İKİLİ TEPKİ VERİLERİ İÇİN LOJİSTİK REGRESYON ANALİZİ VE BİR UYGULAMA

Sema TOKOĞLU

(Yüksek Lisans Tezi)

Eskişehir, 2003

İKİLİ TEPKİ VERİLERİ İÇİN LOJİSTİK REGRESYON ANALİZİ VE BİR UYGULAMA

Sema TOKOĞLU

YÜKSEK LİSANS TEZİ

İşletme Anabilim Dalı

Danışman : Prof. Dr. Emel ŞIKLAR

Eskişehir

Anadolu Üniversitesi Sosyal Bilimler Enstitüsü

Şubat - 2003

YÜKSEK LİSANS TEZ ÖZÜ

İKİLİ TEPKİ VERİLERİ İÇİN LOJİSTİK REGRESYON ANALİZİ VE BİR UYGULAMA

Sema TOKOĞLU

İşletme Anabilim Dalı

Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, Şubat, 2003

Danışman: Prof.Dr. Emel ŞIKLAR

İstatistiksel tekniklerden biri olan lojistik regresyon, önce sağlık alanında kullanılmaya başlandı. Daha sonra bu yöntem, yaygın kabulü sonucunda başka alanlardada kullanıldı.

Bu çalışmada, ikili tepki verileri için lojistik regresyon analizi anlatılmıştır. İlk bölümde regresyon kavramı, basit ve çoklu doğrusal regresyon modelleri ve varsayımları, regresyon kullanım alanları açıklanmıştır. İkinci bölümde lojistik regresyon analizine ilişkin kavram, gösterim, modeller ve varsayımları açıklanmış, analiz için kullanılabilecek yöntemler ve model uygunluğunun test edilmesi üzerinde durulmuştur.

Araştırmanın uygulama bölümünde, Axa Oyak Sigorta Acenteliği'nde trafik sigortası olan 100 müşterinin otomobillerini kasko sigortası yaptırmaları ile, aracın yaşı bireyin yaşı ve aylık gelir arasındaki ilişki ifade edilmeye çalışılmıştır.

ABSTRACT

Logistic regression, which is one of the statistical techniques has been first started to use in health field. Later on, due to its highly acceptance, this method has been used in other fields, too.

In this study, logistic regression analysis is described for the binary response data. In the first chapter, regression concept, ordinary and multiple linear regression models and assumptions, usage of regression areas are explained. In the second chapter, concepts, representations, models and assumptions belong to the regression analysis are explained, methods which can be used for analysis and testing of the fitness of the model are pointed out.

In the application section of this study, it is tried to explain the relationship between the 100 customers who have traffic insurance at Axa Oyak insurance company, of being a car insured with the age of car, the age of person and monthly salary.

JÜRİ VE ENSTİTÜ ONAYI

Sema TOKOĞLU'nun "İkili Tepki Verileri İçin Lojistik Regresyon Analizi ve Bir Uygulama" başlıklı tezi **01 Nisan 2003** tarihinde, aşağıdaki jüri tarafından Lisansüstü Eğitim Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca, **İşletme (Sayısal Yöntemler)** Anabilim Dalında, yüksek lisans tezi olarak değerlendirilerek kabul edilmiştir.

			İmza
Üye (Tez Danışmanı)	:	**Prof.Dr.Emel ŞIKLAR**	
Üye	:	**Doç.Dr.Hasan DURUCASU**	
Üye	:	**Yrd.Doç.Dr.Fikret ER**	

Prof.Dr.Nurhan AYDIN
Anadolu Üniversitesi
Sosyal Bilimler Enstitüsü Müdürü

ÖNSÖZ

Sema Tokoğlu

Çalışmalarım sırasında yardımlarını esirgemeyen, yardımlarıyla beni yönlendiren danışman hocam Sayın Prof. Dr. Emel ŞIKLAR'a (Anadolu Üniversitesi) en içten teşekkürlerimi bir borç bilirim.

Çalışmalarım süresince yardımlarını esirgemeyen hocalarım Sayın Doç. Dr. Hasan DURUCASU'ya (Anadolu Üniversitesi) ve Sayın Yrd. Doç. Dr. Fikret ER'e (Anadolu Üniversitesi), uygulama verilerinin derlenmesi konusunda bana yardımcı olan Axa Oyak Sigorta A.Ş. Acentesi Sayın Beytullah ASİL'e ve çalışma boyunca göstermiş oldukları destek ve anlayıştan dolayı aileme en içten teşekkürlerimi sunarım.

İÇİNDEKİLER

BİRİNCİ BÖLÜM

REGRESYON

İKİNCİ BÖLÜM

LOJİSTİK REGRESYON ANALİZİ

ÜÇÜNCÜ BÖLÜM

LOJİSTİK REGRESYON ANALİZİNİN SİGORTA ALANINDA KULLANILMASINA İLİŞKİN BİR UYGULAMA

TABLOLAR LİSTESİ

GİRİŞ

Toplum içinde yaşayan insanın malı ve hayatı sayılamayacak kadar çeşitli tehlikelerle karşı karşıyadır[1]. Bu tehlikeler evlerinin yanması, evlerini su basması, mallarının çalınması, kazaya uğrama, sakatlanma vs. olabilir. Kişilerin yalnız yaşayamamaları sonucu toplum yaşamı doğmuş ancak bu değişim kişilerin gereksinimlerini değiştirememiştir. Yaşamın devamı için zorunlu ve gerekli olan beslenme, barınma, giyinme vb. gereksinimler çoğalmaktadır. Bu duruma günümüzde kişinin tek bir elbise ile yetinmeyerek mevsimlere göre, bulunduğu ortama göre değişik elbiseler giyme gereksinimi duyması örnek verilebilir[2].

Gereksinimlerin gelişimine ve sayısal artış göstermesine paralel olarak zaman içinde bir sıralamaya tabi tutulması gerekmektedir. Çünkü kişi sadece günlük gereksinimleri ile birlikte gelecektekilerini de düşünmek zorundadır. Bu düşünceye bağlı olarak ta elde edilen gelirin bir kısmı tasarruf edilerek (saklanarak) gelecekteki gereksinimlerin giderilmesine ayrılmaktadır. Belki gelecekte gerçekleşmesi düşünülen gereksinimler hiçbir zaman gerçekleşmeyecektir. Ancak önlem alınması zorunludur. Bir diğer ifade ile gerçekleşmesi kesin olmamakla birlikte gerçekleşme olasılığı bulunan ve gerçekleşmesi tesadüflere bağlı olan gereksinimleri düşünmek ve önlem almak hem bir gereklilik hem de bir zorunluluktur[3].

Zaman ilerledikçe ve olaylar geliştikçe kişinin karşı karşıya kalacağı gerçek kendisini şu iki noktada gösterecektir.

---

[1] **Temel Sigorta Bilgileri**, (İstanbul: Türkiye Genel Sigorta A.Ş.), s.3.

[2] Mehmet Özkan, **Sigorta İşlemleri Ve Muhasebesi**, (İstanbul: Bilim Teknik Yayınevi, 1998), s.8.

[3] **Aynı,** s.9.

-Kişinin gelecekteki gereksinimlerinin tümünü tek başına karşılaması olanaksızdır. Çünkü sahip bulunduğu olanaklar bu gereksinimlerin karşılanmasında tek başına yetersiz kalmaktadır.

-Aynı gereksinimleri duyan başka kişilerde vardır. Ancak bu gereksinimlerin tümünün birlikte ve aynı zamanda gerçekleşmesi zayıf bir olasılıktır[4].

Kişinin sahip olduğu tüm malı, eşyası, gelirinin kaynağı olan sağlık ve yaşamının, hastalık, kaza ve ölüm gibi zarar verici olayların tehlikesine uğrama olasılığı her zaman söz konusudur. Bu olasılıklardan birisinin gerçekleşmesi kişinin ekonomik varlığının yada yaşamının sona ermesine neden olabilir[5]. İnsanlar kazançlarının bir bölümünü tasarruf ederek bu gibi olayların zararlarını gidermeye çalışabilirler. Ancak mevcut tasarruf her zaman zararı karşılamayabilir. İnsanlar zararları tek başlarına karşılamak yerine, birleşerek aralarında bölüşebilirler[6].

Kişinin zarar verici olayların tehlikesine uğrama olasılığı kaza olarak, kazaların gerçekleşme olasılığıda riziko veya risk olarak nitelendirildiğinde kazanın ne zaman gerçekleşeceği önceden hiçbir şekilde kestirilememekte ve buna bağlı olarakta önlenememektedir. Bunun yanısıra önceden önlenmesi olanaksız olan kazaların sorumlusunu belirlemekte olanaksızdır. Bu nedenle kişiler önceden önlem alma gereksinimini duymaktadır[7].

---

[4] Özkan, **a.g.e.,** s.9.

[5] **Aynı,** s.9.

[6] Temel Sigorta Bilgileri, **a.g.e.,** s.3.

[7] Özkan, **a.g.e.,** s.9.

Şu halde sigorta kişinin gelecekteki gereksinimlerini karşılamak arzu ve isteği sonucu oluşan bir davranış biçimi[8] olup insanların zararları birlikte karşılama ihtiyacından doğmuştur[9].

Kişiler ölüm, hastalık, sakatlık, iş gücünün kaybı, yangın, deprem, su baskını gibi olaylara karşı prim ödeyerek kendilerini ve ekonomik değerlerini güvence altına almakta, böylece de sigortacılık endüstrisine yardım etmektedirler[10].

Bu yönü ile değerlendirildiğinde sigortanın ekonomik yaşamın emniyet - güvenlik sübabı olduğu yada ekonomik yaşamın yayları olduğu ifade edilmektedir[11].

Ekonomik yaşamda karşılaşılan olayların bazıları var-yok, başarılı-başarısız, az-orta-çok, olumsuz-olumlu-çok olumlu ya da daha çok kategoriler biçiminde sonuçlanabilir. Bu sonuçların ortaya çıkmasında bir çok faktör rol oynar.

Bilimsel araştırmalarda güvenilir sonuçlara varılabilmesi için, kararların istatistik yöntemlere dayandırılması gerekmektedir. İncelenen olayların karmaşık ve bu olayların çözümü için önerilen yolların fazla olması, olayı açıklamada kullanılacak değişken sayısını artırmaktadır. Bu amaçla olayların çözümünde, birden fazla değişkeni konu alıp, bunların analizleriyle uğraşan istatistiksel modellerin kullanılması gerekmektedir.

Elde edilen verilerin analiziyle ilgilenildiğinde, kuramsal istatistik bir modelin matematiksel fonksiyonlarla ifade edilmesi gerekmektedir. Çünkü

---

[8] Özkan, **a.g.e.,** s.8.

[9] Temel Sigorta Bilgileri, **a.g.e.,** s.3.

[10] Özkan, **a.g.e.,** s.v.

[11] **Aynı,** s.10.

bu fonksiyonlar, gözlenen verilerden ileride gerçekleşmesi muhtemel olaylar hakkında tahmin yapılmasına ve olaylara etki eden faktörlerin belirlenmesine olanak sağlarlar[12]. Bu faktörlerle, sonuçlar arasındaki neden sonuç ilişkisini belirlemede lojistik regresyon modelleri kullanılmaktadır.

