi
$SE(\hat{\beta}_p)$	$\hat{\beta}_p$ ' nin standart hatası
$\widehat{SE}(\hat{\beta}_p)$	$\hat{\beta}_p$ ' nin standart hatasının tahmin edicisi
ST	Skor testi değeri
$\bar{Y}$	Y_i ' lerin ortalaması
$Var(e_i)$	i inci gözlemin hata değerinin varyansı
$\pi(x_i)$	x_i ' ye karşılık gelen değer
$\hat{\pi}(x_i)$	x_i ' ye karşılık gelen değer tahmin edicisi
$g(x_i)$	x_i ' ye karşılık gelen lojik dönüşüm değeri
$\hat{g}(x_i)$	x_i ' ye karşılık gelen lojik dönüşüm değerinin tahmin edicisi
$L(\beta_0, \beta_1)$	Olabilirlik fonksiyonu
$\ln L(\beta_0, \beta_1)$	Log olabilirlik fonksiyonu
W	Wald testi değeri

ÇİZELGELER DİZİNİ

Çizelge 4.1. Öğrencilerin sınav puanı ve not ortalamasını gösteren tanımlayıcı istatistikler	30
Çizelge 4.2. Öğrencilerin mezun oldukları kurumlara göre dağılımı	30
Çizelge 4.3. Lisansüstü eğitime kabul durumunun kurumlara göre dağılımı	31
Çizelge 4.4. Başlangıçta gözlenen duruma göre öngörülerek oluşturulmuş sınıflandırma tablosu	31
Çizelge 4.5. Eşitlikte bulunan değişkenlerin durumu	32
Çizelge 4.6. Eşitlikte bulunmayan değişkenlerin durumu	32
Çizelge 4.7. Kurulan ilk model ve son modelin incelenmesi.....	32
Çizelge 4.8. Model özeti	33
Çizelge 4.9. Enter yöntemi kullanılarak oluşturulan modelin bulunduğu sınıflandırma tablosu	33
Çizelge 4.10. Enter yöntemi ile eşitlikteki değişkenlerin analizi.....	35
Çizelge 4.11. Öğrencilerin matematik sınav puanı ve not ortalamasını gösteren tanımlayıcı istatistikleri	36
Çizelge 4.12. Oluşturuların modelin uyum iyiliği test sonuçları	36
Çizelge 4.13. Omnibus test sonuçları.....	37
Çizelge 4.14. Değişkenlerin modele etkisi.....	37
Çizelge 4.15. Modele alınan değişkenler için parametre tahmini	37

1. GİRİŞ

İstatistik analizlerinin temel amacı herhangi bir konu hakkında birtakım araştırma sonucu ortaya çıkan verileri elde ederek bu veriler ışığında konu hakkında genel bir bilgiye ulaşmaktır. Sosyal bilimler, ekonomi, biyoloji ve tıp gibi birçok alanda araştırması yapılan olayları çözümlemede kullanılabilecek değişkenlerin sayısı fazla olabilir. Bu olayların çözümünde kullanılacak olan değişkenler ve bu değişkenlerin özellikleri araştırmacının kullanmak istediği yöntemin seçiminde çok önemlidir. Kullanılmak istenen yöntemlerden model oluşturabilmek için değişkenler tanımlanır, bu değişkenlerin tek başına modele katkısı ve diğer değişkenler ile arasındaki uyum incelenmek istenebilir. Değişkenler arasındaki ilişkilerin daha iyi yorumlanabilmesi için birtakım ihtiyaçlar ortaya çıkmıştır. Bu ihtiyaçlar sayesinde bazı istatistiksel yöntemler ortaya çıkmıştır. Bağımlı değişkenin, bir veya birden fazla bağımsız değişken arasındaki ilişkinin incelenmesi regresyon analizi ile mümkündür.

Bir konu hakkında araştırma yapan kişilerin sahip oldukları veriler her zaman sayısal bir veri olmayabilir. Bu durumlarda veriler iki seçenekli, seviye olarak artış gösteren ya da eşit düzeyli seçeneklerden oluşabilir. Bu tür veriler karşımıza kategorik veri olarak çıkar. Kategorik verilerin analizinde birçok istatistiksel yöntem kullanılmıştır. Lojistik regresyon bu istatistiksel yöntemlerden sadece biridir. Lojistik regresyon modelinde doğrusal regresyon analizi için aranan verilerin normal dağılması, verilerin sürekli dağılması vs. gibi varsayımlar olmadığı için son yıllarda daha çok kullanılan istatistiksel bir yöntem olmuştur. Doğrusal regresyon analizinde bağımlı değişken sürekli bir dağılıma sahip iken; lojistik regresyon analizinde bağımlı değişken kesikli bir dağılıma sahiptir.

Poisson regresyon bağımlı değişkenin 0, 1, 2, ... gibi kesikli değer aldığı fakat kategorik olmadığı durumlarda kullanılabilir. Daha çok günlük hayatımızda karşımıza çıkan olaylar üzerinde kullanabileceğimiz, kesikli ve kategorik olmayan, nadir olaylarla ilgili bağımlı değişkenli modeller, bazı varsayımlar altında Poisson regresyon modeli olarak adlandırılır. Bu regresyon modeli daha çok sayma verilerini analiz etmek için kullanılmaktadır. Araştırması yapılan birçok deneyde bağımlı değişken tamsayı veya miktar sayısı olarak incelenir. Çok terimli (multinomial) lojistik regresyon analizinde olduğu gibi bağımlı değişkenin kategorik bir değer

aldığı durumlar ile sıklıkla karıştırılmaktadır. Burada en önemli unsur bu bağımlı değişkenin Poisson dağılımına sahip olmasıdır. Lojistik regresyondan sonra en genel doğrusal model Poisson regresyon modelidir. Poisson regresyon analizinde amaç açıklayıcı değişkenler ile sayıma dayalı elde edilen cevap değişkeni ile arasındaki ilişki göstermektir. Burada açıklayıcı değişkenlerin doğrusal yapısını cevap değişkenin beklenen değerine bağlayan bağlantı fonksiyonu, logaritmik dönüşüm ile verilmiştir (Tamar, 2013).

Yapılan bu tez çalışmasında Lojistik regresyon ve Poisson regresyon analizi detaylı bir şekilde ele alınmış, Lojistik regresyon ve Poisson regresyon için literatür taraması yapılmıştır. Lojistik regresyon ve Poisson regresyon analizi hakkında teorik olarak konu anlatımı yapılmıştır. Yapılan araştırma sonu elde edilen veriler kullanılarak değişkenler oluşturulmuştur. Lojistik regresyon analizi ve Poisson regresyon analizi için uygun parametreler tahmin edilmiştir. Modele uygun değişkenler modelde kalırken modele uygun olmayan değişkenler modelden çıkarılmıştır. Parametrelerin güven aralıkları oluşturularak uygun modeller bulunmuştur. Yapılan bu çalışmada parametre tahminleri ve değişkenlerin uyumunun hesaplanmasında SPSS istatistiksel paket programından yararlanılmıştır.

2. LOJİSTİK REGRESYON

2.1. Lojistik Regresyon Analizi

2.1.1. Lojistik regresyon tanımı ve tercih nedenleri

Regresyon analizi bir tane bağımlı değişkenin bağımsız olan değişken veya değişkenler ile arasındaki ilişkinin durumunu açıklayan bir analiz türüdür. Lojistik regresyon analizi de bu analiz türlerinden biri olarak karşımıza çıkar. Lojistik regresyon analizi modellerinde bulunan değişkenler kesikli dağılıma sahip olduğundan dolayı bağımlı değişkenin Bernoulli veya Binom dağılıma sahip olduğunu belirtir. Burada ortaya çıkan beklenen değer ve varyansların birbirine eşit olmadığıdır. Buradan hareketle ortaya çıkan gözlemlerin birbirinden farklı dağılımlara sahip olduğudur. Lojistik regresyon analizinde bağımlı değişken kesikli bir dağılıma sahip olduğu için sıfır (0) ve bir (1) değerini almaktadır. Alınan bu değerler ile oluşturulan lojistik regresyon modeli bağımlı değişken için bir olasılık tahmini yapabilmelerini sağlamaktadır.

Bağımsız değişkenler ile bağımlı değişkenler arasındaki bağıntı lineer olmadığından lojik dönüşüm yapılır ve bu şekilde bağımlı ve bağımsız değişkenler arasındaki bağıntının lineer olması sağlanır. Burada bağıntının lineer olmasını sağlamak için yapılan lojik dönüşümde en küçük kareler yöntemi değil maksimum olabilirlik tahmin edici yöntemi kullanılmaktadır. Maksimum olabilirlik tahmin edicisi yöntemiyle bilinmeyen parametrelerin olasılıklarını maksimum olacak şekilde tahmin etmektedir. Tahmin edilen parametreler ile oluşturulan model denkleminde bulunan bağımsız değişkenlerin katsayıları için tahmin ediciler oluşturulur ve uygun bir model ortaya çıkmaktadır. Modelde katsayılar için anlamlılık testleri yapılarak bağımsız değişkenlerin model için önemli olup olmadığı test edilir. Lojistik regresyonun temel amacı; bağımlı ve bağımsız değişkenler arasındaki ilişkiyi, en az değişken ile en iyi uyuma sahip olacak biçimde tanımlayabilen, kabul edilebilir bir model kurmaktır (Atasoy, 2001).

Lojistik regresyon yönteminin daha çok kullanılmasının nedenlerinden bazıları şunlardır:

- Bağımlı değişken ve bağımsız değişken ya da değişkenler arasındaki ilişki doğrusal olmak zorunda değildir.
- Negatif olasılık ile karşılaşma sorunu yoktur.
- Bağımsız değişkenlerin olasılık fonksiyonlarının dağılımıyla alakalı şart bulunmaz.
- Doğrusal regresyon analizinde bulunan varsayımların hiçbiri aranmaz.
- Günümüzde lojistik regresyon analizi yapan paket programlar mevcuttur (Örneğin; SPSS, R, SAS vb. gibi) .

2.1.2. Basit Lojistik regresyon modeli

Lojistik regresyon modeli oluşturulurken kullanılan denklem doğrusal bir denklem değildir. Bunun yerine lojistik denklem kurulmaktadır. Bunun nedeni bağımlı değişkenin kategorik bir değişken olmasıdır. Kullanılan dağılımda değişkenler sıfıra eşit ya da sıfırdan büyük değer alacak şekilde sınırlandırılmıştır. Bu dağılım lojistik dağılım olarak adlandırılır. Modeli oluşturmada kullanılan parametreler aritmetik ortalama ve standart sapmadır. Böylece x_i verilmişken Y 'nin beklenen değeri;

$$\pi(x_i) = E(Y/x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \quad (2.1)$$

şeklindedir. Eşitlik (2.1)'e göre lojik dönüşüm yapıldığında lojistik regresyon modeli;

$$\begin{aligned} g(x_i) &= \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) \\ &= \ln(\pi(x_i)) - \ln(1 - \pi(x_i)) \\ &= \ln\left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}\right) - \ln\left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}}\right) \\ &= \ln(e^{\beta_0 + \beta_1 x_i}) \\ &= \beta_0 + \beta_1 x_i \end{aligned} \quad (2.2)$$

doğrusal regresyon modeli kurulmuş olur.

