

Çoklu İç İlişki ve Aykırı Değer Olduğu Durumda Lojistik Regresyon Yaklaşımları

Ebru Gündođan

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı

Haziran 2015

Logistic Regression Approaches in case of Multicollinearity and Outlier Existence

Ebru Gündođan

MASTER OF SCIENCE THESIS

Department of Statistics

June 2015

Çoklu İç İlişki ve Aykırı Değer Olduğu Durumda Lojistik Regresyon Yaklaşımları

Ebru Gündoğan

Eskişehir Osmangazi Üniversitesi
Fen Bilimleri Enstitüsü
Lisansüstü Yönetmeliği Uyarınca
İstatistik Anabilim Dalı
İstatistik Teorisi Bilim Dalında
YÜKSEK LİSANS TEZİ
Olarak Hazırlanmıştır

Danışman: Doç. Dr. Arzu Altın Yavuz

Haziran 2015

ONAY

İstatistik Anabilim Dalı Yüksek Lisans öğrencisi Ebru Gündoğan'ın YÜKSEK LİSANS tezi olarak hazırladığı “Çoklu İç İlişki ve Aykırı Değer Olduğu Durumda Lojistik Regresyon Yaklaşımları” başlıklı bu çalışma, jürimizce lisansüstü yönetmeliğin ilgili maddeleri uyarınca değerlendirilerek kabul edilmiştir.

Danışman : Doç. Dr. Arzu ALTIN YAVUZ

İkinci Danışman :...

Yüksek Lisans Tez Savunma Jürisi:

Üye : Doç. Dr. Arzu ALTIN YAVUZ

Üye : Prof. Dr. Zeki YILDIZ

Üye : Doç. Dr. Zeynep FİLİZ

Üye :Doç. Dr. Sema BEHDİOĞLU

Üye :Yrd. Doç. Dr. Sinan SARAÇLI

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun tarih ve
.....sayılı kararıyla onaylanmıştır.

Prof. Dr. Hürriyet ERŞAHAN
Enstitü Müdürü

ETİK BEYAN

Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kılavuzuna göre, Doç. Dr. Arzu ALTIN YAVUZ danışmanlığında hazırlamış olduğum “Çoklu İç İlişki ve Aykırı Değer Olduğu Durumda Lojistik Regresyon Yaklaşımları” başlıklı YÜKSEK LİSANS tezimin özgün bir çalışma olduğunu; tez çalışmamın tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; tezimde verdiğim bilgileri, verileri akademik ve bilimsel etik ilke ve kurallara uygun olarak elde ettiğimi; tez çalışmamda yararlandığım eserlerin tümüne atıf yaptığımı ve kaynak gösterdiğimi ve bilgi, belge ve sonuçları bilimsel etik ilke ve kurallara göre sunduğumu beyan ederim.

30/06/2015

Ebru Gündoğan

İmza

ÖZET

Regresyon analizi, bağımlı bir değişken ile bağımsız değişken/değişkenler arasındaki fonksiyonel ilişkiyi açıklamak ve bağımlı değişken değerlerini tahmin etmek için kullanılan istatistiksel bir yöntemdir. Değişkenler arasındaki bu ilişkiyi belirlemek için birçok regresyon yöntemi geliştirilmiştir. Regresyon analizinde bağımlı değişken genellikle sürekli ve ölçülebilir bir değişken olarak ele alınır. Ancak bazı durumlarda bağımlı değişken kesikli bir değişken olabilir ve sınırlı sayıda değer alabilir. Bağımlı değişkenin kesikli olması durumunda uygulanan regresyon analizi yöntemine lojistik regresyon analizi denir.

Lineer regresyon modelinde olduğu gibi lojistik regresyon modelinde de açıklayıcı değişkenler arasında çoklu iç ilişki söz konusu olabilir. Bu durumda klasik lojistik regresyon analizi sonucunda elde edilen regresyon katsayıları gerçek değerlerini yansıtmayacaktır. Ayrıca elde edilen modelin tahmin yeteneği düşük olabilecektir. Bu durum için lojistik regresyon analizine alternatif olarak Ridge lojistik regresyon modeli kullanılmalıdır.

Regresyon analizinde karşılaşılan bir diğer sorun veri setinde aykırı değerlerin yer almasıdır. Bu durumda model sağlaması gereken varsayımların bazılarını sağlayamamaktadır. Veri setinde aykırı değerlerin bulunması durumunda bu değerlerin olumsuz etkilerini en aza indireyecek robust yöntemler tercih edilmelidir. Lojistik regresyon analizinde veri setinde aykırı değerlerin bulunması halinde yaygın olarak Bianco-Yohai tahmin edicisi kullanılmaktadır.

Bu çalışmada 132 ülkeye ait sosyal gelişmişlik endeks değeri kullanılarak lojistik regresyon modelleri kurulmuştur. Oluşturulan modeller doğru sınıflandırma performansları bakımından karşılaştırılmıştır.