Sağlık ve yaşam üzerine etki eden faktörlerin neden sonuç ilişkilerinin doğru biçimde ortaya konması için sağlıklı ve yeterli verilere gereksinim vardır.

Lojistik regresyon analizinin ikili tepki verileri için incelenmesi ve bu analizin sigorta alanında bir uygulama ile gösterilmesinin amaçlandığı bu çalışma üç bölümden oluşmaktadır.

Birinci bölümde regresyon kavramı, basit ve çoklu doğrusal regresyon modelleri ve varsayımları, regresyon kullanım alanları açıklanmıştır.

İkinci bölümde lojistik regresyon analizinin teorik yapısına yer verilmiştir. Bu bölümde lojistik regresyon kavramı, odds değeri, odds oranı, lojit fonksiyon, lojit dönüşüme ilişkin özellikler, doğrusal model ve lojistik regresyon, ikili lojistik regresyon modeli, iki grup lojistik modele ilişkin varsayımlar, analiz için kullanılabilecek yöntemler ve model uygunluğunun test edilmesi üzerinde durulmuştur.

Son bölüm teorik olarak incelenen lojistik regresyon analizinin gerçek hayattaki geçerliliğini görebilmek amacıyla sigorta alanında ikili veri yapısına uygulanan lojistik regresyon yönteminin incelendiği uygulama bölümüdür.

---

[12] Ertuğrul Çolak, **"Koşullu ve Sınırlandırılmış Lojistik Regresyon Yöntemlerinin Karşılaştırılması ve Bir Uygulama."** (Yüksek Lisans Tezi , Osmangazi Üniversitesi, Sağlık Bilimleri Enstitüsü, Eskişehir, 2002), s.2.

BİRİNCİ BÖLÜM

REGRESYON

Lojistik regresyon analizinin anlaşılması açısından, bu bölümde regresyon kavramı, basit ve çoklu doğrusal regresyon modelleri ve varsayımları, regresyon kullanım alanları açıklanmıştır.

1. REGRESYON KAVRAMI

Regresyon (Bağlanım), sözlük anlamıyla bir şeyi başka bir şeye bağlama işi ve biçimidir. Bilimsel olarak regresyon terimi, bir değişken ile başka bir ya da birden çok değişken arasında ilişki kurma işini ve ilişkinin biçimini anlatır[13].

İstatistiksel anlamda iki değişken arasındaki ilişki, bunların değerlerinin karşılıklı değişimleri arasında bir bağlılık şeklinde anlaşılır. Örneğin; pancar üretimi arttığında fiyatı düşüyorsa veya azaldığında fiyatı yükseliyorsa, bu iki değişken arasında ilişki olduğunu gösterir. Aslında değişkenler arasındaki bu ilişki "neden-sonuç" ilişkisidir. İşte değişkenler arasındaki neden-sonuç ilişkisinin matematiksel bir fonksiyonla ifade edilmesi regresyon analizinin konusunu oluşturmaktadır. Regresyon analizi, biri bağımlı (açıklanan) değişken, diğeri bağımsız (açıklayıcı) değişken olmak üzere en az iki değişken arasındaki ortalama ilişkinin

---

[13] Emel Şıklar, **Regresyon Analizine Giriş,** (Eskişehir: Anadolu Üniversitesi Fen Fakültesi Yayınları, No:16, 2000), s.1.

matematik bir fonksiyon şeklinde yazılmasıdır. Bu fonksiyona regresyon denklemi adı verilmektedir[14].

Bağımlı değişken; değeri başka değişkenler tarafından etkilenen değişkene bağımlı değişken denir. İncelenen bir olayda, sonuç değişken bağımlı değişkendir. Bağımlı değişken genellikle Y harfi ile gösterilir.

Bağımsız değişken; değeri rasgele koşullara göre belirlenen, bağımsız olarak değişim gösteren ve başka değişkenlerin de değişimi üzerine etkide bulunan değişkenlere, bağımsız değişken denir. Bağımsız değişken genel olarak X harfi ile gösterilir[15].

Regresyon denklemi yardımıyla bağımsız değişkenlerin çeşitli değerlerine karşılık bağımlı değişkenin alacağı değer tahmin edilir. Bağımlı değişkeni etkileyen bağımsız değişkenlerin saptanmış olması da bağımlı değişken üzerinde geliştirilecek politikalarda hangi değişkenlerin önem kazandığını ortaya çıkarmaktadır[16].

Değişkenlerden birinin değerleri azalıp artarken, diğerinin değerleri de azalıp artıyorsa (veya ters yönde değişmeler gösteriyorsa) söz konusu bu değişkenler arasında bir ilişki olduğu açıktır. Buna karşılık birinin değeri azalır veya çoğalırken, diğerinin değeri değişmiyorsa bu değişkenler arasında bir ilişkinin varlığından söz edilemez. Değişkenler arasındaki ilişkinin fonksiyonel şekli, yönü ve derecesi bu noktada önem kazanır. "İlişkinin fonksiyonel şekli" değişkenler arasındaki ilişkiyi açıklayan matematiksel fonksiyon tipini belirtir. Bu ise regresyonda doğrusal ilişki olarak karşımıza çıkar. "İlişkinin yönü" ise iki değişkenin değişim yönünü

---

[14] Şıklar, **a.g.e.**, s.1-2.

[15] Çolak, **a.g.e.**, s.5.

[16] Şıklar, **a.g.e.**, s.2.

ortaya koyar. Son olarak "İlişkinin derecesi", iki değişken arasındaki bağlılığın kuvvetli ya da zayıf olduğunu belirler.

Korelasyon, bağımlı değişkenle bağımsız değişken veya değişkenler arasındaki ilişkinin derecesini gösteren bir katsayıdır[17]. r ile gösterilen bu katsayı Y ile X değişkeni arasında, ilgili veriler çerçevesinde ne derece sıkı bir ilişki bulunduğunu belirtir[18].

Korelasyon katsayısı –1'den küçük olamayacağı gibi +1'dende büyük çıkamaz, yani

$$-1 \leq r \leq +1$$

olmalıdır. Pozitif işaretli korelasyon katsayısı, değişkenlerden birinin değeri artarken diğerinin değerinin de arttığını, negatif işaretli korelasyon katsayısı ise değişkenlerden birinin değeri artarken diğerinin değerinin azaldığını gösterir. r=0 olduğunda ise, değişkenler arasında doğrusal bir ilişki olmadığı söylenir[19].

2. BASİT VE ÇOKLU DOĞRUSAL REGRESYON MODELLERİ

Regresyon analizinde bağımsız (açıklayıcı) değişken sayısı bir olduğunda basit regresyon modelinden, iki veya daha fazla olduğunda ise çoklu regresyon modelinden söz edilir. Diğer taraftan değişkenler arasındaki ilişki doğrusal olduğunda doğrusal regresyon modeli, doğrusal olmadığında ise doğrusal olmayan regresyon modeli söz konusu olur[20].

---

[17] Şıklar, **a.g.e.**, s.2.

[18] Tümay Ertek, **Ekonometriye Giriş**, (İstanbul: Beta Yayınları, 1996), s.174.

[19] Şıklar, **a.g.e.**, s.32.

[20] **Aynı**, s.5.

Basit ve çoklu doğrusal regresyon analizleri bağımlı değişken ile bağımsız değişken ya da değişkenler arasındaki matematiksel ilişkiyi analiz etmekte kullanılmaktadır[21]. Bu ilişki matematiksel bağıntılar yardımıyla temsil edilir.

Değişkenler veri setinde yer alan ölçümlere göre (kullanılan ölçeklerin ürettiği sayısal değerlerin anlamları bakımından) temel olarak dört gruba ayrılarak incelenirler.

Bunlar ;

- İsimsel (Nominal) ölçekli değişkenler (String variables)

- Sıralı (Ordinal) ölçekli değişkenler

- Aralıklı (İnterval) ölçekli değişkenler

- Oransal (Proportional) ölçekli değişkenler

olarak isimlendirilir[22].

İsimsel Değişken : Değişkenin değerleri isimsel olarak saptanıyorsa değişkene isimsel değişken denir. İsimsel değişkenin seçenekleri isim (karakter, string) olarak belirlenir. Bu seçenekler kodlanarak sayısal değerlere dönüştürülebilir.

İsimsel değişken ancak iki değer alabiliyorsa ikili (Binary, Dichotomous) değişken, üç ve daha fazla farklı değer içeriyorsa Politomous değişken adını alır.

---

[21] Kazım Özdamar, **Paket Programlar İle İstatistiksel Veri Analizi-1,** (Eskişehir: Kaan Kitabevi, 1997), s.461.

[22] **Aynı,** s.101.

Sıralı Değişken : Seçenekleri birbirlerini artan biçimde izleyen değerler içerir.

Aralıklı Değişken : Sınıfları artan sırada ve birbirlerini izleyen sınıfların birbirlerinin alt ve üst değerleri arasında geçişlilik olan değişkenlerdir.

Oransal Değişken : Değerleri Uluslararası Ölçü Birimlerinden herhangi birine göre saptanan değerleri ondalıklı olarak birbirlerine geçişli olan ve ölçü titizliği arttıkça ölçeğin daha alt birimlerine göre ölçüm değerleri saptanabilen değişkenlerdir[23].

Bağımsız değişkenler ortak ya da faktör değişkenler olarak ele alınır[24].

Ortak Değişken: Bağımlı değişkenlerle birliktelik içindeki, aralıklı ya da oransal ölçekli sürekli değişkenlere ortak değişken denir.

Faktör Değişken: Bağımlı değişkenlerle neden sonuç ilişkisi içinde olan sınıflayıcı veya sıralı ölçekli ya da isimsel değişkenlere faktör değişken denir[25].

Basit doğrusal regresyon modeli;

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (1)$$

şeklinde ve çoklu doğrusal regresyon modeli ise;

---

[23] Özdamar, **a.g.e.**, s.101.

[24] **Aynı**, s.463.

[25] Çolak, **a.g.e.**, s.5.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon \qquad (2)$$

şeklinde ifade edilir.

Burada;

Y : Bağımlı değişkeni,

X / $X_1, X_2, \ldots, X_k$: Bağımsız (Açıklayıcı) değişken / değişkenleri,

β_0 : Bağımsız değişken / değişkenler sıfır değerini aldığında bağımlı değişkenin değerini, yani sabiti,

$\beta_1, \beta_2, \ldots, \beta_k$: Bağımsız değişkenlerin regresyon katsayılarını,

ε : Hata terimini,

k : Bağımsız değişken sayısını, göstermektedir.

Basit ve çoklu doğrusal regresyon modelinde, bağımlı değişkenin, verilen bağımsız değişken ya da değişkenlerin değerlerine göre beklenen değeri (ortalama değeri) basit doğrusal regresyon modeli için;

$$E(Y/X) = \beta_0 + \beta_1 X \qquad (3)$$

şeklinde ve çoklu doğrusal regresyon modeli için ise;

$$E(Y/X_1, X_2, \ldots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k \qquad (4)$$

şeklinde gösterilir. Hata teriminin beklenen değeri sıfırdır. Parametre tahminleri yukarıdaki bağımlı değişkenin beklenen değerlerini veren modellere göre yapılmaktadır[26].