2.1.3. Basit Lojistik regresyon tahminsel modeli

$(x_i, y_i), i = 1, 2, \dots, n$ için bağımsız n tane gözlem olsun. y_i bernoulli dağıldığından olasılık fonksiyonu;

$$P(Y = y_i) = \pi (x_i)^{y_i} [1 - (x_i)]^{1-y_i}$$

$$y_i = 1 \text{ veya } y_i = 0 \text{ ve } i = 1, 2, \dots, n \quad (2.3)$$

şeklinde olur. Olabilirlik fonksiyonu gözlemlerin birbirlerinden bağımsız olmasından dolayı olasılık fonksiyonların çarpımına eşittir.

$$L(y_i; \beta_0, \beta_1) = \prod_{i=1}^n \pi (x_i)^{y_i} [1 - (x_i)]^{1-y_i} \quad (2.4)$$

Eşitlik (2.2)'de maksimum olabilirlik tahmin edici yöntemi kullanıldığında maksimize olurken fonksiyonun logaritması da maksimize olursa fonksiyon maksimize edilmiş olur ve işlemler daha kolay yapılır. Bu logaritmanın monoton fonksiyon olmasından kaynaklanmaktadır.

$$\ln L(y_i; \beta_0, \beta_1) = \ln \prod_{i=1}^n \pi (x_i)^{y_i} [1 - (x_i)]^{1-y_i}$$

$$\sum_{i=1}^n \left(y_i \ln \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) - y_i \ln \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right) + \ln \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \right)$$

$$\sum_{i=1}^n (y_i(\beta_0 + \beta_1 x_i) - \ln(1 + e^{\beta_0 + \beta_1 x_i})) \quad (2.5)$$

Burada β_0 ve β_1 parametrelerine göre türev alıp sıfıra eşitlenir ise;

$$\frac{\partial \ln L(\beta_0, \beta_1)}{\partial \beta_0} = \sum_{i=1}^n \left(y_i - \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) = 0$$

$$\sum_{i=1}^n (y_i - \hat{\pi}(x_i)) = 0$$

$$\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i)) = 0 \quad (2.6)$$

$$\frac{\partial \ln L(\beta_0, \beta_1)}{\partial \beta_1} = \sum_{i=1}^n \left(x_i y_i - x_i \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right)$$

$$\sum_{i=1}^n x_i (y_i - \hat{\pi}(x_i)) = 0$$

$$\sum_{i=1}^n x_i (y_i - (\beta_0 + \beta_1 x_i)) = 0$$

$$\sum x_i y_i = \beta_0 \sum x_i + \beta_1 \sum x_i^2 \quad (2.7)$$

Eşitlik (2.6) ve (2.7) denklemleri basit doğrusal regresyon denklemleridir ve bu denklemlerden β_0 ve β_1 parametrelerinin tahmin edicileri bulunur. Lojistik regresyon modelinde parametreler yerine yazıldığında tahminsel modelimiz;

$$\hat{\pi}(x_i) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}} \quad (2.8)$$

şeklindedir. Lojik dönüşüm şeklinde yazılmak istenirse;

$$\hat{g}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (2.9)$$

elde edilir.

2.1.4. Lojistik regresyon modeli için önemlilik testleri

Katsayıların önemliliğinin test edilebilmesi için kullanılabilecek 3 tane test bulunmaktadır. Bunlar; olabilirlik oranı testi, wald testi ve skor testidir. Lojistik regresyonda katsayıların testi için temel amaç tahmin değerlerinin sonuç değişkeninin gözlenen değerler ile karşılaştırılmasıdır. Burada modele alınan değişkenlerin önemlilik testi sonucu model için anlamlı veya önemli olmayan değişkenleri modelden çıkarılarak yeni bir model elde edilir. Modele alınan değişkenlerin önemliliği;

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum (x_i - \bar{x})^2} \text{ ve } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

şeklinde test edilir.

Olabilirlik oranı testi

Bir araştırma için yapılan gözlemlerde ve burada tahmin edilen değerlerin log olabilirlik fonksiyonu ile karşılaştırılması ilkesine dayanır.

$$D = -2 \ln \left(\frac{\text{tahminsel modelin olabilirliği}}{\text{doğru modelin olabilirliği}} \right) \quad (2.10)$$

şeklinde ifade edilir. Birçok araştırmacı bu formülü sapma olarak ifade eder. Burada doygun model, araştırma sonucu elde edilen gözlem sayısı kadar parametre içermektedir.

$$\ln L(\beta_0, \beta_1) = \ln \left(\prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \right) \\ \sum_{i=1}^n \left(y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i)) \right) \quad (2.11)$$

Buradan tahminsel modelin olabilirliği;

$$\sum_{i=1}^n \left(y_i \ln(\hat{\pi}(x_i)) + (1 - y_i) \ln(1 - \hat{\pi}(x_i)) \right) \quad (2.12)$$

olarak bulunur. Burada toplam içindeki ifadenin maksimum olabilirlik tahmin edici yöntemi kullanıldığında, $\pi(x_i)$ ' ye göre türev alınıp sıfıra eşitlendiğinde;

$$\frac{y_i}{\hat{\pi}(x_i)} - \frac{(1 - y_i)}{1 - \hat{\pi}(x_i)} = 0 \\ \hat{\pi}(x_i) - \hat{\pi}(x_i)y_i = y_i - \hat{\pi}(x_i)y_i \\ \hat{\pi}(x_i) = y_i \quad (2.13)$$

olarak bulunur. Bulduğumuz değeri eşitlik (2.12)' de yerine yazdığımız zaman doygun modelin olabilirliği;

$$\sum_{i=1}^n (y_i \ln(y_i) + (1 - y_i) \ln(1 - y_i)) \quad (2.14)$$

olarak elde edilir.

$$D = -2 \left[\sum_{i=1}^n \left(y_i \ln(\hat{\pi}(x_i)) + (1 - y_i) \ln(1 - \hat{\pi}(x_i)) \right) - \sum_{i=1}^n (y_i \ln(y_i) + \ln(1 - y_i)) \right] \\ D = -2 \sum_{i=1}^n \left(y_i \ln \left(\frac{\hat{\pi}(x_i)}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}(x_i)}{1 - y_i} \right) \right) \quad (2.15)$$

şeklinde yazılır. Burada bulunan parametrelerin önemli veya önemsiz olduğuna karar vermek için modelin değişkensiz hali ile değişkenli halinin sapmalarının karşılaştırılması gereklidir (Türker, 2016) .

$$G = D(\text{değişkensiz model}) - D(\text{değişkenli model}) \quad (2.16)$$

şeklinde tanımlanır. Buradan hareketle formül;

$$G = -2\ln \left(\frac{\text{değişkensiz tahminsel modelin olabilirliği}}{\text{değişkenli tahminsel modelin olabilirliği}} \right) \quad (2.17)$$

bulunur. Model X_v^2 dağılır. Burada $v = v_2 - v_1$ şeklindedir. v_2 modelde bulunan tüm parametre sayısını gösterirken v_1 , H_0 hipotezinde bulunan parametre sayısıdır.

$H_0: \beta_1 = 0$ için;

$$\pi(x_i) = \frac{e^{\beta_0}}{1+e^{\beta_0}}, \quad \forall x_i \text{ için} \quad (2.18)$$

eşitlik (2.18)' de bulunan her x_i değeri sabit olduğundan dolayı π ile gösterilebilir. Buradan sonuçla log olabilirlik fonksiyonunda yerine yazıldığında ve π 'ye göre türev alıp sıfıra eşitlenirse;

$$L(\pi) = \ln \pi \sum_{i=1}^n y_i + \ln(1 - \pi) \sum_{i=1}^n (1 - y_i) \quad (2.19)$$

$$\frac{\partial L(\pi)}{\partial \pi} = \frac{\sum_{i=1}^n y_i}{\hat{\pi}} - \frac{\sum_{i=1}^n (1 - y_i)}{1 - \hat{\pi}} = 0$$

π 'nin tahmin edicisi;

$$\hat{\pi} = \frac{\sum_{i=1}^n y_i}{n}$$

olarak elde edilir ve

$$1 - \hat{\pi} = 1 - \frac{\sum_{i=1}^n y_i}{n} \quad (2.20)$$

olarak yazılabilir.

$$G = -2\ln \left(\frac{\left(\frac{\sum_{i=1}^n y_i}{n} \right) \left(\frac{\sum_{i=1}^n (1 - y_i)}{n} \right)^{\sum_{i=1}^n (1 - y_i)}}{\prod_{i=1}^n \hat{\pi}(x_i)^{y_i} (1 - \hat{\pi}(x_i))^{(1 - y_i)}} \right) \quad (2.21)$$

şeklinde G istatistiği bulunmuş olur. $v_2 - v_1 = 2 - 1 = 1$ 'den χ_1^2 dağılır. Testin önem seviyesi ise $P(\chi_1^2 > G) = p$ eşitliği ile hesaplanmış olur. Burada bulunan p değeri verilmiş olan α değerinden küçükse H_0 hipotezi reddedilmiş olur ve bu hipotez için kurulmuş olan model anlamsızdır sonucu çıkar. Eğer p değeri α değerinden büyük ise H_0 hipotezi kabul edilir (reddedilemez) ve bu hipotez için kurulmuş olan model anlamlıdır sonucu çıkar.

Wald test istatistiği

Wald test istatistiği tahmini β_1 ve tahmin edilen standart hatanın karşılaştırılmasıdır. $H_0: \beta_1 = 0$ hipotezi altında standart normal dağılım göstermektedir. Standart normal dağılımın karesi χ_1^2 dağılımı gösterdiğinden dolayı wald test istatistiği;

$$W = \left(\frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \right)^2 = Z^2 \quad (2.22)$$

olarak yazılır ve olma olasılığı $P(\chi_1^2 > W) = p$ ile bulunur. Bize verilen α değeriyle bulunan p değeri kıyaslanır. Eğer α değeri bulunan p değerinden büyük ise H_0 hipotezi reddedilir, α değeri p değerinden büyük ise H_0 hipotezi kabul edilir (reddedilemez).

Score testi

Score testinin hesaplanması , $U = \frac{\partial \ln L(\beta_0, \beta_1)}{\partial \beta_1}$ olmak üzere tahmini U değeri ve tahminin standart hatasının karşılaştırılmasına dayanmaktadır.

$H_0: \beta_1 = 0$ için;

$$\pi(x_i) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$$

Her x_i değeri sabit olduğundan dolayı π ile gösterilebilir.

$$\sum_{i=1}^n (y_i - \pi(x_i)) = 0$$

olarak bulunur. Buradan hareketle

$$\hat{\pi}(x_i) = \bar{y} \quad (2.23)$$

olarak bulunur. Buradan $\bar{y}$ değeri eşitlik (2.24)'te yerine yazılırsa;

$$\sum_{i=1}^n x_i (y_i - \pi(x_i))$$
$$U = \sum_{i=1}^n x_i (y_i - \bar{y}) \quad (2.24)$$

olarak elde edilir. U 'nun varyansı ise;

$$Var(U) = Var\left(\sum_{i=1}^n (x_i - \bar{x})(y_i - \pi(x_i))\right)$$

$$\sum_{i=1}^n (x_i - \bar{x})^2 Var(e_i) \quad (2.25)$$

$$Var(U) = \sum_{i=1}^n (x_i - \bar{x})^2 \hat{\pi}(x_i) (1 - \hat{\pi}(x_i)) \quad (2.26)$$

eşitlik (2.23)'te bulduğumuz $\bar{y}$ değeri yerine yazıldığında

$$Var(U) = \bar{y}(1 - \bar{y}) \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.27)$$

elde edilir. Buradan hareketle score test istatistiği;

$$ST = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sqrt{\bar{y}(1 - \bar{y}) \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (2.28)$$

elde edilmiş olur. Çift taraflı p değeri hesaplanır. Burada verilen α değeri hesaplanan p değerinden büyük ise H_0 hipotezi reddedilir, α değeri hesaplanan p değerinden küçük ise H_0 hipotezi reddedilemez (Türker, 2016) .