Anahtar Kelimeler: Lojistik regresyon, Lojistik ridge regresyon, Robust lojistik regresyon (Bianco-Yohai tahmincisi), Çoklu iç ilişki, Aykırı değer

SUMMARY

Regression analysis explains functional relationships between a dependent variable and an independent variable (or variables). It is a statistical method which is used to predict values of the dependent variable. There are several regression techniques to determine the relationship between the variables. In regression analysis, the dependent variable is usually considered as continuous and measurable. However, in some cases, the dependent variable could be discrete and restricted. When the dependent variable is discrete, performed regression analysis is called as logistic regression analysis.

In logistic regression models, multicollinearity can be valid between the explanatory variables as in the linear regression models. In this case, the classical logistic regression coefficients obtained from the analysis results will not reflect the actual values. Moreover, the prediction ability of the model will be lower. In that case Ridge logistic regression model should be used as an alternative to logistic regression analysis.

Another problem encountered in the regression analysis of the data set is the existence of outliers. The model cannot satisfy some of the assumptions in this situation. One of the best solutions for this problem is to use robust methods to minimize the negative effects of outliers on the model. When the data includes outliers, Bianco-Yohai estimator is used in logistic regression analysis in common.

In this study, logistic regression models were established by using social development index of 132 countries. These models were compared in terms of correct classification performance.

Keywords: Logistic regression, Logistic ridge regression, Robust logistic regression (Bianco-Yohai estimator), Multicollinearity, Outliers.

TEŐEKKÜRLER

Yüksek lisans çalışmam boyunca, gerek derslerimde ve gerekse tez çalışmamda, bana danışmanlık ederek, beni yönlendiren ve her türlü olanağı sağlayan değerli hocam sayın Doç. Dr. Arzu Altın Yavuz' a en içten saygı ve teşekkürlerimi sunarım.

Lisans ve Yüksek lisans eğitimim boyunca her konuda desteklerini esirgemeyen ihtiyaç duyduğum her anda sorularıma yılmadan cevap veren değerli hocalarım Yakup Murat Bulut, Talha Arslan ve Barış Ergül'e sonsuz teşekkürler sunarım.

Ayrıca eğitim hayatım boyunca maddi ve manevi bana destek olan aldığım her kararı sorgulamadan arkamda duran kocaman ailem ve özellikle annem Ayten Gündoğan, babam Münir Gündoğan, abilerim Hakan Gündoğan, Hüseyin Gündoğan ve ikizim Emre Gündoğan'a sonsuz teşekkürler.

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	vi
SUMMARY	vii
TEŞEKKÜR	viii
İÇİNDEKİLER	ix
ŞEKİLER DİZİNİ	xi
ÇİZELGELER DİZİNİ	xii
SİMGELER VE KISALTMALAR DİZİNİ	xiii
1. GİRİŞ	1
2. LİTERATÜR ARAŞTIRMASI	4
3. LOJİSTİK REGRESYON MODELİ	7
3.1. Lojistik Regresyonun Lineer Regresyon İle İlişkisi.....	7
3.2. Lojistik Regresyon Modeli.....	8
3.3. Lojistik Regresyon Modelinde Parametre Tahmin Yöntemleri.....	11
3.3.1.En çok olabilirlik ve Newton-Raphson yöntemleri.....	11
3.3.2.En küçük kareler yöntemi.....	15
3.3.3. Yeniden ağırlıklandırılmış en küçük kareler yöntemi.....	16
3.4. Lojistik Sınıflandırma ve Katsayı Yorumları.....	18
3.5. Lojistik Regresyon Çeşitleri.....	17

İÇİNDEKİLER (devam ediyor)

3.5.1. İkili lojistik regresyon.....	17
3.5.2. Ordinal lojistik regresyon.....	17
3.5.3. Nominal lojistik regresyon.....	17
3.5.4. Koşullu ve sıralı lojistik regresyon.....	18
3.6. Modelin Uyum İyiliğinin Değerlendirilmesi.....	18
3.6.1. Hosmer-Lemeshow (H-L) testi.....	18
3.6.2. Sınıflandırma tabloları.....	18
3.6.3. Yapay Cox&Snell ve Nagelkerke R2 değerleri.....	19
4. LOJİSTİKREGRESYONDA ALTERNATİF TAHMİN YÖNTEMLERİ.....	20
4.1. Lojistik Ridge Regresyon.....	21
4.2. Robust Lojistik Regresyon.....	27
5. UYGULAMA.....	29
5.1. Giriş.....	29
5.2. Sosyal Gelişmişlik Endeksi.....	30
5.3. Lojistik Ridge Regresyon Modeli Sonuçları.....	36
5.4. Robust Lojistik Regresyon Modeli Sonuçları.....	38
5.5. Yöntemlerin Karşılaştırılması.....	41
6. SONUÇ VE ÖNERİLER.....	42
KAYNAKLARDİZİNİ.....	44