3. DOĞRUSAL REGRESYON MODELİNİN VARSAYIMLARI

Söz konusu varsayımlar:

i) Hata terimi ile ilgili varsayımlar

ii) Bağımsız değişken X ile ilgili varsayımlar

iii) Diğer varsayımlar olarak üç grupta toplanabilir.

Hata terimi ile ilgili varsayımlar:

- Hata terimi ε_i ortalaması sıfıra eşit rassal bir değişkendir.
- Hata terimi ε_i'nin dağılımı normaldir.
- Hata terimi ε_i 'nin değerleri arasında ilişki yoktur.
- Hata terimi ε_i'nin varyansı her X_i değeri için eşittir.

Bağımsız değişken X ile ilgili varsayımlar:

- X değişkeni hata terimi ε ile ilişkili olmayıp, stokastik değildir.
- X değişkeni tekrar eden örnek değerlerine göre sabittir.

Diğer varsayımlar:

- Model belirleme hatası taşımamaktadır.
- Bağımsız değişkenler arasında tam veya kuvvetli bir ilişki yoktur. (Çoklu doğrusal regresyon modellerinde geçerlidir.)[27]

---

[26] Çolak, **a.g.e.**, s.6-7.

[27] Şıklar, **a.g.e.**, s.6-7.

Varsayım-1- Hata Terimi ε_i'nin Rassal Değişken Olması

Hata teriminin alabileceği değer şansa bağlıdır. X'in her bir değeri için hata terimi pozitif, negatif veya sıfır değerlerini belli bir olasılıkla alabilmektedir. Yani ε, stokastik bir değişkendir ve değerleri önceden kesin olarak bilinmemektedir.

Bazı bağımsız değişkenlerin modele alınamaması matematiksel modelin yanlış seçilmesi, değişkenlerdeki ölçme hatalarından kaynaklanan çeşitli hataların hesaba katılması amacıyla hata terimine yer verilmektedir. Bu gibi durumlarda ε'un artı değer alabileceği gibi eksi değer de alabileceğini gösterir. Modele dahil edilmeyen değişkenlerin etkisi bazen Y'yi gözlenebilecek olan değerinden daha büyük bazen de daha küçük olması yönünde gerçekleşecektir. Bu da ε'un stokastik olduğunu göstermektedir[28].

Varsayım-2- Hata Terimi ε'un Ortalamasının Sıfır Olması

X'in her bir değeri için hata terimi pozitif, negatif veya sıfır değerlerini belli olasılıklarla alabilmektedir. İşte ε'un bu özelliğinden yararlanarak beklenen değerinin (ortalamasının) sıfır olduğu varsayımı yapılmaktadır,

$$E(\varepsilon_i) = 0 \qquad i=1, 2, ..., N$$

$$E(\varepsilon_i|X) = 0$$

koşullu beklenen değer şeklinde de ifade edilebilir, yani verilen bir X_i değeri için ε_i'nin ortalaması sıfıra eşittir varsayımıyla aynıdır. Bu varsayım sayesinde

---

[28] Şıklar, **a.g.e.**, s.7.

$$E(Y_i|X_i) = \beta_0 + \beta_1 X_i$$

sonucuna ulaşılır. (1) ifadesinin iki tarafının verilen bir X_i değeri için beklenen değeri aşağıdaki gibidir.

$$E(Y_i|X_i) = E(\beta_0 + \beta_1 X_i + \varepsilon_i)$$

$$= E(\beta_0 + \beta_1 X_i) + E(\varepsilon_i)$$

$$= \beta_0 + \beta_1 X_i$$

Bu eşitlik sonuçta $E(\varepsilon_i|X_i) = 0$ varsayımına denktir[29].

Varsayım-3- Hata Terimi ε'un Normal Dağılıma Sahip Olması

Her X_i için hata teriminin değerleri kendi ortalamaları etrafında çan eğrisi şeklinde simetrik bir dağılım gösterir.

$$\varepsilon \sim N(0,\sigma_\varepsilon^2)$$

Regresyon çözümlemesinde istatistiksel çıkarsamalarda gerekli olan bir varsayımdır[30].

Varsayım-4- Hata Terimi ε_i Değerleri Arasında İlişki Olmaması

Hata terimlerinin ardışık değerleri birbirinden bağımsızdır. Başka bir ifadeyle, hata terimi değerleri arasında otokorelasyon yoktur. Bu varsayım ε_i ve ε_j'nin kovaryanslarının sıfıra eşit olmasını gerektirir.

---

[29] Şıklar, **a.g.e.**, s.7-8.

[30] Aynı, **a.g.e.**, s.8.

$$Kov\,(\varepsilon_i\,\varepsilon_j) = E[(\varepsilon_i) - E(\varepsilon_i)]\,[(\varepsilon_j) - E(\varepsilon_j)]$$

$E(\varepsilon_i) = 0$ olması nedeniyle

$$Kov\,(\varepsilon_i\,\varepsilon_j) = E(\varepsilon_i\,\varepsilon_j) = 0 \qquad i \neq j \text{ için}$$

Bu varsayıma göre; Y_i rassal değişkeninin aldığı değerler birbirinden istatistiksel olarak bağımsızdır. Her bir gözlem diğer gözlemlerden bağımsız olarak elde edilmelidir. Bu ifade $Kov(Y_i\,Y_j) = 0$, $i \neq j$ şeklinde gösterilebilir[31].

Varsayım-5- Sabit Varyanslılık Varsayımı

Hata teriminin varyansı X değerlerine göre değişmeyip sabit kalmaktadır. ε_i'nin varyansının her X_i için eşit olduğu varsayımı aşağıdaki şekilde ifade edilebilir.

$$Var(\varepsilon_i|X_i) = E[\varepsilon_i - E(\varepsilon_i)]^2$$

$E(\varepsilon_i) = 0$ olduğundan

$$Var(\varepsilon_i|X_i) = E[\varepsilon_i^2]$$

$$Var(\varepsilon_i|X_i) = \sigma^2 \text{ veya}$$

$Var(\varepsilon_i) = \sigma^2$ ile gösterilmektedir.

[31] Şıklar, **a.g.e.**, s.8-9.

Aynı zamanda $f(Y|X=X_i)$ olasılık dağılımlarının varyansı $Var(\varepsilon_i) = \sigma_i^2$ ile gösterilir. Regresyon çözümlemesinde tümünün σ^2 gibi aynı varyansla dağılmaları istenir. $Var(\varepsilon_i) = Var(Y_i) = \sigma^2$ dir. Bu varsayım (homoskedasticity) sabit varyanslılık adını alır[32].

Varsayım-6- Hata Terimiyle X Bağımsız Değişkenin Bağımsız Olması Varsayımı

X_i bağımsız değişkenle hata terimi ε_i arasında ilişki yoktur, yani kovaryansları sıfıra eşittir.

$$Kov(\varepsilon_i, X_i) = 0$$

X rassal değişken ise varsayım (2) geçerli ise, ki bu durumda $Kov(\varepsilon_i, X_i) = [X_i - E(X_i)]E[\varepsilon_i - E(\varepsilon_i)] = 0$'dır. Dolayısıyla ε ve X arasındaki bağımsızlık varsayımı açıkça gösterilmiş olmaktadır[33].

Varsayım-7- Bağımsız Değişken X'in Değerlerinin Sabit Olması Varsayımı

X_i ile ε_i arasında ilişki olmaması yani $Kov(\varepsilon_i, X_i) = 0$ varsayımı X'in stokastik bir değişken olmamasını gerektirir. Diğer bir ifadeyle istatistiksel olarak anakütleden çekilebilecek tüm örnekler için X_i değerleri aynıdır (sabittir), fakat ε_i ve dolayısıyla Y_i değerleri farklı rassal sebeplerle örneklemden örnekleme değişir. Örneğin, yapılan bir gelir-tüketim harcamaları anketinde X_1, X_2, ..., X_k gibi gelir düzeyleri belirlendiği ve bu gelir düzeylerindeki tüketim harcamaları Y'lerin saptandığı varsayılsın. X'ler değişmemektedir, yani bir sabit değerler kümesidir, diğer taraftan her

---

[32] Şıklar, **a.g.e.,** s.9-10.

[33] **Aynı**, s.10-11.

hane halkının tüketim düzeylerini temsil eden Y_i'ler farklı rassal etmenler nedeniyle aynı gelir düzeyine sahip hane halkları arasında değişmektedir. Dolayısıyla X'lerin sabit olduğu veri iken

$$Kov(\varepsilon_i X_i) = E[(\varepsilon_i - E(\varepsilon_i))(X_i - E(X_i))]$$

Varsayım 2'ye göre $E(\varepsilon_i) = 0$ olduğundan;

$$Kov(\varepsilon_i, X_i) = E[\varepsilon_i(X_i - E(X_i))]$$

$$= E[\varepsilon_i X_i - \varepsilon_i E(X_i)]$$

ve

$$E[E(X_i)] = E(X_i)$$

$$Kov(\varepsilon_i X_i) = E(\varepsilon_i X_i) - E(\varepsilon_i)\, E(X_i)$$

$$= E(\varepsilon_i X_i)$$

$$= 0 \text{ (Varsayım-6 gereği)}$$

şeklinde gösterilebilir[34].

Varsayım-8- X Bağımsız Değişkenin Ölçme Hatalı Olmaması Varsayımı

Bağımsız değişken X ölçme hatası içerdiğinde daha önce açıklanan hata terimiyle olan bağımsızlık varsayımının gerçekleşmemesine neden olacaktır.

Bu varsayımdan sapmalar uygulamalarda sık sık görülmektedir. Örneğin; belli bir olayla ilgili olarak doğrudan veri olmaması halinde ona

[34] Şıklar, **a.g.e.**, s.11.

yakın bağımsız bir değişken alınması, indekslerin bağımsız değişken olarak alınması halinde ölçme hatası içerirler[35].

Varsayım-9- Model Belirleme Hatasının Olmaması Varsayımı

Doğrusal regresyon modelinin varsayımlarından birisi de modelin doğru belirlenmesidir. Bu varsayımın uygulamada sağlanması oldukça güçtür. Model belirlenmesinde;

- Modele hangi değişkenler alınacaktır?
- Modelin fonksiyonel şekli nasıldır?
- Y, X ve ε hakkındaki varsayımlar hangileridir?

Soruları yanıtlanmalıdır. Modele alınması gereken değişkenlerin alınmaması, üstel fonksiyon alınması gibi hatalar yapılması modelin yanlış belirlenmesine neden olacaktır[36].

Varsayım-10- Bağımsız Değişkenler Arasında İlişki Olmaması Varsayımı

Birden fazla bağımsız değişken olan çoklu regresyon için geçerli olan bu varsayıma göre bağımsız değişkenler arasında ilişki yoktur[37].

Lojistik regresyon analizinde normal dağılım ve süreklilik varsayımı önkoşulu yoktur. Bağımlı değişken üzerinde açıklayıcı değişkenlerin etkileri olasılık olarak elde edilir[38].

---

[35] Şıklar, **a.g.e.**, s.12.

[36] **Aynı**, s.12.

[37] **Aynı**, s.12.

[38] Özdamar, **a.g.e.**, s.461.