2.1.5. Tahmin edicilerin güven aralığı

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \sum_{i=1}^n c_i y_i, \quad c_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$Var(\hat{\beta}_1) = Var\left(\sum_{i=1}^n c_i y_i\right) = \sum_{i=1}^n c_i^2 Var(y_i) + 2 \sum_{i=1}^n \sum_{j=i+1}^n c_i c_j Cov(y_i, y_j)$$

$$Var(y_i) = \pi(x_i)(1 - \pi(x_i))$$

$$Cov = (y_i, y_j) = Cov(\beta_0 + \beta_1 x_i + e_i, \beta_0 + \beta_1 x_j + e_j) = Cov(e_i, e_j) = 0$$

buradan,

$$Var(\hat{\beta}_1) = \sum_{i=1}^n c_i^2 \pi(x_i) (1 - \pi(x_i))$$

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2 \pi(x_i) (1 - \pi(x_i))}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} \quad (2.29)$$

olur.

$$Var(\hat{\beta}_0) = Var(\bar{y} - \bar{x}\hat{\beta}_1) = Var(\bar{y}) + \bar{x}^2 Var(\hat{\beta}_1) - 2\bar{x}Cov(\bar{y}, \hat{\beta}_1)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

olmak üzere,

$$Var(\bar{y}) = \frac{1}{n^2} \left(\sum_{i=1}^n Var(y_i) + 2 \sum_{i=1}^n \sum_{j=i+1}^n Cov(y_i, y_j) \right) = \frac{1}{n^2} \sum_{i=1}^n (1 - \pi(x_i))$$

$$Cov(\bar{y}, \hat{\beta}_1) = Cov\left(\bar{y}, \sum_{i=1}^n c_i y_i\right) = \sum_{i=1}^n c_i Cov(\bar{y}, y_i) = 0$$

$$Var(\hat{\beta}_0) = \frac{1}{n^2} \sum_{i=1}^n \pi(x_i)(1 - \pi(x_i)) + \frac{\bar{x}^2 \sum_{i=1}^n (x_i - \bar{x})^2 \pi(x_i)(1 - \pi(x_i))}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} \quad (2.30)$$

olur. Buradan hareketle tahmin edicilerin güven aralığını yazacak olursak;

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \sim N(0,1) \text{ ve } P\left(-Z_{1-\frac{\alpha}{2}} < \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} < Z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

olarak ifade edilir. Tahmin ediciler ifade de yerine yazılırsa;

$$P\left(\hat{\beta}_1 - Z_{1-\frac{\alpha}{2}} SE(\hat{\beta}_1) < \beta_1 < \hat{\beta}_1 + Z_{1-\frac{\alpha}{2}} SE(\hat{\beta}_1)\right) = 1 - \alpha \quad (2.31)$$

elde edilir. Benzer şekilde β_0 için güven aralığı;

$$P\left(\hat{\beta}_0 - Z_{1-\frac{\alpha}{2}} SE(\hat{\beta}_0) < \beta_0 < \hat{\beta}_0 + Z_{1-\frac{\alpha}{2}} SE(\hat{\beta}_0)\right) = 1 - \alpha \quad (2.32)$$

olur (Türker, 2016) .

2.2. Çoklu Lojistik Regresyon

2.2.1. Çoklu Lojistik regresyon modeli

Temelde basit lojistik regresyona benzemektedir. Modelde bulunan bağımsız değişken sayısının birden fazla olduğu hallerde kullanılmaktadır. Örneğin p adet bağımsız değişkenli lojistik regresyon modeli;

$x_i = (x_{i1}, x_{i2}, \dots, x_{ip}) \quad i = 1, 2, \dots, n$ olmak üzere;

$$\pi(x_i) = E(Y/x_i) = \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}} \quad (2.33)$$

olarak tanımlanır ve lojik dönüşüm yapılırsa;

$$g(x_i) = \ln \left(\frac{\pi(x_i)}{1 - \pi(x_i)} \right)$$
$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad (2.34)$$

şeklinde olmaktadır. Burada not edilmelidir ki; $\frac{\pi(x_i)}{1-\pi(x_i)}$ oranına odds oranı da denir.

Odds: Bir olayın gerçekleşme olasılığının gerçekleşmemesi olasılığına oranıdır.

Odds oranı: İki odds'un birbirine olan oranıdır. İlişkiyi özetlemektedir.

Logit: Odds oranının doğal logaritmasının alındığı durumdur. Doğal logaritması alındığında simetrik hale gelir ve doğrusal regresyon analizinde β katsayısının karşılığı olmaktadır.

2.2.2. Çoklu Lojistik regresyon tahminsel modeli

Çoklu regresyon tahminsel modeli basit lojistik regresyon tahminsel modeli gibi bulunur. Burada basit lojistik regresyondan tek farkı birden fazla değişken bulundurmasıdır. Buna göre;

$$x_i = (x_{i1}, x_{i2}, \dots, x_{ip}) \text{ ve } \beta = (\beta_0, \beta_1, \dots, \beta_p) \text{ olsun,}$$

$(x_i, Y_i), i = 1, 2, \dots, n$ için bağımsız n tane gözlem olduğu varsayılırsa; Y_i Bernoulli dağıldığından, olasılık fonksiyonu;

$$P(Y = y_i) = \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \quad y_i = 1 \text{ veya } y_i = 0 \text{ ve } i = 1, 2, \dots, n$$

şeklindedir. Olabilirlik fonksiyonu gözlemleri birbirinden bağımsız olduğundan dolayı olasılık fonksiyonlarının çarpımlarına eşit olduğu görülmüştür. Buradan hareketle;

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \quad (2.35)$$

(2.35)'te maksimum olabilirlik tahmin edici yöntemi kullanılarak maksimize edilirken fonksiyonun logaritması da maksimize edilirse hem fonksiyonun kendisi maksimize edilmiş olur hemde işlemler daha kolay bir şekilde yapılabilir. Çünkü logaritma monoton bir fonksiyondur (Türker, 2016) .

$$\begin{aligned}
\ln L(\beta) &= \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} = \sum_{i=1}^n \left[\ln \left[\pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \right] \right] \\
&= \sum_{i=1}^n \left[\ln[\pi(x_i)^{y_i}] + \ln[(1 - \pi(x_i))^{1-y_i}] \right] \\
&= \sum_{i=1}^n [y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i))] \\
&= \sum_{i=1}^n \left[y_i \ln \left(\frac{\pi(x_i)}{1 - \pi(x_i)} \right) + \ln(1 - \pi(x_i)) \right] \\
&= \sum_{i=1}^n \left[y_i \ln \left(\frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}} \right) - y_i \ln \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}} \right) \right. \\
&\quad \left. + \ln \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}} \right) \right] \\
&= \sum_{i=1}^n \left[y_i (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) - \ln(1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}) \right] \tag{2.36}
\end{aligned}$$

Burada her bir parametreye göre türev alınır ve sıfıra eşitlenir ise; β_j , $j=0,1,2,\dots,p$ ve $x_{i0} = 1$ için türev alınır;

$$\begin{aligned}
\frac{\partial \ln L(\beta)}{\partial \beta_j} &= \sum_{i=1}^n \left[x_{ij} y_i - \frac{x_{ij} e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}} \right] \\
&= \sum_{i=1}^n [x_{ij} (y_i - \pi(x_i))] \tag{2.37}
\end{aligned}$$

$j=0,1,2,\dots,p$ ve $x_{i0} = 1$ şeklinde bulunur. Burada $p+1$ denklem sıfıra eşitlenerek çözülür ve $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$ bulunur. $\hat{\beta}$, çoklu lojistik regresyon modelinde bulunan ifade yerine yazılırsa modelin tahminsel modeli elde edilmektedir (Türker, 2016).

$$\hat{\pi}(x_i) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip}}} = \frac{e^{\hat{g}(x_i)}}{1 + e^{\hat{g}(x_i)}} \tag{2.38}$$

tahminsel model elde edilmiş olur.

2.2.3. Çoklu Lojistik regresyon modeli için önemlilik testleri

Olabilirlik oranı testi

Basit lojistik regresyonda olduğu gibi gözlem değerleri ve tahmin edilen değerlerin log olabilirlik fonksiyonuna göre karşılaştırılmasına dayanmaktadır. Burada;

$$D = -2 \ln \left[\frac{\text{tahminsel modelin olabilirliği}}{\text{doygun modelin olabilirliği}} \right] \quad (2.39)$$

olarak yazılır ve buna sapma adı verilir. Doygun model gözlem sayısı kadar parametre içermektedir.

$$\ln L(\beta_0, \beta_1, \dots, \beta_p) = \ln \left(\prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \right) \\ \sum_{i=1}^n [y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i))] \quad (2.40)$$

ve tahminsel modelin olabilirliği;

$$\sum_{i=1}^n y_i \ln(\hat{\pi}(x_i)) + (1 - y_i) \ln(1 - \hat{\pi}(x_i)) \quad (2.41)$$

şeklinde yazılır.

Wald testi

Wald testi ile modelin önemi test edildikten sonraki adımda değişkenlerin en az biri veya hepsinin p katsayıları sıfırdan farklı olabilir. Wald testi istatistikleri hangi değişkenlerin anlamlı olduğunu görmek için kullanılır ve parametreleri tek tek test edebiliriz. Bu istatistikler standart normal dağılıma sahiptir ve şu şekilde değerlendirilir:

$$W_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \sim Z\left(\frac{\alpha}{2}\right) \quad (2.42)$$

$P(|Z| > W_j) = p$ çift yönlü p değeri hesaplanır ve burada p değeri verilen α değerinden küçük ise H_0 hipotezi bu değer altında reddedilir (Vupa, 2004).

Score testi

$S = U^T(\hat{\beta})I^{-1}(\hat{\beta})U(\hat{\beta})$ olarak hesaplanmaktadır. Burada S değeri χ^2_{s-t} dağılır. Burada parametre tahminleri Wald testinde olduğu gibi teker teker veya istenilen

gruplar baz alınarak test edilebilir. Sonuç olarak p olasılık değeri hesaplanır. Hesaplan değer verilen α değerinden küçük ise H_0 hipotezi reddedilir büyük ise H_0 kabul edilir ($P(\chi^2_{s-t} > S) = p$).

2.2.4. Çoklu Lojistik regresyonda tahmin edicilerin güven aralığı

$$\frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} \sim N(0,1)$$

olmak üzere, her bir β_j için $100(1-\alpha)\%$ güven aralığı;

$$P\left(-Z_{1-\frac{\alpha}{2}} < \frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} < Z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(\hat{\beta}_j - Z_{1-\frac{\alpha}{2}}SE(\hat{\beta}_j) < \beta_j < \hat{\beta}_j + Z_{1-\frac{\alpha}{2}}SE(\hat{\beta}_j)\right) = 1 - \alpha \quad (2.43)$$

olarak bulunmuştur.

2.2.5. Model geliştirme yöntemleri

Araştırmacının yapmış olduğu çalışma içinde birtakım değişkenlerden oluşturmuş olduğu model geliştirilmek istenir. Bundan dolayı modelde olması gereken değişkenler ve modelden çıkarılması gereken değişkenlerin belirlenmesi amaçlanır. Yapılan çalışmada kurulan model için bir veya birden fazla değişkenin bağımlı değişkeni etkileyip etkilemediği ya da modelde bulunan herhangi bir değişkenin bağımlı değişkeni nasıl etkilediği bulunmak istenebilir. Bunun için dikkat edilmesi gereken asıl yer oluşturulan modelin sonucu etkileyip etkilemediğini bilmektir. Çoklu regresyon modelleri; tam regresyon modeli, ileri doğru değişken seçim yöntemi, geriye doğru değişken seçim yöntemi ve son olarak adımsal değişken seçim yöntemidir.

Tam regresyon modeli

Tam regresyon modelinde; bağımsız tüm değişkenlerin aynı anda modele alındığı durumdur. Değişken sayısı model için önemli bir etkidir. Eğer veride değişken sayısı fazla olursa tanımlayıcı modelinde sayısı fazla olacaktır. Değişken sayısının fazlalığı, modele alınan değişkenlerin anlamlılığını belirlerken analiz içinde geçen sürenin uzamasına neden olacaktır. Araştırmacılar değişken sayısının fazlalığı

nedeniyle çalışma süresini de uzattığından dolayı tam regresyon modelinin yerine başka yöntemlerin arayışına gireceklerdir.

İleri doğru değişken seçim yöntemi

İleri doğru değişken seçim yöntemi tam regresyondan farklıdır. Burada bağımsız değişkenler modele aynı anda konulmaz. Modelin başlangıç aşamasında modele yalnızca sabit terim eklenerek analize başlanır. Başka bir değişkenin modele eklenmesi için öncelikli temel düşünce maksimum olabilirlik oranı üzerinde en fazla modele katkısı olabilecek bağımsız değişkenin modele eklenmesidir. Bir sonraki adım ise maksimum olabilirlik oranına göre modele katkısı olan değişken sayısı bitene kadar değişken seçim süreci devam eder. Burada dikkat edilmesi gereken en önemli husus ise oluşturulan modelin en uygun model olup olmadığı konusunda bir garantisinin olmamasıdır.

Geriye doğru değişken seçim yöntemi

Geriye doğru değişken seçim yönteminde temel amaç tüm bağımsız değişkenler ile model kurulduktan sonra model için anlamsız olan değişkenlerin sırasıyla modelden çıkarılması durumudur. Dikkat edilecek en önemli husus; modelden atılan değişkenin sapma değeridir. Bu sapma değeri minimum artış sağlayan değişkenlerin modelden çıkarılarak sapmada yüksek bir artışa neden olan bir değişken belirleninceye kadar devam eder.

Adımsal değişken seçim yöntemi

Adımsal değişken seçim yönteminde modele eklenen birçok değişkenin incelenmesi durumunu daha pratik hale getirmektedir. Daha önceki iki seçim yönteminin (ileri-geriye doğru seçim yöntemi) birlikte hareket etmesi durumudur. Buradan hareketle anlatılmak istenen durum ileri doğru değişken seçimi yaparken modele eklenen bir değişkenin bir sonraki adımda modele katkısının anlamsız bulunması durumunda modelden çıkarılabilmektedir.

2.2.6. Model uygunluğunun değerlendirilmesi

Model uygunluğunun değerlendirme sürecinde araştırmacılar genel olarak modelde bulunan değişkenlerin gerçek olasılık değerleri ile tahmin edilen olasılıklar arasındaki standart farklara bakarlar. Lojistik regresyonda hesaplanabilen hatalar; standart olmayan hatalar, standart hatalar, sapma değerleri, gözlem uzaklık değerleri, cook uzaklığı olarak incelemek mümkündür.

Standart olmayan hatalar

Standart olmayan hatalar (e_i) gerçek olasılıklar ile tahmin edilen olasılıklar arasındaki farka eşittir. Modelin hataları logit ölçekle hesaplanmış ise bu hatalara logit hatalar adı verilmektedir. Logit hatalar Eşitlik (2.44)' teki gibi hesaplanmaktadır (Albayrak, 2006).

$$\text{Logit Hata}_i = \frac{e_i}{\pi_i(1-\pi_i)} \quad (2.44)$$

Standart hatalar

Standart hatalar standart olmayan hataların kendi standart sapmalarına bölünerek elde edilmektedir. Standart hatalar eşitlik (2.45)' te verilmiştir. Burada her birimin standart hatası χ^2 uygunluk istatistiğinin bir bileşeni olarak görülebilir. Büyük örnekler i.in standart hatalar sıfır ortalama ve bir standart sapma ile normal dağılıma uymaktadır (Albayrak, 2006, s. 439-462).

$$Z_i = \frac{e_i}{\sqrt{\pi_i(1-\pi_i)}} \quad (2.45)$$

Sapma değerleri

Doğrusal regresyonda her gözlem için artıkların karesi $(y - \hat{y})^2$, o gözlemin regresyon modelinin tümel uyum eksikliği olan hata kareler toplamının katsayısını gösterir. Lojistik regresyonda sapma değerleri, her gözlemin sapma istatistiği D'ye olan katkısını belirtir. Sapma değeri eşitlik (2.46)' daki gibidir (Alpar, 2013).

$$d_i = \text{sign}(y_i - \hat{\pi}_i) \sqrt{-2[(-y_i \ln(\hat{\pi}_i)) - (1 - y_i) \ln(1 - \hat{\pi}_i)]} \quad (2.46)$$

Gözlem uzaklık değerleri

Tahmini yapılan değişkenlerin uzaklık değerleri arasında hangisinin daha büyük bir etkiye sahip olan birimlerin belirlenmesi amaçlanır. Model için oluşturulmuş değişkenlerin uzaklık değerleri sıfır ile bir arasında bir değer almaktadır. Uzaklık değerleri sabit terim dahil model için tahmin edilen parametrelerin oranının örneklem hacmine oranı ile hesaplanabilmektedir.

Cook uzaklığı

Cook uzaklığı modelde bulunan değişkenlerin etkili gözlem olup olmadığını belirlemek amacıyla kullanılan bir istatistik yöntemidir. Genel olarak bir'in üzerindeki cook değerleri etkili gözlem olarak değerlendirilmiştir (Alpar, 2013) .

Lojistik regresyon analizinde ilişki ölçüsü

Doğrusal regresyon analizinde belirtme katsayısı (R^2), bağımlı değişkendeki değişimin ne kadarının regresyon modelindeki bağımsız değişkenler tarafından açıklandığını göstermesi amaçlanan bir ölçüdür. Ancak lojistik regresyon analizinde bağımlı değişken nominal ölçekli bir değişken olduğundan doğrusal regresyon analizindeki R^2 istatistiği ile lojistik regresyon analizindeki R^2 değerini birebir karşılaştırmak çok doğru olmayacaktır. Bununla birlikte literatürde lojistik regresyon analizinde bir kaç R^2 istatistiğine yer verilmiştir. Bunlar Mc Faden, Cox-Snell ve Nagelkerke R^2 istatistikleri en sık kullanılan istatistiklerdir. Bu istatistikler genel olarak küçük çıkma eğiliminde olduğundan, model uyumunu değerlendirmekten çok en uygun modeli seçmek amacıyla kullanılması önerilmektedir (Albayrak, 2006) . Burada R-kare benzeri istatistiği daha büyük olan modelin daha uygun bir model olduğu kabul edilmektedir.

- **Cox ve Snell R^2 :** Olabilirlik durumu taban alındığında çoklu R^2 istatistiği ile birbirine oldukça benzerdir. İstatistiğin maksimum değerinin genelde birden küçük olması istatistiksel yorumları zor hale getirebilmektedir.
- **Nagelkerke R^2 :** Cox ve Snell R^2 hesaplanan istatistik değerinin 0-1 aralığında değer almasını sağlamak için geliştirilmiş yöntemidir.

2.2.7. Uyum iyiliği

Modelde bulunması gereken bütün değişkenler ele alındıktan sonra modelin sonuç değişkenini açıklamadaki etkinliğini araştırmaya "uyum iyiliği" adı verilmektedir (Ediz, 1997). Lojistik regresyon analizinde Pearson ki-kare testi, Hosmer-Lemeshow testi ve sapma istatistikleri en çok tercih edilen türlerdir.

$$Pearson \chi^2 = \sum_{i=1}^j r_i^2 \quad (2.47)$$

$$Sapma(D) = \sum_{i=1}^j d_i^2 \quad (2.48)$$

Hosmer-Lemeshow ki-kare uyum iyiliği testi lojistik regresyon analizinin tamamının uyumu incelenir.

$$\chi_h^2 = \sum_{i=1}^c \sum_{j=1}^r \frac{(G_{ij} - B_{ij})^2}{B_{ij}} \quad (2.49)$$

Pearson ki-kare testi, genellikle bağımsız değişkenlerin sürekli değişkenler olduğu durumda ya da örneklem sayısı az olduğu durumlarda çalışıldığında, genel olarak hesaplanan ki-kare testinden çok daha güçlüdür. Formülde G, gözlenen değerleri, B ise beklenen değerleri ifade etmektedir. Bu anlamda da genel olarak ki-kare yöntemi ile hesaplanan Omnibus testinin daha güçlü bir alternatifidir. Bu teste ilişkin sonucun anlamlı olmaması ($p > 0.05$), model-veri uyumunun yeterli düzeyde olduğunu gösterir (Garson, 2012) .

2.3. Lojistik Regresyon Analizi Kullanılarak Yapılan Çalışmalar: Literatür Taraması

Literatür taramasında 1999-2017 yılları arasında lojistik regresyon analizi kullanılarak yapılmış olan çalışmalar incelenmiştir.

“Hile denetimi ve tespitinde lojistik regresyon analizinin kullanımı; Sağlık sektöründe bir uygulama” adlı tez çalışmasında bir kamu hastanesinde 2015 yılında gerçekleştirilen laparoskopik kolesistektomi operasyonlarına ait gerçek ameliyat verilerinin içine gerçek olmayan veriler dâhil edilmiş ve oluşturulan lojistik regresyon modelinin bu gerçek olmayan verileri tespit etme başarısını test etmiştir (Boztepe, 2017) .

“Use of binary logistic regression model in Iraq patients of breast cancer” adlı çalışma da regresyon modellerinden biri olan lojistik regresyon analizi incelenmiştir. Diğer istatistik model teknikleri ile aynı amaca sahip olan Lojistik regresyon analizinde amaç, en az değişkenle en iyi uyuma sahip modeli kurabilmeyi amaçlar. Bu çalışmada Irak Süleymaniye’de hizmet veren Hiwa Hastanesine ait 463 meme kanseri teşhisi konmuş kadın hastaya ait veriden yararlanılmış olup, tez için belirlenen risk faktörlerine göre hastalara ait binary lojistik regresyon modelleri incelenmiştir (Mahmood, 2017) .

“İki durumlu (binary) lojistik regresyon modeline Bayesci bir yaklaşım: OECD örneği” adlı tez çalışmasında Bayesci yaklaşımın temel özelliklerini açıklamak ve lojistik regresyon analizini Bayesci ilkelere göre uygulamaktır. Klasik ve Bayesci analiz tekniklerinin birbirlerine karşı zayıf ve güçlü yönlerinden, Bayesci istatistiğin en temel özelliği olan önsel ve sonsal dağılıma değinmektedir. Aynı zamanda karmaşık istatistiksel analizlerin hesaplanmasını kolaylaştıran MCMC algoritmalarında Markov zincirleri Gibss örnekleme gibi bazı simülasyon tekniklerine değinmektedir. Çalışmanın uygulam bölümünde OECD verilerinden bir lojistik regresyon modeli kurulmuştur. Bu model klasik ve bayesci regresyona göre analiz edilmiş, çıkan sonuçları karşılaştırılarak etkili bir analiz oluşturulmaya çalışılmıştır (Yılmaz, 2016) .

“Türkiye’de resmi kurumlara duyulan güvenin kategorik regresyon ve lojistik regresyon analizi ile incelenmesi” adlı tez çalışmasında Bilgiç ve Akyürek (2009) tarafından hazırlanan "Güneydoğu Sorununun Sosyolojik Analizi" isimli çalışmadaki veriden yararlanmıştır. Ülkemizde Türk Silahlı Kuvvetlerine olan güven düzeyi ile çeşitli demografik değişkenler arasındaki ilişkiler, kategorik regresyon analizi ve lojistik regresyon analizi kullanılarak incelenmek istemiştir. Benzer veri tipine sahip modelleri açıklamada kullanılan bu iki regresyon modelinin benzerlikleri ve farklılıkları ortaya konulmaya çalışmıştır (Güç, 2015) .