4. REGRESYON KULLANIM ALANLARI

Regresyon modelleri stokastik ve bağımlı-bağımsız değişkenler arasındaki ilişkiyi örnekleyen modellerdir.

Bir regresyon modelinin genelde kullanım amaçları şunlardır:

1) Salt Tanımlama

2) Parametre Kestirimi

3) Ön Kestirim

4) Kontrol[39]

Salt tanımlama; bağımlı değişkeni en iyi tanımlayacak etkin bağımsız değişkenlerin modele alınmasıdır.

Regresyon modelinin kullanılmasında en önemli amaç parametre kestirimleridir. Parametre kestirimleri ise, bağımlı değişkenleri etkileyen değişkenlerin etkilerinin kestirimleri olan regresyon katsayılarıdır. Sıfırdan farklı değere sahip regresyon katsayıları ilk anda bağımlı değişkeni etkileyen değişkenlerin var olduğunu gösterir.

Regresyon modellerinde parametrelerin kestirimleri ön kestirim amacı ile de yapılır. Bağımlı değişken değerlerinin önceden kestirimleri elde edilir. Bu değerler ya verilerin kullanıldığı bölgenin içerisinde bulunan herhangi bir gözlem içindir ya da dışındaki bir gözlem için bulunan değerlerdir.

---

[39] Şıklar, **a.g.e.**, s.3.

Regresyon modelleri kontrol amacıyla da kullanılabilir. Örneğin; bir kimya mühendisi kağıdın dayanıklılık düzeyini sıkıştırma aşamasındaki tahtanın konsantrasyonu ile ilişkilendirilen bir model geliştirmek için regresyon analizini kullanabilir. Elde edilen eşitlik çeşitli konsantrasyon düzeyleri için uygunluk dayanıklılık düzeyini kontrol etmek amacı ile kullanıldığı zaman değişkenlerin bir nedensellik bağlamında ilişkilendirilmesi önemli bir husustur[40].

Gözlemlenmiş bir oran veya orantı ile bir açıklayıcı değişkenler kümesi arasındaki ilişkinin inceleneceği regresyon analizi ise bu çalışmada kullanılan "Lojistik Regresyon Analizi" dir.

---

[40] Şıklar, **a.g.e.**, s.4.

İKİNCİ BÖLÜM

LOJİSTİK REGRESYON ANALİZİ

Lojistik regresyon analizinin amacı, bir bağımlı (sonuç veya tepki değişkeni) ile bir bağımsız (tahmin edici veya açıklayıcı) değişkenler kümesi arasındaki ilişkiyi tanımlamak için en uygun modeli bulmaktır[41].

Lojistik regresyon analizinin açıklanmasında kullanılan kavram, gösterim ve modeller aşağıda açıklanmıştır.

1. LOJİSTİK REGRESYON KAVRAMI

Lojistik regresyon, önce sağlık alanında kullanılmaya başlandı. Daha sonra bu yöntem, yaygın kabulü sonucunda başka alanlardada kullanıldı. Bu yöntemin uygulama biçimleri konuyla ilgili kaynaklarda tanımlanmıştır. Örneğin; White, Pearson ve Wilson (1999) lojistik regresyon modellerini "Just in Time" çalışmalarının tanımlanmasında incelemişlerdir.

Palma, Beja ve Rodrigues (1999) "Vaşak" hayvanının görünümünü model almıştır ve özellikle çağdaş bir uygulamada Foong, Hu ve Heisey (1999) lojistik regresyonu internette nesnelerin gizli değerlerini tahmin

---

[41] www.wiley.co.uk/eob/sample4.pdf, s.1, (24/04/2003).

etmek için kullanmıştır[42]. Uygulamanın yer aldığı ana bölümler ekonometri, biyoistatistik ve eğitimsel testlerdir[43].

Lojistik Regresyon; bağımlı değişkenin ikili, üçlü ve çoklu sınıflar halinde gözlendiği durumlarda, açıklayıcı değişkenlerle neden sonuç ilişkisini belirlemede yararlanılan bir yöntemdir. Açıklayıcı değişkenlere göre, tepki değişkenin beklenen değerinin, olasılık olarak elde edildiği bir regresyon yöntemidir[44].

2. ODDS DEĞERİ

İncelenen bir olayın olasılığının kendi dışında kalan diğer olayların olasılığına oranına ODDS değeri denir ve

$$\mathrm{ODDS_P} = \frac{P}{1-P} \tag{5}$$

şeklinde gösterilir. Burada P incelenen olayın olasılığını göstermektedir. ODDS değeri 0 ile $+\infty$ arasında değerler alır[45].

3. ODDS ORANI

İncelenen iki farklı olayın ODDS değerlerinin birbirine oranına ODDS oranı denir. ODDS oranı incelenen iki olayın gözlenme olasılıklarından birinin diğerine oranla kaç kat daha fazla veya kaç kat daha az olarak ortaya çıkabileceğini gösterir.

---

[42] Thomas P. Ryan, "Some Issues In Logistic Regression" **Communications In Statistics Theory And Methods.** (Vol:29, No:9 &10, 2000), s.2020.

[43] James M. Landwehr, Darly Pregıbon and Anne C. Shoemaker, "Graphical Methods For Assessing Logistic Regression Models", **Journal Of The American Statistical Association**, (March 1984), s.61.

[44] Özdamar , **a.g.e.**, s.461.

[45] Çolak, **a.g.e.**, s.8.

Eğer incelenen bir A olayının E kümesi içinde ortaya çıkma olasılığı;

$$P(A / E) \tag{6}$$

ile gösterilir ise, A olayının E kümesi içindeki ODDS değeri;

$$ODDS_{P(A/E)} = \frac{P(A/E)}{1-P(A/E)} = \frac{P(A/E)}{P(\bar{A}/E)} \tag{7}$$

şeklinde ifade edilir. İncelenen A olayının E kümesi dışında ortaya çıkma olasılığının ODDS değeri ise;

$$ODDS_{P(A/\bar{E})} = \frac{P(A/\bar{E})}{1-P(A/\bar{E})} = \frac{P(A/\bar{E})}{P(\bar{A}/\bar{E})} \tag{8}$$

şeklinde ifade edilir. Buradan, bu iki ODDS değeri birbirlerine oranlanır ise;

$$ODDS\,ORANI = \frac{P(A/E)/P(\bar{A}/E)}{P(A/\bar{E})/P(\bar{A}/\bar{E})} \tag{9}$$

ifadesi elde edilir. ODDS oranı 0 ile $+\infty$ arasında değerler alır. Eğer bu oran 1'den büyük çıkarsa, incelenen A olayının, E kümesi içinde ortaya çıkma olasılığının E kümesi dışında gözlenme olasılıklarına göre o kadar kat artacağını, eğer 1'den küçük çıkarsa, incelenen A olayının, E kümesi içinde ortaya çıkma olasılığının E kümesi dışında gözlenme olasılığına göre o kadar kat azalacağını gösterir[46].

[46] Çolak, **a.g.e.**, s.8-9.

4. LOJİT FONKSİYON

Lojit fonksiyon, incelenen bir olasılığın (P), ODDS değerinin doğal logaritmasını verir. İncelenen olasılığın (P) lojit fonksiyonunda ki gösterimi;

$$\mathrm{Logit}[P] = \ln\left(\frac{P}{1-P}\right) = \ln\left(\mathrm{ODDS}_P\right) \qquad (10)$$

şeklindedir.

İncelenen olasılığın ODDS değeri 0 ile $+\infty$ arasında değer alırken aynı olasılığın lojit değeri $-\infty$ ile $+\infty$ arasında değerler alır[47].

5. LOJİT DÖNÜŞÜME İLİŞKİN ÖZELLİKLER

Lojit(P)=ln(P/(1-P)) lojit dönüşümünün bazı özellikleri şöyle sıralanabilmektedir.

-P arttıkça, lojit (P) de artar.

-P, 0-1 arasında iken lojit (P) tüm gerçel değerleri alır.

-Eğer P<0,5 ise lojit (P)<0 dır.

-Eğer P>0,5 ise lojit (P) >0 dır[48].

---

[47] Çolak, **a.g.e.**, s.9.

[48] Hüseyin Tatlıdil, **Uygulamalı Çok Değişkenli İstatistiksel Analiz,** (Hacettepe Üniversitesi, 1996). s.292-293.

6. DOĞRUSAL MODEL VE LOJİSTİK REGRESYON

Olasılıkların lojit fonksiyonunun kullanılmasının amacı, doğrusal bir model elde edilerek, parametre tahminlerinin yapılmasıdır. Doğrusal modeli;

$$\beta_0+\beta_1X_1+.....+\beta_kX_k \qquad (11)$$

şeklinde gösterirsek ve incelenen bir olasılığın (P) lojit değerini bu doğrusal modele eşitlersek;

$$\text{Logit}[P]=\ln\left(\frac{P}{1-P}\right)=\beta_0+\beta_1X_1+\beta_2X_2+...+\beta_kX_k \qquad (12)$$

eşitliğini elde ederiz. Elde ettiğimiz bu eşitlik, basit ve çoklu doğrusal regresyon modellerindeki bağımlı değişkenin beklenen değerini veren ve parametre tahminlerinde kullanılan eşitliğe benzer bir eşitliktir. Çünkü $E(Y/X_1,X_2,...,X_k)$ değeri ile Lojit[P] değeri $-\infty$ ile $+\infty$ arasında değerler almaktadır. Bu eşitlikten incelenen olasılık (P);

$$P=\frac{e^{\beta_0+\beta_1X_1+...+\beta_kX_k}}{1+e^{\beta_0+\beta_1X_1+...+\beta_kX_k}} \qquad (13)$$

şeklinde elde edilir. Bu eşitliğe lojistik regresyon modeli denir.

Burada;

P : İncelenen olayın gözlenme olasılığını,

β_0 : Bağımsız değişkenler sıfır değerini aldığında bağımlı değişkenin değerini, yani sabiti,

$\beta_1, \beta_2, \ldots, \beta_k$: Bağımsız değişkenlerin regresyon katsayılarını,

$X_1, X_2, \ldots, X_k$: Bağımsız değişkenleri,

k : Bağımsız değişken sayısını,

e = 2,71 sayısını göstermektedir[49].

7. İKİLİ LOJİSTİK REGRESYON MODELİ (BINARY LOGISTIC REGRESSION)

Bağımlı değişkenin ikili kategorik cevap içerdiği lojistik regresyon modelidir. Bağımlı değişken var-yok, hasta-sağlam, başarılı-başarısız gibi değerler alır. Bir ya da daha fazla bağımsız değişken ile ikili cevap değişken arasındaki bağıntıyı ortaya koyar. Bağımsız değişkenler ya faktör değişken ya da ortak değişkenlerdir. Bağımlı değişkende incelediğimiz kategori genel olarak Y=1 ile kodlanır. Diğer kategoride Y=0 ile ifade edilir. Bu durumda incelediğimiz kategorinin olasılık değerini bağımsız değişkenlerle analiz eden ikili lojistik regresyon modeli;

$$P(Y_j = 1) = \frac{e^{\beta_0 + \beta_1 X_{j1} + \ldots + \beta_k X_{jk}}}{1 + e^{\beta_0 + \beta_1 X_{j1} + \ldots + \beta_k X_{jk}}} \tag{14}$$

şeklinde ifade edilebilir.