“Penalized logistic regression” adlı çalışmada Lojistik regresyon analiz türü kategorik verilerin modellenmesinde sıklıkla kullanılan bir istatistiksel tekniktir. Kategorik verilerin en yaygın formu başarılı veya başarısız, evet veya hayır gibi ikili kategorilerin olduğu durumlardır. Regresyon modelini oluşturan değişkenler arasında çoklu doğrusal bağlantı olması durumunda, regresyon modelinin başarı oranı önemli ölçüde düşmektedir. Bu çalışmada, çoklu doğrusal bağlantı sorununu gidermek ve

modelin başarı oranını arttırmak için karesel cezalandırılmış lojistik regresyon modeli kullanılmıştır. En uygun cezalandırma miktarını belirlemek için çeşitli ölçüler kullanılmıştır. Bu iki yöntem gerçek veri setine (koroner kalp krizi verileri) uygulanmış ve performansları bakımından karşılaştırmaları yapılmıştır (Göksülük, 2011).

“İlköğretim 5. sınıf öğrencilerinin okuduğunu anlamada başarılı olup-olmama durumlarının kestirilmesinde etkili olan değişkenlerin incelenmesi” isimli çalışma da okuduğunu anlama başarısını sınıflamada (başarılı, başarısız) etkili olan değişkenlerin modeldeki önem düzeylerini ortaya koymaya çalışmışlardır. 10 ilköğretim okulunun 5. sınıfında okuyan 279 öğrenci üzerinde yürütülmüştür. Bu öğrencilerden elde edilen veriler, lojistik regresyon analizi kullanılarak incelenmiştir. Analiz sonucunda öğrencilerin okuduğunu anlamada başarılı olup olmama durumlarının kestirilmesi olasılığını; cinsiyet, öğrenciye ait kitap sayısı, evde bulunan kitap sayısı, öğretmenin okunan metinlerle ilgili özet yazdırması ve öğretmenin okumayla ilgili ev ödevi vermesi değişkenlerinin anlamlı düzeyde etkili olduğu bulunmuştur (Kutlu vd, 2011).

“Lojistik regresyon analizi ile Eskişehir’in sis kestiriminin incelenmesi” isimli çalışma da bağımlı değişkenin iki düzeyli olması durumunda demografik, davranış ve risk faktörüyle ilgili tahmin çalışmalarında oldukça sık kullanılan lojistik regresyon analizi teorik olarak kısaca incelenmiştir. Sınıflandırma analizlerinin en çok kullanılanlarından biri olan ve birçok konuda uygulama alanı bulunan lojistik regresyonun parametre tahminine ilişkin kestirim yöntemi detaylı bir şekilde incelenmiş, katsayıların yorumu için bahis oranı anlatılmıştır. Daha sonra Eskişehir’de sis olayı ve havacılık faaliyetleri üzerine etkileri incelenmiştir. Araştırmanın uygulama bölümünde ise, Eskişehir’de sis meydana gelmesi olasılığını verecek en uygun denklem belirlenmiş ve lojistik regresyon analizi kullanılarak çalışma somut hale getirilmiştir (Aktaş ve Erkuş, 2009).

“Bebek ölüm nedenleri üzerine bir araştırma” isimli çalışmalarında bebek ölümleri üzerine etkili olan faktörler araştırılmaya çalışılmıştır. Annenin eğitim durumu, gebelik yaşı, bebeğin doğduğu yer ve doğumda bir sağlık personeli yardımı olması ya da olmaması, bebeğin sağlık kontrolünden geçmesi, bebeğin cinsiyeti, emzirilme durumu, annenin doğum sayısı ve bebeğin doğum zamanı, kilosu bağımsız

değişkenler olarak alınıp, bebeğin ölümüne etkisi için iki kategorili bağımlı değişken üzerinde lojistik regresyon analizi uygulanmıştır (Özkan, 2009) .

“Tramvay yolcu memnuniyetinin lojistik regresyon analiziyle ölçülmesi: ESTRAM örneği” isimli çalışma da toplu taşıma araçlarından biri olan tramvaya yönelik yolcu memnuniyeti, Eskişehir tramvay sistemi (ESTRAM) örneğinde, binominal lojistik regresyon analizi ile incelenmektedir. İki üniversiteye sahip olan Eskişehir’de öğrenci nüfusun fazla olması ve tramvay için önemli bir yolcu kitlesi olacakları düşüncesiyle çalışma, her iki üniversiteden basit tesadüfî örnekleme yoluyla seçilen 300 öğrenci üzerinde gerçekleştirilmiştir. Öğrencilerin memnuniyetleri ile ilgili binominal düzeyde, gizil bir değişken kullanılmıştır. Uygulanan binominal lojistik regresyon analizi sonucunda; öğrencilerin ESTRAM’dan memnuniyetleri üzerinde modele alınan tüm bağımsız değişkenlerin negatif etkileri olduğu belirlenmiştir (Girginer ve Cankuş, 2008) .

“Lise öğrencisi ergenlerde depresyonun yaygınlığı ve ilişkili olduğu etmenler” isimli çalışmada çocukluk depresyon envanteri değişkeninde belirlenen kesme noktası üzerinde değere sahip olan verileri etkileyen bağımsız değişkenler arasındaki durumu incelemiştir. Bu sonucun oluşmasında ne gibi etkilerinin olduğunu araştırmıştır. Depresyon olup olmama durumu bağımlı değişken olarak belirlenmiş ve diğer etkenlerin bağımlı değişken üzerindeki etkisi lojistik regresyon analizi ile incelenmiştir (Eskin, 2008) .

“Tıp fakültesi son sınıf öğrencilerinin kaygı düzeyi ve etkileyen faktörler” isimli çalışma da tıp fakültesi son sınıf öğrencilerinin kaygı düzeylerini etkileyen faktörler uygulanan ölçeklerden elde edilen değişkenlerin birbiri ile ilişkileri lojistik regresyon analizi ile incelenmiştir (Canbaz, 2007) .

“Mali başarısızlığın çok değişkenli istatistiksel yöntemlerle belirlenmesi: tekstil sektörü örneği” isimli çalışma da mali başarısızlık riski olan işletmeleri tespit edebilecek bir model bulmaya çalışmışlardır. Tekstil sektöründe faaliyet gösteren İMKB’ye kayıtlı işletmelere ait veriler araştırma çerçevesini oluşturmaktadır. Çalışmada öncelikle bu işletmelerin 2001 yılına ait bilançoları yardımıyla mali oranları hesaplanmış, dönem sonu kâr-zarar durumuna bakılarak, işletmeler o dönem için mali başarısız ya da başarılı olarak değerlendirilmiştir. Krizlere karşı dirençsiz olan tekstil sektörü işletme riski modelinin oluşturulması ve modelde anlamlı olan değişkenlerin belirlenmesi için öncelikle elde edilen mali oranlara faktör analizi

uygulanmış, elde edilen faktör skorları bağımsız değişken olarak alınarak uygulanan lojistik regresyon analizi sonuçlarına göre mali başarısızlığı etkileyen faktörler belirlenmiştir. Araştırma sonunda elde edilen bulgular yorumlanmıştır (Altaş ve Giray, 2005) .

“Lojistik regresyon ve biyolojik alanlarda kullanımı” adlı çalışma da lojistik model katsayılarının en çok olabilirlik tahminleri, önem testleri, desen değişkenlerini tanımlanması, olabilirlik oran testleri, model yeterliliği ve modelin uyum iyiliğinin belirlenmesi incelenmiştir (Şahin, 1999) .



3. POISSON REGRESYON

3.1. Poisson Regresyon Analizi

Poisson regresyonun normal regresyondan farkı modelde bulunan bağımlı değişkenin olasılık dağılımının negatif olmayan sayılardan ve kesikli bir dağılımdan oluşmaktadır. Belirli bir süreç içinde meydana gelen olayları inceler. Verilerin kesikli olması doğrusal regresyon modelleri için yapılacak analizlerin etkisiz, tutarsız ve çelişkili sonuçlar vereceği bilinir. Bundan dolayı kesikli veriler için istenilen koşullar sağlandığında kullanılabilecek regresyon türlerinden biridir. Kesikli ve negatif olmayan veri yapısına sahip olmasından dolayı beklenen sayıların logaritmasının açıklayıcı değişkenlerin doğrusal bir fonksiyonu olduğu varsayımından yola çıkmaktadır.

Poisson regresyon analizleri için en önemli özelliğinden biri de ortalama ve varyans değerlerinin birbirine eşit olmasıdır. Bazı uygulamalarda ortalama ve varyans değerlerinin eşitliği mümkün olmamaktadır. Varyansın ortalamadan büyük olması durumuna aşırı yayılım, varyansın ortalamadan küçük olması durumu da az yayılım olarak adlandırılır. Ortalama ve varyans değerlerinin birbirine eşit olmaması durumunda verilere poisson regresyon analizi uygulanmaz.

3.2. Poisson Regresyon Modeli

Poisson regresyon modeli sahip olduğu özelliklerinden dolayı modelde bulunan parametrelerin tahmini için en küçük kareler yöntemi kullanılması mümkün değildir. En küçük kareler yöntemi yerine yaygın olarak en çok olabilirlik yöntemi kullanılır (Deniz, 2005).

Poisson regresyon modelinin sahip olduğu özelliklerden bazıları şunlardır:

- Yanıt değişkeni olan Y 'nin Poisson dağılıma sahip olması gerekir.
- En çok olabilirlik standart hataları ve t istatistikleri kullanılarak hesaplanan parametrelerin, koşullu ortalama ve varyansını doğru bir şekilde göstermesi gerekir. Burada istenilen özellik koşullu varyans ile ortalamanın eşit olmasıdır.
- Koşullu varyans ve ortalamanın eşit olmaması durumunda daha iyi sonuçlar veren tahmin ediciler kullanılmalıdır.

Poisson regresyon modeli, bağımsız değişken ile μ parametresi ile arasında bulunan ilişki durumunu ifade eden Poisson dağılımından çıkarılır. Poisson regresyon modeli;

$$P(y_i/x_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \quad y_i = 0, 1, .. \quad (3.1)$$

olarak verilir.

3.3. Poisson Regresyon Analizinde Katsayıların Tahmini

Poisson regresyon analizinde parametrelerin tahminleri için en sık kullanılan yöntem en çok olabilirlik yöntemidir. Bağımsız gözlemler için maksimum olabilirlik fonksiyonu;

$$L(y, \beta) = \prod_{i=1}^n f(y_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} = \frac{(\prod_{i=1}^n \mu_i^{y_i}) e^{-\sum_{i=1}^n \mu_i}}{\prod_{i=1}^n (y_i)!} \quad (3.2)$$

şeklinde olmaktadır. Log-olabilirlik fonksiyonu ise;

$$\log(\beta) = \sum_{i=1}^n y_i \ln(\mu_i) - \sum_{i=1}^n \mu_i - \sum_{i=1}^n \ln y_i! \quad (3.3)$$

olarak yazılır.

3.4. Poisson Regresyonda Uyum İyiliği Testi

Bazı veriler ile elde edilen parametrelerin tahminlerinden sonra oluşturulan modelin uyum iyiliğinin sınanması gerekir. Burada istenen tahmin edilen parametrelerin model için ne kadar verimli ise modelin uyum iyiliği de o kadar yüksek çıkacaktır.

Poisson regresyonda uyum iyiliğinin sınanması için kullanılan ölçütler bulunmaktadır. Bu ölçütler: Pearson χ^2 , yapay R^2 ölçümü, sapmalar istatistiği, Akaike bilgi ölçütü ve bayes bilgi ölçütü en yaygın olarak kullanılan ölçütlerdir.