Burada;

[49] Çolak, **a.g.e.**, s.9-10.

n : Birim sayısını,

j : 1,2,...,n

P(Y_j=1) : j. birimin incelenen kategoriye eşit olma olasılığını ya da incelenen olay ile ilgili pozitif cevap verme olasılığını,

β_0 : Bağımsız değişkenler sıfır değerini aldığında bağımlı değişkenin değerini, yani sabiti,

$\beta_1, \beta_2, ..., \beta_k$: Bağımsız değişkenlerin regresyon katsayılarını,

$X_{j1}, X_{j2}, ..., X_{jk}$: j.birime ait bağımsız değişkenleri,

k : Bağımsız değişken sayısını,

e=2,71 sayısını göstermektedir[50].

8. İKİ GRUP LOJİSTİK MODELE İLİŞKİN VARSAYIMLAR

Lojistik model, "Genelleştirilmiş Doğrusal Modeller" olarak bilinen çok geniş model ailesinin bir üyesidir. Lojistik model ikili tepki (sonuç) verileri için varsayıldığında modele ilişkin varsayımlar şöyle sıralanabilmektedir.

-Açıklayıcı değişkenler (x_k) birbirinden bağımsızdır.

-y_1,..........y_n değerleri istatistiksel olarak bağımsızdır.

-$y_j \in (0,1)$

[50] Çolak, **a.g.e.**, s.10-11.

$-P(y_j=1 / x_j)=P_j$ [51].

9. KATSAYI KESTİRİM YÖNTEMLERİ

9.1. En Çok Olabilirlik Kestirim Yöntemi

Lojistik regresyon modeli (14)'ü bir veri kümesine uyarlamak için, bilinmeyen parametrelerin değerlerini tahmin etmemiz gerekir[52].

Lojistik modelde gözlem sayısı yeteri kadar büyük değilse (n<30) elde edilen tahminlere güvenilmez[53].

En çok olabilirlik yönteminin yaygın kabulü, bir yandan bunun teorik özellikleriyle ilgilenen büyük miktardaki araştırma, bir yandan hemen hemen sınırsız uygulamaların listesi üzerindeki görülmektedir[54].

En çok olabilirlik yöntemi lojistik regresyon parametrelerini tahmin için kullanılan yöntemdir. Çok genel bir anlamda en çok olabilirlik yöntemi gözlenen veri kümesini elde etme olasılığını en büyükleyen bilinmeyen parametreler için değerler meydana getirir. Bu yöntemi uygulamak için ilk önce olabilirlik fonksiyonu olarak adlandırılan bir fonksiyon oluşturulur. Bu fonksiyon gözlemlenen verilerin olasılığını bilinmeyen parametrelerin bir fonksiyonu olarak ifade eder. Bu parametrelerin en çok olabilirlik tahmin edicileri bu fonksiyonu en büyükleyen değerler olarak seçilir. Bu yüzden

---

[51] Tatlıdil, **a.g.e.**, s.292-293.

[52] www.wiley.co.uk/eob/sample4.pdf, s.1, (24/04/2003).

[53] Şahin Akkaya ve Vedat Pazarlıoğlu, **Ekonometri 2,** (İstanbul: 1998), s.90.

[54] Samuel Kotz and Norman L. Johnson, **Encyclopedia Of Statistical Sciences,** (Newyork: J. Wiley & Sons, Vol:5, 1982), s.340.

sonuç tahmin ediciler gözlemlenen verilerle en yakın uyum içinde olanlardır[55].

Eğer Y, 0 veya 1 olarak kodlanırsa, (14) de verilen $P(Y_j=1)$ için ifade ($\beta'=(\beta_0,\beta_1,\ldots\beta_k)$) nın rasgele bir değeri için (parametrelerin bir vektörü) belirli bir x için Y nin 1 e eşit olma şartlı olasılığını sağlar. Bu P(Y=1/x) olarak ifade edilecektir.

Buradan 1-P(Y=1) niceliğinin belirli x için 0 a eşit olma şartlı olasılığı P(Y=0/x) i verir. Böylece, $y_j=1$ olan bu (x_j,y_j) çiftleri için olabilirlik fonksiyonuna katkı $P(Y_j=1)$ dir ve $y_j=0$ olan çiftler için olabilirlik fonksiyonuna katkı $1-P(Y_j=1)$ dir. Burada $P(Y_j=1)$ niceliği x_j de hesaplanan P(Y=1) in değerini ifade eder.

(x_j,y_j) çifti için olabilirlik fonksiyonuna katkıyı ifade etmenin uygun bir yolu şu terimle gerçekleştirilir;

$$\xi(x_j)=P_j^{y_j}[1-P_j]^{1-y_j} \qquad (15)$$

Gözlemlerin bağımsız olduğu varsayıldığından olabilirlik fonksiyonu aşağıdaki gibi (15) de verilen terimlerin çarpımı olarak elde edilir.

$$L(\beta)=\prod_{j=1}^{n}\xi(x_j) \qquad (16)$$

En çok olabilirliğin ilkesi β nın bizim tahminimiz olarak (16)'daki ifadeyi en büyükleyen değeri kullanmamızdır. Bununla beraber (16)'nın logaritmasıyla çalışmak matematiksel olarak daha kolaydır. Bu ifade, yani log olabilirlik şu şekilde tanımlanabilir.

[55] www.wiley.co.uk/eob/sample4.pdf, s.2, (24/04/2003).

$$\ln [L(\beta)] = \sum_{j=1}^{n} \{ y_j \ln P_j + (1-y_j) \ln [1-P_j] \} \qquad (17)$$

In [L(β)]'yi en büyükleyen β'nın değerini bulmak için β_0 ve β_1 e göre In [L(β)]'nin türevini alırız ve sonuç ifadeleri 0'a eşitleriz. Bu denklemler aşağıdaki gibidir;

$$\sum_{j=1}^{n} [y_j - P_j] = 0 \qquad (18)$$

ve

$$\sum_{j=1}^{n} x_j [y_j - P_j] = 0 \qquad (19)$$

ve olabilirlik denklemleri olarak adlandırılır.

Doğrusal regresyonda, sapma(hata payı, bağımlı değişkenin gerçek değeri ile tahmin edilen değeri arasındaki fark)ların kareleri fonksiyonunun toplamının β ya göre türevi alınarak elde edilen olabilirlik denklemleri bilinmeyen parametrelerde doğrusaldır ve bu yüzden kolaylıkla çözülürler. Lojistik regresyon için (18) ve (19) daki ifadeler β_0 ve β_1 de doğrusal değildir ve bu yüzden çözümleri için özel yöntemler gerektirirler. Bu yöntemler doğasında tekrarlamalıdır ve uygun logistik regresyon yazılımında programlanmaktadır. McCullagh ve Nelder pek çok program tarafından kullanılan tekrarlı yöntemleri anlatmaktadırlar.

(18) ve (19) un çözümleriyle verilen β nın değeri $\hat{\beta}$ olarak yazılır. En çok olabilirlik tahmini şeklinde adlandırılır[56].

---

[56] www.wiley.co.uk/eob/sample4.pdf, s.2, (24/04/2003).

Benzer şekilde;

$\hat{P}(y_j=1)$, $P(y_j=1)$ nin en çok olabilirlik tahminidir. Bu nicelik, x'in x_j'ye eşit olması şartıyla y'nin 1'e eşit olma şartlı olasılığının bir tahminini sağlamaktadır. Böylece bu lojistik regresyon model için uyarlanmış veya öngörülmüş değeri temsil etmektedir.

(18)'in ilginç bir sonucu $\sum_{j=1}^{n} y_j = \sum_{j=1}^{n} \hat{P}(y_j = 1)$ dir.

Yani y'nin gözlemlenen değerlerinin toplamı öngörülen (beklenen) değerlerin toplamına eşittir[57].

En çok olabilirlik tahmin yönteminin el ile başarılması oldukça güçtür. Yineleme (iterasyon) uygulayarak tahminlerin en iyi olmasının sağlanması gerektiğinden hesaplamalar el ile yapılamaz duruma gelmektedir[58].

Katsayıları tahmin ettikten sonra modeldeki değişkenlerin önemine erişmek standart bir uygulamadır. Bu genellikle modeldeki bağımsız değişkenlerin "önemli derecede" tepki değişkeniyle ilişkili olup olmadığını belirlemede istatistiksel bir hipotezi test etmeyi içerir[59].

---

57 www.wiley.co.uk/eob/sample4.pdf, s.2, (24/04/2003).

58 Kazım Özdamar, "Biyolojik Denemelerin Oransal Cevaplarının Analizinde Probit ve Lojit Regresyon Yöntemlerinin Etkinliğinin Araştırılması" **Anadolu Tıp Dergisi**,(1988), s.352.

59 www.wiley.co.uk/eob/sample4.pdf, s.2, (24/04/2003).

$\hat{\beta}_{m,t+1}$: m'inci bağımsız değişkenin (t+1)'inci iterasyondaki tahmin değerini,

$\hat{\beta}_{m,t}$: m'inci bağımsız değişkenin t'inci iterasyondaki tahmin değerini,

ε : Tolerans seviyesini göstermektedir.

Tolerans seviyesi 0,0001'e eşit ya da daha küçük bir sayı olabilir. Sonuçta, (t+1)'inci iterasyonda hesaplanan tahmin değerleri ile t'inci iterasyonda hesaplanan tahmin değerleri arasındaki mutlak fark 0,0001'den daha küçük ise iterasyon durdurulmuştur ve (t+1)'inci iterasyondaki tahmin değerleri kullanılmıştır[61].

9.3. Fisher Skorlaması

Nelder ve Wedderburn (1972); genelleştirilmiş modellerde $\hat{\beta}$'nın sayısal çözümlemesi için genel bir teknik olarak Fisher Skorlaması'nı önermektedirler[62].

Fisher skorlama yönteminde, Newton Raphson eşitliğinde ikinci derece türevler matrisi yerine beklenen değerler kullanılır.

$$\hat{\beta}_{t+1} = \hat{\beta}_t + \left\{ E\left(-\frac{\partial^2 \log\{L(\beta)\}}{\partial\beta\, \partial\beta} \right) \right\}^{-1} \Bigg|_{\beta=\beta_t} \frac{\partial \log\{L(\beta)\}}{\partial\beta} \Bigg|_{\beta=\beta_t} \qquad (22)$$

---

[61] Çolak, **a.g.e.**, s.30.

[62] Peter J. Green, **Nonparametric Regression and Generalized Linear Models,** (London: Chapman & Hall, 1994), s.94.

Sonuç olarak kestirilen katsayılara ilişkin kovaryans matrisinin beklenen değerinin, kendisine eşit olması nedeniyle bu iki yöntem aynı olur[63].

Çoğu istatistiki paket ve alt programlar bu hesaplama için ihtiyaç duyulan temel yöntemleri sağlar[64].

10. PARAMETRELERİN ÖNEMLİLİĞİNİN VE MODELİN UYGUNLUĞUNUN TEST EDİLMESİ

Tahmin edilen β katsayılarının önemliliğinin test edilmesi H_0: β=0 hipotezinin test edilmesidir. Bu amaçla ileri sürülen testler;

1. Wald testi

2. Benzerlik Oranı (Likelihood Ratio) testidir. Benzerlik oranı testi LR ile gösterilir.

Bu testler asimptotik olarak aynı sonucu verirler. Örnek hacmi büyüdükçe testlerin sonuçları çok yakın çıkmaktadır[65].