3.4.1. Pearson İstatistiği

Yanıt değişkeni y_i 'nin sahip olduğu μ_i ortalamalı ve ω_i varyanslı bir model için uyum iyiliği ölçüm yöntemi Pearson istatistiğinin formülü;

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\omega}_i} \quad (3.4)$$

olarak ifade edilir. Bu formül ile seride yayılım durumunun aşırı olup olmadığına karar vermek için kullanılır. Formülde bulunan $\hat{\mu}_i$ ve $\hat{\omega}_i$ değerleri μ_i ve ω_i 'nin tahmin edicileridir. Burada χ^2 değeri; $\hat{\mu}$ için hesaplanan serbestlik derecesine göre karşılaştırılması yapılır. Poisson regresyon için uygulanan formülde $\omega_i = \mu_i$ yerine yazılırsa formül;

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \quad (3.5)$$

şeklinde olur. Burada hesaplanan χ^2 değerinin serbestlik derecesine oranının birden büyük değer alması verilerin modele uygun olup olmadığını ve aşırı yayılım durumunun varlığını ifade etmektedir (Deniz, 2005).

3.4.2. Yapay R^2 İstatistiği

Uyum iyiliği ölçütlerinden birisi de yapay R^2 istatistiğidir. Normallik varsayımı gerektirmeyen poisson regresyon modeline en çok olabilirlik tahmini ve sapma değeri ile benzerlik göstermesinden dolayı önerilen R_p^2 ölçüsü,

$$R_p^2 = 1 - \frac{\ell(y) - \ell(\hat{\mu})}{\ell(y) - \ell(\bar{y})} \quad (3.6)$$

şeklinde tanımlanmaktadır. Burada $\ell(y)$ doygun modelin log-olabilirliğini, $\ell(\hat{\mu})$ ilgilenilen modelin log-olabilirliği ve $\ell(\bar{y})$ sadece sabit terimin bulunduğu modelin log-olabilirliğini gösterir. $y_i \geq 0$ gözlenen değerler, $\hat{\mu}_i = \exp(x_i \hat{\beta})$ tahmin edilen değerleri ve $\bar{y}_i = \exp(\hat{\beta}_0)$ ortalama değerleri olmak üzere log-olabilirlikleri sırasıyla,

$$\begin{aligned} \ell(y) &= \sum_{i=1}^n (y_i \ln(y_i) - y_i - \ln(y_i!)) \\ \ell(\hat{\mu}) &= \sum_{i=1}^n (y_i \ln(\hat{\mu}_i) - \hat{\mu}_i - \ln(y_i!)) \\ \ell(\bar{y}) &= \sum_{i=1}^n (y_i \ln(\bar{y}_i) - \bar{y}_i - \ln(y_i!)) \end{aligned} \quad (3.7)$$

biçiminde elde edilmektedir. Eşitlik (2.56)'da verilen log olabilirlik fonksiyonları düzenlenirse yapay R^2 ölçüsüne ulaşılmaktadır (Deniz, 2005) .

3.4.3. Sapma İstatistiği

Uyum iyiliği ölçütlerinden birisi de sapma istatistigidir. Literatürde G^2 istatistiği olarak bilinmektedir ve formül edilmiş hali;

$$G^2 = 2 \sum_i^n y_i \ln \left(\frac{y_i}{\mu_i} \right) \quad (3.8)$$

şeklinde ifade edilmektedir. Bu istatistik değeri 0'a yakınsıyor ise model uyumunda artış gösteriyor denebilir. Eğer bu istatistik değeri tam 0'a eşit ise model uyumunun mükemmel olduğu söylenebilir (Deniz, 2005) .

3.4.4. Akaike bilgi ölçütü

Uygun değerlerin belirli beklenen değerler açısından gerçek değerler olma eğilimine ne kadar yakın bir model olduğunu belirler. Basit model gerçek modelden daha karmaşık olmasına rağmen gerçek modelin bazı özelliklerin tahminlerini daha iyi belirlediği için tercih edilebilir. Böylece, uygun model eğilimi gösteren en iyi modelden gerçeğe en yakın olanıdır. Akaike ölçütü en küçük olan ile model;

$$AIC = 2(p - \ell) \quad (3.9)$$

olarak verilir ve burada p tahmin edilen parametre sayısını gösterir (Agresti, 2002) .

3.4.5. Bayes bilgi ölçütü

Bayes bilgi ölçütü model için oluşturulan parametreler ile model arasındaki uygunluğu test eden yöntemdir.

$$BIC = -2 \ln(L) + q \ln(N) \quad (3.10)$$

olarak ifade edilir. Eşitlikte L; log olabilirlik fonksiyonun maksimum değerini, q açıklayıcı değişken sayısını N örnek büyüklüğünü göstermektedir (Ercanlı vd, 2012).

3.5. Poisson Regresyon Analizi Kullanılarak Yapılan Çalışmalar: Literatür Taraması

Literatür taramasında 2006-2017 yılları arasında poisson regresyon analizi kullanılarak yapılmış olan çalışmalar incelenmiştir.

“Klinik arařtırmalarda poisson regresyon ve negatif binomiyal regresyon yöntemlerinin kullanımı” adlı alıřma da Poisson ve negatif binomiyal regresyon modellerinin performansının rnek geniřlięi, baęımsız deęiřken sayıları ve baęımlı deęiřkenin varyans/ortalama oranı sorunundan nasıl etkilendięi simlasyon alıřması ile aıklamaya alıřmıřtır (Adalıtař, 2017) .

“Poisson regresyon tahmin yöntemlerinin karřılařtırılması” adlı tez alıřmasında Poisson regresyon analizinde tahmin yöntemlerinden; Poisson en ok olabilirlik tahmin yöntemi ve genelleřtirilmiř doęrusal modeller tahmin yöntemlerinin karřılařtırılarak hangi yöntemin daha uygun olduęu konusunda arařtırmacılar a yol göstermek amacıyla yapılmıřtır. Yöntemlerin karřılařtırılması iin 100, 500 ve 1000 rnek byklklerinde yapay veri kullanılmıřtır. Yapılan bu alıřmada sonu olarak kullanılan parametre tahmin yöntemleri arasında uyum iyilięi bakımından farklılık olmadıęı tespit edilmiřtir. Ancak, en ok olabilirlik tahmin edicisinin rettięi standart hata deęerlerinin daha yksek olmasından dolayı genelleřtirilmiř doęrusal modeller yöntemin daha gvenilir olduęu ve kk rnek byklklerinde de daha gvenilir tahmin yapabildięi bulunmuřtur. Sonu olarak, Poisson regresyon analizinde genelleřtirilmiř doęrusal modeller yönteminin kullanılması nerilmiřtir (Memiř, 2013) .

“Poisson regresyonu” adlı tez alıřmasında poisson regresyon analizinin arařtırılması ve bu analiz yardımıyla Aksaray ilinde meydana gelen trafik kazalarının belirleyici nemli nedenlerinin incelenmesi amalanmaktadır. Tez alıřması temel olarak giriř, materyal ve metot, bulgular, sonu ve tartıřma kısımlarından oluřmaktadır. Giriř kısmında tez alıřması hakkında genel bilgiler verilerek uygulamada seilen deneyin nemi vurgulanmıřtır. Materyal ve metot kısmında tezde kullanılan teorik bilgiler tekrar edilmiř ve kullanılan bazı istatistiksel zellikler vurgulanmıřtır. Bulgular kısmında Poisson regresyon analizinin uygulanıřı ve incelenen verinin genel karakteristik yapısı anlatılarak gzlemler hakkında yeterli bilgi sunulmuřtur. alıřmanın sonu ve tartıřma kısmında analiz edilen verilerin sonuları aıklanmıř, analizde nemli bulunan faktrler belirtilmiřtir. İncelenen trafik kaza sayısı verisi  farklı model yardımıyla aıklanmıř ve bu modellerin AIC kriterlerine gre kıyaslaması yapılmıřtır (Tamar, 2013) .

“Doęrusal olmayan poisson regresyon” adlı tez alıřmasında sayıma dayalı elde edilen veriler iin kullanılan Poisson regresyon, Poisson regresyona ait uyum

iyiliği testleri, artıklar, katsayıların anlamlılık testleri ayrıca Poisson dağılımına yaklaşım testi ve Poisson regresyonun bazı özel durumları için kullanılan özel Poisson regresyon modelleri olan Negatif Binom regresyon, birleşik Poisson regresyon, genelleştirilmiş ve kısıtlanarak genelleştirilmiş Poisson regresyon ve yinelenmiş verilerde Poisson regresyon ile birlikte doğrusal olmayan regresyon analizi incelenmiştir. Ayrıca Poisson regresyon ve doğrusal olmayan Poisson regresyona ait tahmin ediciler tanıtılmış ve uygulama olarak doğrusal olmayan bir model kullanılarak Poisson regresyona uyarlanmıştır. Elde edilen doğrusal olmayan regresyon modelinin en küçük kareler ve en çok olabilirlik tahmin edicileri kullanılarak farklı gözlem sayılarında elde edilen değerler ile karşılaştırma yapmıştır (Körez, 2012) .

“Olay zamanı analizinde tesadüfî etkiler Poisson regresyon modeli ile gözlemlenemeyen heterojenliğin incelenmesi” isimli tez çalışmasında yöneticilerin organizasyonla ilgili bağlarının sonlanmasında tesadüfî etkiler Poisson regresyon modeli kullanılarak gözlemlenemeyen heterojenlik kontrol altında tutulmaya çalışılmıştır (Köleoğlu, 2006) .

4. BULGULAR ve TARTIŞMA

Çalışmanın bu bölümünde lojistik regresyon analizi ve poisson regresyon analizi için uygun veri setleri bulunmuş ve bulunan veri setleri ile oluşturulmak istenen modeller için parametre tahminleri yapılmış ve güven aralıkları oluşturulmuştur. Yapılan analizler sonucu modeller oluşturularak yorumlanmıştır.

4.1. Lojistik Regresyon Analizi Uygulaması

Lojistik regresyon analizi için; Dijital Araştırma ve Eğitim Enstitüsü adlı internet sitesinden alınan veriler kullanılarak yapılan çalışmada; belirli bir eğitim seviyesinde bulunan kurumlardan mezun olmuş öğrencilerin, mezun oldukları lisans not ortalaması ve lisansüstü eğitim için girdikleri mülakat sınavı puanları göz önüne alınarak lisansüstü eğitime kabul edilip edilmeme durumunu inceleyen bir çalışmadır.

Çizelge 4.1. Öğrencilerin sınav puanı ve not ortalamasını gösteren tanımlayıcı istatistikler

Değişkenler	Sayı(n)	En küçük-En büyük	Ortalama±S.S
Mülakat sınavı puanı	400	220-800	587,70±115,51654
Lisans not ortalaması		2,26-4,00	3,3899±0,38057

Öğrencilerin mülakat sınavı puanı ortalaması 587,70±115,516 ve lisans not ortalamalarının ortalaması ise 3,38±0,380 olarak hesaplanmıştır. Mülakat sınavında alınan en düşük puanı 220 iken; en yüksek puanın 800 olduğu görülmüştür. Lisans not ortalaması minimum 2,26 iken; en yüksek not ortalaması 4,00 olarak belirlenmiştir.

Çizelge 4.2. Öğrencilerin mezun oldukları kurumlara göre dağılımı

Kurum Derecesi	Sayı(n)	%' de
Çok yüksek	61	15,25
Yüksek	151	37,75
Orta	121	30,25
Düşük	67	16,75
Toplam	400	100,0

Oluşturulan veri setine göre öğrencilerin mezun olduğu kurumların başarı derecesi çok yüksek, yüksek, orta ve düşük dereceli olarak sınıflandırılmıştır. Buna göre lisansüstü eğitime başvuran öğrencilerin mezun oldukları kurumların başarı derecelerine göre en fazla katılımın %37,75 ile başarı derecesi yüksek olan kurumlardan mezun öğrencilerin ve en az katılımın ise %15,25 ile başarı derecesi çok yüksek olan kurumlardan mezun öğrencilerin oluşturduğu gözlenmiştir.