10.1. Wald Testi

Wald testi en çok olabilirlik tahminlerinin (Maksimum Likehood Estimations) asimptotik olarak normal dağılım gösterdiği varsayımına dayanır. Regresyon katsayılarının, standart hatasına oranı;

---

[63] Tatlıdil, **a.g.e.**, s.308.

[64] Green, **a.g.e.**, s.95.

[65] Çolak, **a.g.e.**, s.16.

$$z = \frac{\beta}{SE_\beta} \tag{23}$$

Wald istatistiği olarak adlandırılır. Bu durumda Wald istatistiği Standart Normal Dağılım (SND) gösterir ve SND'nin kritik değerleri ile karşılaştırılarak önemliliği belirlenir. Aynı zamanda Wald istatistiği;

$$w = z^2 = \left(\frac{\beta}{SE_\beta}\right)^2 \tag{24}$$

olarak da kullanılabilir. Bu durumda ise Wald istatistiği 1 serbestlik dereceli ki-kare dağılımı gösterir ve 1 serbestlik dereceli ki-kare dağılımının kritik değerleri ile karşılaştırılarak önemliliği belirlenir[66].

10.2. Benzerlik Oranı Testi

X_1, X_2,.., X_k tane bağımsız değişken içeren bir regresyon analizinde, oluşturulan iki regresyon modelinin birinci modelinde X_1, X_2, ..., X_v tane bağımsız değişken, ikinci modelinde ise X_1, X_2, ..., X_v, X_{v+1}, X_{v+2}, ..., X_{v+m} (v+m=k) olmak üzere tüm bağımsız değişkenler olsun. Bu iki modelde de bağımsız değişkenlerin katsayılarını birinci model için β_v vektörü, ikinci model için de β_{v+m} vektörü göstersin. X_{v+1}, X_{v+2}, ..., X_{v+m} bağımsız değişkenlerinin katsayılarını sıfıra eşit olup olmadığını eşanlı (simültane) olarak test eden benzerlik oran test istatistiği;

$$LR = -2\left(\log\left[\frac{L(\beta_v)}{L(\beta_{v+m})}\right]\right) = -2(\log[L(\beta_v)] - \log[L(\beta_{v+m})]) \tag{25}$$

şeklinde ifade edilir.

[66] Çolak, **a.g.e.**, s.16-17.

Burada;

$L(\beta_v)$: v tane bağımsız değişken içeren birinci model için en çok olabilirlik fonksiyonu,

$L(\beta_{v+m})$: v+m tane bağımsız değişken içeren ikinci model için en çok olabilirlik fonksiyonunu göstermektedir.

Benzerlik oran istatistiği, (v+m)-v=m serbestlik dereceli ki-kare dağılımı gösterir ve m serbestlik dereceli ki-kare dağılımının kritik değerleri ile karşılaştırılarak β_{v+m} vektörünün en son m tane katsayısının sıfıra eşit olup olmadığı test edilir.

Verilere uygulanan modelin uygunluğunu test etmek için model parametrelerinin tümünün sıfır olduğu hipotezi test edilir. Burada uygulanan test benzerlik oranı testidir. Birinci model olarak sabit dışında (β_0) hiçbir parametre modele katılmaz. İkinci model olarak ise tüm parametreler modele katılır. Birinci ve ikinci model için hesaplanan en çok olabilirlik fonksiyonları elde edilerek benzerlik oranı LR test istatistiği yukarıdaki gibi hesaplanır. Uygunluk testi için hesaplanan benzerlik oranı test istatistiği toplam parametre sayısının 1 eksiği kadar ya da modele katılan bağımsız değişken sayısı kadar serbestlik dereceli ki-kare dağılımını gösterir ve ki-kare dağılımının kritik değerleri ile karşılaştırılarak modelin uygunluğu test edilir[67] .

[67] Çolak, **a.g.e.**, s.17-18.

ÜÇÜNCÜ BÖLÜM

LOJİSTİK REGRESYON ANALİZİNİN SİGORTA ALANINDA KULLANILMASINA İLİŞKİN BİR UYGULAMA

Bu bölümde ikili veri yapısına uygulanan lojistik regresyon yöntemi incelenmiştir. Uygulamada lojistik regresyon analizinin ikili tepki verileri için incelenmesi amacıyla, Axa Oyak Sigorta A. Ş. Acenteliği bilgisayarından alınan, 2002-2003 yılında trafik sigortası olan 100 müşterinin otomobilleri için kasko sigortası var ve yok şeklinde sonuçlanan, aracın yaşı bireyin yaşı ve 2002 ocak ortalama aylık gelir verileri kullanılmıştır.

1. UYGULAMAYA İLİŞKİN TANIM, KAVRAMLAR VE GENEL BİLGİ

Riziko (Risk): Sigortalının teminat altına alınmasını istediği olayın sigortalıya zarar verme ihtimalidir[68].

Prim: Sigorta şirketinin üstlendiği riske karşılık, sigorta ettiren'in ödediği güvence satın alma fiyatı ya da ücretidir[69].

---

[68] Temel Sigorta Bilgileri, **a.g.e.**, s.3.

[69] Şebnem Duman, **Sigorta Tekniği**, (Eskişehir: Anadolu Üniversitesi Yayın no:816, Aralık 1995), s.42.

Sigortalı (Sigorta Ettiren): Taşıdığı riski belli bir prim karşılığında sigortacıya devreden özel yada tüzel kişiye sigortalı yada sigorta ettiren denir[70].

Sigortacı: Sigortalı adı verilen şahsın, sigorta konusu olan unsur üzerindeki menfaatini sigortalamaktadır.

Sigorta Sözleşmesi (Poliçe): Sigortacının sigortalıya vermek zorunda olduğu bir belgedir.

Bir sigorta sözleşmesinde sigortacı, sigorta ettiren, sigorta konusu olan menfaat ve özellikleri, azami tazminat, sigortanın geçerli olduğu tarih, risk ve prim unsurları yer almaktadır[71].

Sigorta: Sigortanın çeşitli tanımları yapılmıştır. Bunlardan bazıları aşağıda verilmiştir:

Sigorta, kişilerin canlarına, sağlıklarına ve sahip oldukları ekonomik değerlerin taşıdığı risklere karşı, teminat veren, ekonomik riskleri kendi üyeleri arasında paylaştırma, olabilecek zarar ve ziyanlara karşı önlem alınmasına olanak sağlama faaliyetidir[72].

Sigorta, aynı yada benzer rizikolara maruz bulunan kişiler topluluğunda rizikonun gerçekleşmesi sonucunda ortaya çıkacak ihtiyacın, belli bir prim karşılığında giderilmesine yönelik bağımsız bir hukuki talebe sahip olmayı temin eden kurum ve sözleşmedir[73].

---

[70] Duman, **a.g.e.**, s.31.

[71] **Aynı,** s.30.

[72] Özkan, **a.g.e.,** s.v.

[73] **Aynı,** s.10.

Sigorta, sigortacının belirli bir prim karşılığında diğer bir kimsenin para ile ölçülebilen bir menfaatini zarara uğratan bir tehlikenin gerçekleşmesi halinde tazminat ödemeyi taahhüt etmesidir[74].

Sigortanın teknik tanımı ise şöyle yapılmıştır:

Sigorta, herhangi bir yangın, kaza, ölüm ve benzeri gibi doğal felaketler sonucunda hasara uğrayan bina, eşya, mal veya can'dan dolayı zarar gören sigortalının zararının, sigortacı tarafından tazminini amaçlayan bir akittir[75].

Acente: Her ne ad altında olursa olsun sigorta şirketine tabii bir sıfatı olmaksızın bir sözleşmeye dayanarak belirli bir yer veya bölge içinde, daimi suretle Türkiye'deki sigorta şirketinin sigorta sözleşmelerine aracılık eden veya bunları sigorta şirketi adına yapan gerçek veya tüzel kişilere sigorta acentesi, sigorta acentelerinin acentelik faaliyetleriyle ilgili olarak tayin ettikleri gerçek veya tüzel kişi acentelere ise tali acente denir[76].

Kasko Sigortası: Sadece sigortalı taşıtın uğrayacağı hasarlar teminat altına alınmış olup sigortalı taşıtın 3. kişilere verebileceği zararlardan dolayı sigorta ettirenin sorumlu bulunduğu tazminatı ödemekle yükümlü olmayan bir sigorta türüdür[77].

Bu sigorta ile sigortacı, sigortalının, karayolunda kullanılabilen motorlu, motorsuz taşıtlardan römork veya karavanlar ile iş makinelerinden

---

[74] Özkan, **a.g.e.**, s.10.

[75] Cevat Yücesoy, **Sigorta İşletmeleri ve Muhasebesi**, (İstanbul: Çağlayan Muhasebe Serisi No:8, Çağlayan Kitabevi B.1, 1966), s.1.

[76] Duman, **a.g.e.**, s.32.

[77] Ergun Orhunöz, **Uygulamada Karayolları Trafik Kanununa Göre Sorumluluk Tazminat Sigorta,** (Ankara: Seçkin Yayınevi,1998), s.250.

ve lastik tekerlekli traktörlerden doğan menfaatinin aşağıdaki tehlikeler dolayısıyla ihlali sonucu uğrayacağı maddi zararları temin eder.

-Gerek hareket gerek durma halinde iken sigortalının veya aracı kullananın iradesi dışında araca ani ve harici etkiler neticesinde sabit veya hareketli bir cisim çarpması veya aracın böyle bir cisme çarpması, müsademesi, devrilmesi, yuvarlanması gibi kazalar ile üçüncü kişilerin kötü niyet veya muziplikle yaptıkları hareketler,

-Aracın yanması,

-Aracın çalınması veya çalınmaya teşebbüs[78].

Kasko sigortası ile sigortacı, sigorta ettirene, sigortalı motorlu kara nakil vasıtasının tamir masraflarını ya da bu vasıta yerine yenisinin alınmasını sağlamak bakımından sigorta bedelini ödemeyi taahhüt etmektedir[79].

O halde kasko sigortasının şu şekilde tanımının yapılması yerinde olacaktır:

Sigorta ettirenin ödeyeceği sigorta primi karşılığında, aracın maruz kalacağı kasko rizikolarının, sigortacı tarafından teminat altına alınmasını öngören sigorta sözleşmesine, kasko sigortası denir[80].

---

[78] Hayri Taşyürek, **Kasko Sigortası,** (1. Baskı. Ankara: Seçkin Yayıncılık, 2001), s.188.

[79] **Aynı**, s.21.

[80] **Aynı**, s.22.

Aracın Kullanıcısı Ve İşleticisi: Araçta fiili hakimiyeti olan ve aracı yararı, masrafları, rizikosu kendisine ait olmak üzere işleten kimse olarak tanımlanmaktadır[81].