Çizelge 4.3. Lisansüstü eğitime kabul durumunun kurumlara göre dağılımı

Kabul durumu	Kurumların başarı derecesi			
	Çok yüksek	Yüksek	Orta	Düşük
Kabul edildi	28	97	93	55
Kabul edilmedi	33	54	28	12
Toplam	61	151	121	67

Lisansüstü eğitime kabul durumunun kurumlara göre dağılımına bakıldığında lisansüstü eğitime kabul edilen öğrencilerin büyük çoğunluğunun başarısı derecesi yüksek ve orta olan kurumlardan olduğu görülmüştür. Başarı derecesi yüksek olan kurumdan 151 öğrenciden 97'sinin lisansüstü eğitime kabul edildiği belirlenmiştir. Başarı derecesi orta olan kurumda bulunan 121 öğrenciden 93'ünün lisansüstü eğitime kabul edildiği görülmüştür.

Çizelge 4.4. Başlangıçta gözlenen duruma göre öngörülerek oluşturulmuş sınıflandırma tablosu

Gözlenen Durum		Öngörülen durum		
		Kabul Durumu		Doğruluk yüzdesi
		Kabul edildi	Kabul edilmedi	
Kabul Durumu	Kabul edildi	273	0	100,0
	Kabul edilmedi	127	0	0,0
Tamamının Yüzdesi				68,3

Başlangıçta oluşturulmuş olan sınıflandırma tablosu model için hiçbir tahmin yapılmadan en basit mantıkla ele alınmıştır. Açıklayıcı değişkenlerin modele alınmadığı yalnızca sabit değer modelde bulunduğu ve n=400 kişinin lisansüstü

eđitime kabul edildiđi varsayımı altında ancak %68,3 ile dođrulanabilir olduđu saptanmıřtır.

Çizelge 4.5. Eřitlikte bulunan deđiřkenlerin durumu

Katsayılar	β	Standart Hata	Wald	Serbestlik derecesi	P
β_0 (Sabit)	-0,765	0,107	50,764	1	0,000

Yalnızca sabit deđiřkenin modelde bulunduđu ve diđer açıklayıcı deđiřkenlerin modele alınmadıđı bölümdür. Sabit deđerin katsayısı -0,765 olarak hesaplanmıřtır. Standart hata deđeri ise 0,107 olduđu görölmüřtür.

Çizelge 4.6. Eřitlikte bulunmayan deđiřkenlerin durumu

Deđiřkenler	Puan	Serbestlik derecesi	P
Mülakat puanı	13,606	1	0,000
Mezuniyet ortalaması	12,704	1	0,000
Kurum başarı derecesi	25,242	3	0,000
Kurum başarı derecesi1	16,590	1	0,000
Kurum başarı derecesi2	1,801	1	0,180
Kurum başarı derecesi3	5,934	1	0,015
Toplam İstatistik	40,160	5	0,000

Eřitlikte bulunmayan deđiřkenlerin durumuna göre toplam istatistik p deđerini incelenmiřtir. Burada toplam istatistik deđerini $p < 0,05$ olduđundan dolayı eřitlikte yer almayan deđiřkenlerin modele katkısının anlamlı olduđu görölmüřtür. Kurum başarı derecesi2 p deđerini 0,05 deđerinden büyük olduđu için anlamlı bulunmamıřtır. Mülakat puanı, mezuniyet ortalaması ve kurum başarı derecesinin model için anlamlı olduđu görölmüřtür.

Çizelge 4.7. Kurulan ilk model ve son modelin incelenmesi

Model durumu	Ki-Kare(χ^2)	Serbestlik derecesi	P
Step	41,459	5	0,000
Block	41,459	5	0,000
Model	41,459	5	0,000

Omnibus Test
Adım1

Omnibus test incelendiğinde; ilk kurulan model ile son kurulan model arasındaki farkı gösterir. Buradaki fark χ^2 dağılımına sahiptir. $p < 0,05$ olduğundan dolayı açıklayıcı değişkenlerin oluşturduğu modelin anlamlı olduğu sonucuna varılmıştır. Oluşturulan modelin χ^2 değeri 41,459 olarak hesaplanmıştır. Model durumunda bulunan step ise en son aşamadan bu yana modeldeki iyileşmeyi göstermiştir. Yapılan iyileştirmenin model için anlamlı olduğu görülmüştür ($P < 0,05$).

Çizelge 4.8. Model özeti

Adım	-2Log likelihood	Cox&Snell R ²	Nagelkerke R ²
1	458,517	0,098	0,138

Model özet tablosunda -2 log likelihood değeri 458,517; Cox&Snell R² değeri 0,098 ve Nagelkerke R² değeri 0,138 bulunmuştur. Bulunan R² değeri oluşturulan modelde bulunan bağımlı değişkenin değişiminin %13,8'ini tahmin etmiştir.

Çizelge 4.9. Enter yöntemi kullanılarak oluşturulan modelin bulunduğu sınıflandırma tablosu

Gözlenen Durum		Öngörülen durum		
		Kabul Durumu		Doğruluk yüzdesi
		Kabul edildi	Kabul edilmedi	
Kabul Durumu	Kabul edildi	254	19	93,0
	Kabul edilmedi	97	30	23,6
Tamamının Yüzdesi				71,0

Başlangıçta oluşturulan sınıflandırma tablosunda lisansüstü eğitime kabul edilen 273 kişinin 254'ünü doğru sınıflandırmış ve 19 kişiyi yanlış sınıflandırmıştır. Burada 19 kişiyi hatalı sınıflandırmasına rağmen doğruluk yüzdesi %93,0 olarak hesaplanmıştır. Lisansüstü eğitime kabul edilmeyip başlangıç olarak kabul edildi olarak gösterilen 127 kişinin 97'sini yanlış tahmin etmiş ve 30 kişiyi doğru tahmin etmiştir. Doğruluk yüzdesi ise %23,6 olarak hesaplanmıştır. Genel doğruluk oranı incelendiğinde oluşturulan model %71 oranında doğruluk oranına sahiptir. İlk oluşturulan sınıflandırma tablosuna göre doğruluk oranının tahmin yapıldıktan sonra arttığı görülmüştür.

Çizelge 4.10' da enter yöntemi ile eşitlikteki değişkenlerin analizi incelenmiştir. Mülakat puanı $p < 0,05$ olarak bulunmuştur ve odds oranı 1,002 olarak hesaplanmıştır. Mezuniyet not ortalaması $p < 0,05$ olduğu görülmüştür. Mezuniyet not ortalaması odds oranı ise 2,235 olduğu belirlenmiştir. Mülakat puanı güven aralığına bakıldığında 1 değerini içerdiğini görülmüştür. Buna göre mülakat puanı p değeri anlamlı bulunmamıştır. Mezuniyet not ortalaması güven aralığı incelendiğinde 1 değerini içermediğini ve dolayısıyla p değeri anlamlı bulunmuştur. Odds oranı yorumuna bakıldığında mezuniyet not ortalamasındaki 1 birimlik artış kabul edilme olasılığını 2,235 kat artırdığı görülmüştür. Kurum başarı derecesi 1'in odds oranına bakıldığında kurum başarı derecesi çok yüksek olan kurumdaki 1 birimlik artış kabul edilme olasılığını 4,718 kat artırdığı görülmüştür.

Çizelge 4.10. Enter yöntemi ile eşitlikteki değişkenlerin analizi

Değişkenler	β	Standart Hata	Wald	p	Exp(β)	Güven aralığı	
						Alt sınır	Üst sınır
Mülakat puanı	0,002	0,001	4,284	0,038	1,002	1,000	1,004
Mezuniyet not ortalaması	0,804	0,332	5,872	0,015	2,235	1,166	4,282
Kurum başarı derecesi			20,895	0,000			
Kurum başarı derecesi1	1,551	0,418	13,787	0,000	4,718	2,080	10,702
Kurum başarı derecesi2	0,876	0,367	5,706	0,017	2,401	1,170	4,927
Kurum başarı derecesi 3	0,211	0,393	0,289	0,591	1,235	0,572	2,668
β_0 (Sabit)	-5,541	1,138	23,709	0,000	0,004		

4.2. Poisson Regresyon Analizi Uygulaması

Poisson regresyon uygulaması için Araştırma ve Eğitim Enstitüsü adlı internet sitesinden alınan veriler kullanılarak yapılan çalışmada veri setinde bulunan n=200 kişi yer almıştır. Burada program türü olarak; üç ayrı lise programında eğitim gören öğrencilerin bir dönemde matematik dersinden aldıkları notlar ve bir dönemde aldıkları ödül sayısı ele alınmış ve çalışmanın veri setini oluşturmuştur.

Çizelge 4.11. Öğrencilerin matematik sınav puanı ve not ortalamasını gösteren tanımlayıcı istatistikleri

Değişkenler	n	Minimum-Maximum	Ortalama±S.S
Matematik sınav notu	200	33-75	52,65±9,368
Ödül sayısı		0-6	0,63±1,053

Öğrencilerin matematik dersinde aldıkları sınav notu minimum 33 maksimum 75 olarak hesaplanmıştır. Matematik sınav notu ortalaması 52,65±9,36 olarak belirlenmiştir. Ödül sayısı minimum 0, maksimum 6 olarak görülmüştür. Öğrencilerin almış oldukları ödüllerin ortalaması 0,63±1,053 olarak belirlenmiştir.

Çizelge 4.12. Oluşturuların modelin uyum iyiliği test sonuçları

Değişken	Değer	Serbestlik derecesi	Değer/Serbestlik derecesi
Sapma	189,450	196	0,967
Ölçekli Sapma	189,450	196	
Pearson Ki-Kare	212,144	196	1,082
Ölçekli Pearson Ki-Kare	212,144	196	
Log Likelihood ^b	-182,752		

Oluşturulan model için yapılan uyum iyiliği test sonucuna göre modele eklenen her bir değer modelin katkısını ya da değişiminin farkını görebilmek için Pearson Ki-Kare ve Log likelihood değerlerine bakıldığında Pearson Ki-kare değeri 212,144 ve log likelihood değerinin -182,752 olarak hesaplanmıştır.

Çizelge 4.13. Omnibus test sonuçları

Likelihood Ratio χ^2	Serbestlik derecesi	p
98,223	3	0,000

Omnibus testi incelendiğinde $p < 0,05$ 'ten küçük olduğu için modelin anlamlı olduğu saptanmıştır. Log likelihood oranı 98,223 olarak hesaplanmıştır.

Çizelge 4.14. Değişkenlerin modele etkisi

Değişkenler	Wald χ^2	Serbestlik derecesi	p
Lise türü	12,306	2	0,002
Sınav notu	43,806	1	0,000

Bağımlı değişken: ödül sayısı

Lise türü ve sınav notunun modele etkisi incelendiğinde $p < 0,05$ olarak belirlenmiştir. Buna göre lise türü ve sınav notu model için anlamlıdır.

Çizelge 4.15. Modele alınan değişkenler için parametre tahmini

Parametreler	β	Standart Hata	p	Exp(β)	Exp(β) için güven aralığı	
					Alt sınır	Üst Sınır
Lise türü1	-0,37	0,4411	0,000	0,691	0,291	1,640
Lise türü2	0,714	0,3200	0,402	2,042	1,091	3,824
Sınav notu	0,070	0,0106	0,000	1,073	1,051	1,095

Bağımsız değişken: ödül sayısı

Modele alınan değişkenler için yapılan parametre tahminlerinde lise türü1 anlamlı bulunmuş iken; lise türü2 anlamlı değildir ($p > 0,05$) . Sınav notu ise model için anlamlıdır. Lise türü1 için odds oranına bakıldığında 0,691 olarak bulunmuştur. Burada odds oranı için güven aralığı incelendiğinde 1 değerini kapsadığı için lise türü1 model için anlamlı değildir. Lise türü2 ve sınav notu odds oranı güven aralıkları incelendiğinde 1 değerini kapsamadığı ve model için anlamlı olduğu belirlenmiştir. Lise türü 2 için 1 birimlik artış model için 2,042 kat katkı sağladığı gözlenmiştir. Sınav notundaki bir birimlik artış ise 1,073 kat katkı sağladığı belirlenmiştir.