Trafik Sigortası (Mecburi Mali Mesuliyet Sigortası): Motorlu bir aracın karayolunda işletilmesi sırasında, bir kimsenin ölümüne veya yaralanmasına veya bir şeyin zarara uğramasına neden olması halinde o aracı işletenin zarara uğrayan üçüncü kişilere karşı olan (hukuki) sorumluluğunu belli limitler dahilinde karşılamayı amaçlayan ve yasaca yapılması mecburi kılınan sorumluluk sigortası türüdür[82]. Bu tanımından da anlaşılacağı üzere, mecburi mali mesuliyet sigortası bir motorlu aracın kullanımı sırasında, trafik sonucu oluşan ölüm yada yaralanma veya birşeyin zarara uğratılması gibi hallerde (maddi ve bedensel dahil zararların olması halleri) hukuki zorunluluğu teminata alır.

Kasko sigortasından farklı yanı, motorlu aracın kullanım sonucu üçüncü kişilerde oluşan hasar ve zararların teminata alınmasıdır. Ve bu yönden de mecburi bir mali mesuliyet olmaktadır[83]. Kasko sigortası aracın kendisinde meydana gelen hasarları temin eder. Trafik sigortası ise araç nedeniyle başkalarına verilen hasarlardan doğan sorumluluğu karşılamaktadır. Kasko sigortasının zorunlu olmasına rağmen bu sigortanın zorunlu olmamasının nedeni de budur[84].

İlk defa kasko sigortası poliçesi 1890 yılında yapılmıştır. Kasko sigortalarında, sigorta konusuna (=araca) kullanıcıdan daha fazla önem

---

[81] Özkan, **a.g.e.**, s.116.

[82] Taşyürek, **a.g.e.**, s.52.

[83] Özkan, **a.g.e.**, s.115-116.

[84] Temel Sigorta Bilgileri, **a.g.e.**, s.39.

verilmiştir. İlk kasko tarifesi, İngiltere'de 1920'lerde "Accident Offices Assocation" tarafından kullanılmıştır[85].

Otomobil sahiplerinin otomobillerini diğer mallarından ayrı tutmalarından dolayıdır ki, otomobil kasko sigortası sigortacılıkta ayrı bir tip olarak ortaya çıkmıştır. Bu sigortalara ait idari meseleler, şirketlerin bu branşla iştigal etmek üzere hususi servisler kurmalarını icap ettirecek düzeye gelmiştir.

Çünkü, motorlu kara nakil vasıtaları, mahiyetleri itibariyle, birçok tehlike ile karşı karşıyadır: Gerçekten bir çalınma tehlikesi söz konusu olabileceği gibi, bir yangın ihtimali her zaman mevcuttur. Bunlardan ayrı olarak, nakil vasıtaları daima bir kaza tehlikesi içinde bulunmaktadırlar. İşte birçok tehlike ile karşı karşıya kalan nakil vasıtalarını yaptırılacak bir sigorta ile, bu çeşit tehlikelerin doğuracağı zararların giderilmesi imkanı mümkün olacaktır. Motorlu kara nakil vasıtalarının ekonomik değerlerinin yüksek oluşu da dikkate alındığında bir çalınma ya da kaza anında ortaya çıkan maddi zararın yüksek oluşu, kasko sigortasının önemini göstermeye yeterlidir[86].

Böylece, kasko sigortası önemi gittikçe artan bir sigorta türü haline gelmiştir.

Axa Oyak: Lojistik regresyon analizi uygulamasına ilişkin verilerin derlendiği Axa Oyak şirketi ev, araba, sağlık, hayat, işyeri, gelecek ve kişiye özel branşlarında sigorta yapmaktadır.

Merkezi İstanbul'da bulunan Axa Oyak 5 şubat 1999 tarihinde kurulmuştur[87].

---

[85] Taşyürek, **a.g.e.**, s.17.

[86] **Aynı**, s.18.

[87] www.**axaoyak**.com.tr (24/04/2003).

2. VERİ YAPISI

Veri yapısında;

Lojistik regresyon analizinde kullanılan bağımlı değişken ikili cevap içeren kategorik değişkendir.

j : 1, 2, ..., n

n : Birey sayısı,

Bağımlı değişken;

$$Y_j = \begin{cases} 0 \text{ j.birey kasko sigortası yaptırmamış ise} \\ 1 \text{ j. birey kasko sigortası yaptırmış ise} \end{cases}$$

Bağımsız değişkenler;

1- Aracın yaşı;

X_{j1} ={ j. bireyin aracının yaşı }

2- Bireyin yaşı;

X_{j2} = { j. bireyin yaşı }

3- Bireyin aylık geliri;

X_{j3} = { j. bireyin aylık geliri}

Uygulamada aracın yaşı, bireyin yaşı ve aylık gelir ikili lojistik regresyon modelinde analiz edilmiştir.

3. İKİLİ TEPKİ VERİLERİ İÇİN LOJİSTİK REGRESYON MODELLERİNDE REGRESYON KATSAYILARININ TAHMİNLENMESİ:

İkili tepki verileri için lojistik regresyon modellerinde regresyon katsayılarının tahminlenmesinde en çok olabilirlik fonksiyonu kullanılmıştır.

İkili lojistik regresyon modeli ve parametre tahminleri aşağıdaki eşitliklerle hesaplanmıştır:

İkili lojistik regresyon modeli;

$$P(Y_j = 1) = \frac{e^{\beta_0 + \beta_1 X_{j1} + \ldots + \beta_k X_{jk}}}{1 + e^{\beta_0 + \beta_1 X_{j1} + \ldots + \beta_k X_{jk}}}$$

şeklinde ifade edilir.

Burada;

j : 1, 2, ..., n

n : Birey sayısını,

$P(Y_j = 1)$: j. bireyin incelenen olay ile ilgili pozitif cevap gösterme olasılığını,

β_0 : Bağımsız değişkenler sıfır değerini aldığında bağımlı değişkenin değerini, yani sabiti,

$\beta_1, \beta_2, ..., \beta_k$: Bağımsız değişkenlerin regresyon katsayılarını,

$X_{j1}, X_{j2}, ..., X_{jk}$: j. bireye ait bağımsız değişkenleri,

k : Bağımsız değişken sayısını,

e = 2,71 sayısını göstermektedir.

İkili lojistik regresyon modelinde regresyon katsayılarının hesaplanması için en çok olabilirlik fonksiyonu,

$$L(\beta) = \prod_{j=1}^{n} \xi(x_j)$$

şeklinde ifade edilir.

Uygulamada lojistik regresyon analizinin parametre tahminleri SPSS paket programının 10.0 versiyonu kullanılarak elde edilmiştir[88].

Parametre tahminlerinin anlamlılığı ve modelin uygunluğunun test edilmesi için Wald testinden yararlanılmıştır.

Uygulamaya ilişkin değişkenlere ait veriler tablo 4.1 de verilmiştir.

---

[88] SPSS Inc. **SPSS Statistical Algorithms 2nd Edition,** (Chicago:1991), s.138-145.; M. J. Norusıs, **SPSS Regression Models 10.0**, (Chicago: SPSS Inc., 1999), s.1-14.

4. UYGULAMA ÇALIŞMASINDA KULLANILAN VERİLER

İkili lojistik regresyon modeli ile analiz edilen 100 müşterinin otomobilleri için kasko sigortaları var ve yok şeklinde sonuçlanan, 2002-2003 yılına ait veriler tablo 4.1'de verilmiştir.

Tablo 4.1. Axa Oyak Sigorta A.Ş. Acenteliği 2002-2003 Yılı Kasko Sigortası Verileri

Gözlem No	Ortalama Aylık gelir	Bireyin yaşı	Aracın yaşı	Kasko sigortası
1	4.000.000.000	43	1	1
2	910.000.000	35	4	1
3	480.000.000	43	10	1
4	720.000.000	41	7	0
5	1.200.000.000	38	3	1
6	480.000.000	31	8	0
7	860.000.000	34	7	0
8	1.000.000.000	39	8	1
9	510.000.000	31	2	1
10	880.000.000	32	11	0
11	910.000.000	33	5	1
12	420.000.000	41	9	1
13	4.200.000.000	58	4	1
14	440.000.000	52	5	0
15	900.000.000	28	4	1
16	710.000.000	39	7	1
17	450.000.000	47	9	0
18	790.000.000	33	8	0
19	360.000.000	27	15	0
20	540.000.000	37	3	1
21	690.000.000	57	10	1
22	3.200.000.000	59	3	1
23	710.000.000	32	7	1

Gözlem No	Ortalama Aylık gelir	Bireyin yaşı	Aracın yaşı	Kasko sigortası
24	840.000.000	31	4	1
25	910.000.000	33	7	0
26	1.200.000.000	37	3	1
27	1.310.000.000	39	5	1
28	860.000.000	49	2	1
29	870.000.000	34	1	1
30	750.000.000	51	2	1
31	660.000.000	35	14	0
32	710.000.000	57	15	0
33	2.100.000.000	58	3	1
34	3.000.000.000	43	10	1
35	410.000.000	29	7	0
36	860.000.000	28	4	1
37	550.000.000	41	6	1
38	1.250.000.000	37	3	1
39	480.000.000	48	21	0
40	500.000.000	45	14	0
41	910.000.000	48	2	1
42	980.000.000	50	1	1
43	820.000.000	33	2	1
44	1.350.000.000	37	2	1
45	970.000.000	54	18	0
46	390.000.000	54	19	0
47	750.000.000	51	11	0
48	440.000.000	48	7	1
49	630.000.000	30	6	1
50	560.000.000	54	24	0
51	690.000.000	49	13	0
52	740.000.000	28	7	1
53	820.000.000	50	6	1
54	4.000.000.000	51	3	1
55	1.050.000.000	43	2	1
56	950.000.000	38	7	0
57	845.000.000	34	10	0
58	3.000.000.000	55	1	1
59	700.000.000	46	2	1
60	450.000.000	45	5	0
61	710.000.000	45	4	1
62	560.000.000	34	9	0
63	900.000.000	59	4	1
64	900.000.000	59	3	1

Gözlem No	Ortalama Aylık gelir	Bireyin yaşı	Aracın yaşı	Kasko sigortası
65	1.950.000.000	48	2	1
66	500.000.000	36	19	0
67	410.000.000	45	8	1
68	700.000.000	31	3	1
69	580.000.000	33	6	1
70	470.000.000	25	12	0
71	500.000.000	36	13	0
72	835.000.000	51	3	1
73	750.000.000	38	3	0
74	620.000.000	44	12	0
75	450.000.000	34	5	0
76	560.000.000	49	7	0
77	1.000.000.000	51	3	1
78	1.210.000.000	49	2	1
79	440.000.000	46	17	0
80	570.000.000	31	7	1
81	1.580.000.000	48	1	1
82	1.460.000.000	48	2	1
83	280.000.000	47	23	0
84	740.000.000	32	16	0
85	830.000.000	38	9	1
86	390.000.000	53	15	0
87	420.000.000	37	8	0
88	730.000.000	41	3	0
89	580.000.000	49	1	1
90	530.000.000	43	12	0
91	550.000.000	27	13	0
92	1.335.000.000	54	2	1
93	1.050.000.000	38	2	0
94	590.000.000	37	9	0
95	895.000.000	44	5	1
96	600.000.000	39	8	0
97	1.400.000.000	57	1	1
98	1.800.000.000	48	2	1
99	460.000.000	50	15	0
100	600.000.000	38	6	1

Yukarıdaki tablodaki veriler Axa Oyak'ın Eskişehir Acentesi'nden derlenmiştir.

5. ÇÖZÜM SONUÇLARI

Aracın yaşı, bireyin yaşı ve aylık gelir bağımsız değişkenler olarak ele alınmış ve bu değişkenlerin müşterilerin kasko sigortası yaptırmaları üzerine etkileri ikili lojistik regresyon modelinde analiz edilmiştir. Analiz sonuçları tablo 4.2'de verilmiştir.

Tablo 4.2. Aracın yaşı, bireyin yaşı ve aylık gelir bağımsız değişkenler olarak ele alındığında ikili lojistik regresyon modeli analiz sonuçları

DEĞİŞKEN	KAT-SAYILAR	STD. HATA	WALD İSTATİSTİĞİ	SERBESTLİK DERECESİ	ANLAMLILIK DÜZEYİ	ODDS ORANI
ARACIN YAŞI	-0,423	0,104	16,636	1	0,000	0,655
BİREYİN YAŞI	0,032	0,038	0,675	1	0,411	1,032
AYLIK GELİR	1,98E-09	1,264E-09	2,461	1	0,117	1,000
SABİT	0,483	1,920	0,063	1	0,801	1,622

Wald istatistiği 1 serbestlik dereceli ki-kare dağılımı gösterir. 1 serbestlik dereceli ve % 5 anlamlılık düzeyine ait ki-kare tablo değeri 3,84 dür.

X_1 için : wald istatistiği 16,636 > 3,84 olduğundan, X_1 bağımsız değişkeninin bağımlı değişken üzerinde etkili olmadığı şeklindeki H_0 hipotezi reddedilir. Buna göre aracın yaşı değişkeninin bağımlı değişken üzerinde etkili olduğuna karar verilir.

X_2 için : wald istatistiği 0,675 < 3,84 olduğundan, X_2 bağımsız değişkeninin bağımlı değişken üzerinde etkili olmadığı şeklindeki H_0 hipotezi kabül edilir. Buna göre bireyin yaşı değişkeninin bağımlı değişken üzerinde etkili olmadığına karar verilir.

X_3 için : wald istatistiği 2,461 < 3,84 olduğundan, X_3 bağımsız değişkeninin bağımlı değişken üzerinde etkili olmadığı şeklindeki H_0 hipotezi kabül edilir. Buna göre aylık gelir değişkeninin bağımlı değişken üzerinde etkili olmadığına karar verilir.

Yukarıdaki sonuçlar, anlamlılık düzeylerinin sınanmasından da elde edilebilir.

Aracın yaşı : anlamlılık düzeyi 0,000 < 0,05 olduğundan anlamlıdır
Bireyin yaşı : anlamlılık düzeyi 0,411 > 0,05 olduğundan anlamlı değildir.
Aylık gelir : anlamlılık düzeyi 0,117 > 0,05 olduğundan anlamlı değildir.

Bu durumda modelden bireyin yaşı çıkarılarak, aracın yaşı ve aylık gelir bağımsız değişkenler olarak ele alınmış ve bu değişkenlerin müşterilerin kasko sigortası yaptırmaları üzerine etkileri ikili lojistik regresyon modelinde analiz edilmiştir. Analiz sonuçları tablo 4.3'de verilmiştir.

Tablo 4.3. Aracın yaşı ve aylık gelir bağımsız değişkenler olarak ele alındığında ikili lojistik regresyon modeli analiz sonuçları

DEĞİŞKEN	KAT-SAYILAR	STD. HATA	WALD İSTATİSTİĞİ	SERBESTLİK DERECESİ	ANLAMLILIK DÜZEYİ	ODDS ORANI
ARACIN YAŞI	-0,420	0,104	16,292	1	0,000	0,657
AYLIK GELİR	2,00E-09	1,227E-09	2,662	1	0,103	1,000
SABİT	1,689	1,275	1,755	1	0,185	5,416

X_1 için : wald istatistiği 16,292 > 3,84 olduğundan, X_1 bağımsız değişkeninin bağımlı değişken üzerinde etkili olmadığı şeklindeki H_0 hipotezi reddedilir. Buna göre aracın yaşı değişkeninin bağımlı değişken üzerinde etkili olduğuna karar verilir.

X_3 için : wald istatistiği 2,662 < 3,84 olduğundan, X_3 bağımsız değişkeninin bağımlı değişken üzerinde etkili olmadığı şeklindeki H_0 hipotezi kabül edilir. Buna göre aylık gelir değişkeninin bağımlı değişken üzerinde etkili olmadığına karar verilir.

Aracın yaşı, anlamlılık düzeyi 0,00 < 0,05 olduğundan anlamlıdır. Aylık gelir, anlamlılık düzeyi 0,103 > 0,05 olduğundan anlamlı değildir.

Bu durumda modelden aylık gelir çıkarılarak, aracın yaşı bağımsız değişken olarak ele alınmış ve bu değişkenin müşterilerin kasko sigortası yaptırmaları üzerine etkisi ikili lojistik regresyon modelinde analiz edilmiştir. Analiz sonuçları tablo 4.4'de verilmiştir.

Tablo 4.4. Aracın yaşı bağımsız değişken olarak ele alındığında ikili lojistik regresyon modeli analiz sonuçları

DEĞİŞKEN	KAT-SAYILAR	STD. HATA	WALD İSTATİSTİĞİ	SERBESTLİK DERECESİ	ANLAMLILIK DÜZEYİ	ODDS ORANI
ARACIN YAŞI	-0,480	0,096	24,877	1	0,000	0,619
SABİT	3,623	0,694	27,251	1	0,000	37,444

X_1 için : wald istatistiği 24,877 > 3,84 olduğundan, X_1 bağımsız değişkeninin bağımlı değişken üzerinde etkili olmadığı şeklindeki H_0 hipotezi reddedilir. Buna göre aracın yaşı değişkeninin bağımlı değişken üzerinde etkili olduğuna karar verilir.

Aracın yaşının $\left(\hat{\beta}_1 = -0{,}480 \pm 0{,}096\right)$, müşterilerin kasko sigortası yaptırmaları üzerine etkisi olduğu görülmektedir (ODDS oranı=0,619). Aracın yaşı 1 yaş arttığında, kasko sigortası yaptırma olasılığı

1 / 0,619 = 1,615 kat düşmektedir.

Aracın yaşı 5 yaş arttığında da, kasko sigortası yaptırma olasılığı

$e^{(5 \times -0,480)} = e^{-2,4}$

$= 0,09$

1 / 0,09 = 11,1 kat düşecektir.

SONUÇ

Bağımlı değişkenin ikili, üçlü ve çoklu sınıflar halinde gözlendiği durumlarda, bağımsız değişkenlerle arasındaki ilişkiyi tanımlamak için en uygun modelin bulunmasında lojistik regresyon analizi kullanılır. Bağımlı değişkenin ikili cevap içerdiği lojistik regresyon modeli, ikili lojistik regresyon modelidir.

İkili lojistik regresyon analizinin uygulandığı bu çalışmada müşterilerin kasko sigortası yaptırmaları (y_j) ile aracın yaşı (X_1), bireyin yaşı (X_2) ve bireyin aylık geliri (X_3) arasındaki ilişki ortaya konulmaya çalışılmıştır. Bu amaçla parametre tahminleri hesaplanmış, parametre tahminlerinin anlamlılığı ve modelin uygunluğunun test edilmesi için Wald testinden yararlanılmıştır. Analizin sonucunda bireyin yaşı ve aylık gelirinin, müşterilerin kasko sigortası yaptırmaları üzerine etkisi olmadığı belirlenmiştir. Aracın yaşının, müşterilerin kasko sigortası yaptırması üzerine etkisi olduğu görülmüştür. Aracın yaşı 1 yaş arttıkça, kasko sigortası yaptırma olasılığı 1,615 kat düşmektedir. Buna göre müşterilerin, araçlarının yaşı ilerledikçe sadece trafik sigortası yaptırmayı yeterli görme olasılığıda artmaktadır.

KAYNAKLAR

AKKAYA, Şahin ve PAZARLIOĞLU, Vedat. **Ekonometri 2.** İstanbul: 1998.

ÇOLAK, Ertuğrul. **"Koşullu ve Sınırlandırılmış Lojistik Regresyon Yöntemlerinin Karşılaştırılması ve Bir Uygulama."** Yüksek Lisans Tezi, Osmangazi Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı, Eskişehir: 2002.

DUMAN, Şebnem. **Sigorta Tekniği**. Eskişehir: Anadolu Üniversitesi Yayın no:816, Aralık 1995.

ERTEK, Tümay. **Ekonometriye Giriş**. İstanbul: Beta Yayınları, 1996.

GREEN, Peter J. **Nonparametric Regression and Generalized Linear Models.** London: Chapman & Hall, 1994.

KOTZ , Samuel and JOHNSON, Norman L . **Encylopedia Of Statistical Sciences.** Newyork: J. Wiley & Sons, Vol:5, 1982.

LANDWEHR, James M., PREGIBON, Daryl and SHOEMAKER, Anne C. "Graphical Methods For Assessing Logistic Regression Models", **Journal Of The American Statistical Association**, March 1984.

NORUSIS M. J. **SPSS Regression Models 10.0.** SPSS Inc., Chicago: 1999.

ORHUNÖZ, Ergun. **Uygulamada Karayolları Trafik Kanununa Göre Sorumluluk Tazminat Sigorta.** Ankara: Seçkin Yayınevi,1998.

ÖZDAMAR, Kazım. "Biyolojik Denemelerin Oransal Cevaplarının Analizinde Probit ve Lojit Regresyon Yöntemlerinin Etkinliğinin Araştırılması", **Anadolu Tıp Dergisi**,1988.

------------. **Paket Programlar İle İstatistiksel Veri Analizi-1.** Eskişehir: Kaan Kitabevi, 1997.

ÖZKAN, Mehmet. **Sigorta İşlemleri Ve Muhasebesi.** İstanbul: Bilim Teknik Yayınevi, 1998.

RYAN, Thomas P. "Some Issues In Logistic Regression", **Communications In Statistics Theory And Methods**, Vol:29, No:9 &10, 2000.

ŞIKLAR, Emel. **Regresyon Analizine Giriş.** Eskişehir: Anadolu Üniversitesi Fen Fakültesi Yayınları, No:16, 2000.

TAŞYÜREK, Hayri. **Kasko Sigortası.** 1. Baskı. Ankara: Seçkin Yayıncılık, 2001.

TATLIDİL, Hüseyin. **Uygulamalı Çok Değişkenli İstatistiksel Analiz.** Hacettepe Üniversitesi, 1996.

YÜCESOY, Cevat. **Sigorta İşletmeleri ve Muhasebesi.** İstanbul: Çağlayan Muhasebe Serisi No:8, Çağlayan Kitabevi B.1, 1966.

SPSS Inc. **SPSS Statistical Algorithms 2nd Edition**, Chicago:1991.

Temel Sigorta Bilgileri. İstanbul: Türkiye Genel Sigorta A.Ş.

www.**axaoyak**.com.tr (24/04/2003).

www.wiley.co.uk/eob/sample4.pdf (24/04/2003).