5. SONUÇ ve ÖNERİLER

Lojistik regresyon analizi için yapılan uygulamada iki durum incelenmiştir. Başlangıç olarak açıklayıcı değişkenlerin olmadığı, yalnızca β_0 'ın modelde bulunduğu durum göz önüne alınmıştır. Başlangıç olarak ele alınan modelde referans alınan kategorik veri sınıflandırma tablosunda ele alınmıştır. Sınıflandırma tablosunda öğrenciler lisansüstü eğitime kabul edildiği varsayılmıştır ve toplamda hiçbir tahmin yapmadan %68,3 olarak doğrulanabilir olduğu görülmüştür.

Başlangıçta kurulan modelde bulunan β_0 değerinin modele ters yönde bir etkisi olduğu görülmüştür. β_0 değeri -0,765 olarak hesaplanmıştır. Modele alınan değişkenlerin p değerleri incelenmiştir. Burada p değeri 0,05'ten küçük olduğundan dolayı değişken model için anlamlıdır. Enter yönteminde eldeki bütün değişkenler modele dahil edilmiştir. Başlangıçta oluşturulmuş olan basit modelde 273 kişi lisansüstü eğitime kabul edilmişti fakat enter yönteminde sistem 19 öğrencide hata yapmıştır. Buna rağmen doğruluk yüzdesi lisansüstü eğitime kabul edilen öğrencilerde %93'tür. Lisansüstü eğitime kabul edilmeyen öğrencilerde 97 kişi yanlış tahmin edilirken 30 kişi doğru tahmin edilmiştir. Doğruluk yüzdesi ise %23,6'dır.

Enter yöntemi kullanılarak oluşturulan sınıflandırma tablosunda genel doğruluk oranı model için %71 olarak bulunmuştur. Omnibus testi ile başlangıçta kurulan model ile son model arasındaki fark incelenmiştir. Son durumda modele alınan değişkenlerin anlamlılığı için p değerleri incelenmiş ve çıkarımlar yazılmıştır. Mülakat puanı, kurum başarı derecesi1, kurum başarı derecesi2 model için anlamlı bulunmuştur ($p < 0,05$). Kurum başarı derecesi 3 ise model için anlamlı bulunmamıştır ($p > 0,05$).

Odds oranı için tablodan Exp (β) değerleri incelenmiştir ve değişkenlerin modele katkısı yorumlanmıştır. Mülakat puanı odds oranı 1,002 olarak bulunmuştur. Mülakat puanı odds oranının güven aralığı incelenmiştir ve mülakat puanının güven aralığı 1 değerini içerdiği için modele katkısı olmamıştır. Mezuniyet not ortalaması odds oranı 2,235 olarak bulunmuştur. Mülakat puanı odds oranı güven aralığına bakıldığında 1 değerini kapsamadığı görülmüştür ve burada mezuniyet not ortalamasındaki bir birimlik artışın modele 2,235 kat artış sağladığı görülmüştür. Kurum başarı derecesi1 odds oranı 4,718 olarak bulunmuştur. Kurum başarı

derecesi1 odds oranının güven aralığı 1 değerini kapsamadığından dolayı modele katkısı bulunmuştur. Kurum başarı derecesi1'deki 1 birimlik artışın kabul edilme oranını 4,718 kat artırdığı görülmüştür. Kurum başarı derecesi2 odds oranı 2,401 olarak bulunmuştur. Kurum başarı derecesi2 odds oranının güven aralığı 1 değerini kapsamadığından dolayı modele katkısı bulunmuştur. Kurum başarı derecesi2'deki 1 birimlik artışın kabul edilme oranını 2,401 kat artırdığı görülmüştür. Kurum başarı derecesi3 model için anlamlı bulunmamıştır ($p>0,05$) . Poisson regresyon analizi için yapılan uygulamada; matematik sınav notu ortalaması $52,65\pm9,368$ olarak bulunmuştur. Poisson regresyon için oluşturulmuş olan modele uyum iyiliği testi yapılmıştır. Bu teste göre modele eklenen her değerin modele katkısını Pearson Ki-Kare testi ve log likelihood değerlerine bakarak belirlenmiştir. Omnibus test sonucuna göre log likelihood değeri 98,223 olarak bulunmuştur. Modele eklenen açıklayıcı değişkenlerin anlamlılığı incelendiğinde lise türü ve sınav notu model için anlamlı bulunmuştur ($p<0,05$) .

Ödül sayısının açıklanmasında açıklayıcı değişkenlerin katkısının incelenmesi için parametre tahmini yapılmıştır. Yapılan parametre tahmini sonrası değişkenlerin odds oranı incelendiğinde lise türü1 odds oranı güven aralığı 1 değerini kapsadığı için model için anlamlı değildir. Sınav notu odds oranı 1,073 olarak hesaplanmıştır. Sınav notu odds oranının güven aralığına bakıldığında 1 değerini kapsamadığı görülmüştür. Buna göre sınav notundaki 1 birimlik artış ödül alma sayısını 1,073 kat artırmıştır.

KAYNAKLAR

- Adalıtaş, Ö. 2017. Klinik araştırmalarda Poisson regresyon ve negatif Binomiyal regresyon yöntemlerinin kullanımı. Yüksek Lisans Tezi, Mersin Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, 124, Mersin.
- Agresti, A. 2002. Categorical Data Analysis. John Wiley& Sons. Inc, Publication.
- Aktaş, C. ve Erkuş, O. 2009. Lojistik regresyon analizi ile Eskişehir'in sis kestiriminin incelenmesi. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, Sayı 16, 47-59.
- Albayrak, A. 2006. *Uygulamalı Çok Değişkenli İstatistik Teknikleri*. Asil Yayın Dağıtım, 439-462, Türkiye.
- Alpar, R. 2013. Uygulamalı Çok Değişkenli İstatistiksel Yöntemler, Detay Yayıncılık, 675-700, Türkiye.
- Altaş, D. ve Giray, S. 2005. Mali Başarısızlığın Çok değişkenli İstatistiksel Yöntemlerle Belirlenmesi: Tekstil Sektörü Örneği. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, Cilt 5, Sayı 2, 13-28.
- Atasoy, D. 2001. Lojistik Regresyon analizinin incelenmesi ve bir uygulama. Yüksek Lisans Tezi, Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı Sayısal Yöntemler Bilim Dalı, 73, Sivas.
- Boztepe, E. 2017. Hile denetimi ve tespitinde lojistik regresyon analizinin kullanımı; Sağlık sektöründen bir uygulama. Doktora Tezi, İzmir Kâtip Çelebi Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı, 98, İzmir.
- Canbaz, S. 2007. Tıp Fakültesi son sınıf öğrencilerinin kaygı düzeyi ve etkileyen faktörler. *Genel Tıp Dergisi*, 17, 15-19.
- Deniz, Ö. 2005. Poisson Regresyon Analizi. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, Yıl4, Sayı 7, 59-72.
- Ediz, B. 1997. Lojistik regresyon-ayırma analizi, ayırma sorunu ve kalp hastalarında lojistik model yardımıyla risk ölçütlerinin belirlenmesi. Doktora Tezi, Uludağ Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı, 128, Bursa.
- Ercanlı, İ. , Kahriman, A. ve Yavuz, H. 2012. Trabzon orman bölge müdürlüğü doğu ladini-sarıçam karışık meşcereleri için karışık doğrusal olmayan regresyon denklemleri ile doğu ladini çap-boy modellerinin geliştirilmesi. *Türkiye Süleyman Demirel Üniversitesi Orman Fakültesi Dergisi* 13, 75-84.
- Eskin, M. 2008. Lise öğrencisi ergenlerde depresyonun yaygınlığı ve ilişkili olduğu etmenler. *Türk Psikiyatri Dergisi*, 19(4), 382-389.

- Garson, D. 2012. Testing Statistical Assumptions. Asheboro; Statistical Associated Publishing. North Caroline State University School of Public and International Affairs.
- Girginer, N. ve Cankuş, B. 2008. Tramvay yolcu memnuniyetinin Lojistik regresyon analiziyle ölçülmesi: Estram Örneği. *Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 15(1), 181-193.
- Göksülük, D. 2011. Penalized Logistic Regression. Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, 62, İzmir.
- Güç, K. 2015. Türkiye’de resmi kurumlara duyulan güvenin kategorik regresyon ve Lojistik regresyon analizi ile incelenmesi. Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, 92, Ankara.
- Köleoğlu, N. 2006. Olay zamanı analizinde tesadüfi etkiler Poisson regresyon modeli ile gözlemlenemeyen heterojenliğin incelenmesi. Doktora Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, 73, İstanbul.
- Körez, M. K. 2012. Doğrusal olmayan Poisson regresyon. Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, 73, Konya.
- Kutlu, Ö. , Yıldırım, Ö. , Bilican, S. ve Kumandaş, H. 2011. İlköğretim 5. Sınıf Öğrencilerinin okuduğunu anlama da başarılı olup-olmama durumlarının kestirilmesinde etkili olan değişkenlerin incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 132-139.
- Ucla, 2016, Logistic regression and Poisson regression SPSS data analyzsis examples. Institute for Digital Research&Education Link (<http://stats.idre.ucla.edu/spss/dae/logit-regression/> Erişim Tarihi 04.06.2019 ve <http://stats.idre.ucla.edu/spss/dae/poisson-regression/> Erişim Tarihi 04.06.2019) .
- Mahmood, O. M. 2017. Use of binary Logistic regression model in Iraq patients of breast cancer. Yüksek Lisans Tezi, Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, 73, Van.
- Memiş, M. 2013. Poisson regresyon tahmin yöntemlerinin karşılaştırılması Yüksek Lisans Tezi, Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü Zootečni Anabilim Dalı, 58, Samsun.
- Özkan, S. 2009. Bebek ölüm nedenleri üzerine bir araştırma. *Gazi Tıp Dergisi, Cilt 20, Sayı 4*, 169-172.
- Şahin, M. 1999. Lojistik regresyon ve biyolojik alanlarda kullanımı. Yüksek Lisans Tezi, Kahraman Maraş Sütçü İmam Üniversitesi Fen Bilimleri Enstitüsü Zootečni Anabilim Dalı, 88, Kahraman Maraş.
- Tamar, M. 2013. Poisson Regresyon. Yüksek Lisans Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, 40, Elazığ.

- Türker, H. 2016. Lojistik regresyon ve uygulamaları. Yüksek Lisans Tezi, Çanakkale Onsekiz Mart Üniversitesi Fen Bilimleri Enstitüsü Matematik Anabilim Dalı, 120, Çanakkale.
- Vupa, Ö. 2004. Model Building of logistic regression models. Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, 96, İzmir.
- Yılmaz, A. 2016. İki durumlu (binary) lojistik regresyon modeline Bayesci bir yaklaşım: OECD örneği. Yüksek Lisans Tezi, Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, 105, Van.



ÖZGEÇMİŞ

Adı Soyadı : Mehmet MEŞE
Doğum Yeri : Şişli/İSTANBUL
Doğum Tarihi : 8 Mart 1992
Yabancı Dil : İngilizce

Eğitim Durumu

Lise : Çavuşbaşı Çok Programlı Lisesi (2010)

Lisans : Ondokuz Mayıs Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü (2011-2015)

Yüksek Lisans : Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı (Eylül 2016-)