ŞEKİLLER DİZİNİ

Şekil

Sayfa

5.1. Aykırı değerler için grafik gösterimi.....39

ÇİZELGELER DİZİNİ

<u>Cizelge</u>	<u>Sayfa</u>
3.1. Bernoulli olasılık dağılımı.....	8
5.1. Tanımlayıcı istatistikler.....	32
5.2. Korelasyon matrisi.....	33
5.3. Bağımsız değişkenlerin bağımlı değişkeni açıklama oranları.....	34
5.4. Hosmer-Lemeshow anlamlılık testi.....	34
5.5. Modelin sınıflandırma tablosu.....	35
5.6. Lojistik regresyon modeli analiz sonuçları.....	36
5.7. Lojistik ridge regresyon model sonuçları.....	37
5.8. Aykırı değerler.....	38
5.9. Robust lojistik regresyon analiz sonuçları.....	40
5.10. Doğru sınıflandırma oranı.....	41

SİMGELER VE KISALTMALAR DİZİNİ

<u>Kısaltmalar</u>	<u>Açıklamalar</u>
İE	İleri Eğitim
Tİİ	Temel İnsan İhtiyaçları
Bİ	Bilgi İhtiyaçları
TTİ	Temel Tıp İhtiyaçları
ES	Ekosistem Sürdürülebilirliği
EÇO	En Çok Olabilirlik
EKK	En Küçük Kereler
Rİ	Refah İhtiyaçları
MG	Milli Gelir
H-L	Hosmer-Lemeshow
Sİ	Sağlık İhtiyaçları
BKV	Bilgi Keşfi Veritabanları
BİE	Bilgi ve İletişime Erişim
IRLS	Yeniden Ağırlıklandırılmış İteratif En Küçük Kareler
İGE	İnsani Gelişme Endeksi
İGG	İnsani Gelişme Göstergelerini
İGR	İnsani Gelişme Raporları
FR	Fırsat

SİMGELER VE KISALTMALAR DİZİNİ (devam ediyor)**Kısaltmalar**

KÖ

KH

Gİ

SGE

BİH

TOL

VİF

SİH

Açıklamalar

Kişisel Özgürlük

Kişisel Haklar

Güvenlik İhtiyaçları

Sosyal Gelişmişlik Endeksi

Barınak İhtiyaçları

Tolerans

Varyans Şişirme Faktörü

Su İhtiyaçları

1.Giriş

Hemen hemen her alanda teknolojinin gelişmesiyle ilgili toplanan verilerin boyutları artmaktadır. Elde edilen verilerin, çok boyutlu olması da istatistiksel yöntemleri etkilemektedir. Bu durumda çok değişkenli istatistiksel yöntemlerin tercih edilmesi daha doğru bir yaklaşım olmaktadır. Çok değişkenli istatistiksel verilerin sınıflandırılması için, bu verilere uygulanabilecek çeşitli istatistiksel yöntemler bulunmaktadır. Gözlemleri, verilerin yapısında bulunan olası gruplara atamak için kullanılan analiz yöntemlerinden bazıları ise kümeleme (clustering), diskriminant (discriminant) ve lojistik regresyon (logistic regression) analiz yöntemleridir (Atabey, 2010).

Kümeleme analizinde verilerin yapısındaki grup sayısı bilinmemekte, gözlemler uzaklık ya da benzerlik ölçütlerine göre kümelenebilmektedir. Veriler veya uzaklık değerleri normal dağılımlı olmalıdır. Kümeleme analizinde kovaryans matrisi için varsayım bulunmamaktadır. Burada amaç yalnızca gözlemlerin oluşturduğu küme yapısını belirlemektir (Çoşkun vd, 2004).

Diskriminant ve lojistik regresyon analizinde ise verilerin grup sayısı bilinmemekte ve bu verilerden faydalanarak bir model elde edilmektedir. Bu model yardımıyla veri kümesine yeni alınan gözlemlerin gruplara atamaları yapılmaktadır. Ancak, diskriminant analizindeki veriler çok değişkenli normal dağılıma sahip olmalıdır. Uygulamada normallik varsayımının sağlanamadığı durumda diskriminant analize alternatif olarak görülen lojistik regresyon analizi, bu özelliği nedeniyle son yıllarda kullanımı oldukça artan bir ayırma modelidir (Atabey, 2010; Çoşkun vd, 2004).Lojistik regresyon analizi çeşitli varsayım bozulmaları (normal dağılıma ve ortak kovaryansa sahip olmama) durumunda diskriminant analizi ve çapraz tablolara bir alternatif olurken, bağımlı değişkenin 0 ve 1 gibi ikili ya da ikiden çok düzey içeren kesikli değişken olması durumunda normallik varsayımının bozulması nedeniyle doğrusal regresyon analize alternatif olmaktadır (Çoşkun vd, 2004).

Regresyon analizi, bağımlı bir değişken ile bağımsız değişken/değişkenler arasındaki fonksiyonel ilişkiyi açıklamak ve bağımlı değişken değerlerini tahmin etmek için kullanılan istatistiksel bir yöntemdir. Değişkenler arasındaki bu ilişkiyi belirlemek için birçok regresyon yöntemi geliştirilmiştir. Regresyon analizinde bağımlı değişken genellikle sürekli ve ölçülebilir bir değişken olarak ele alınır. Ancak bazı durumlarda bağımlı değişken kesikli

bir deęişken olabilir ve sınırlı sayıda deęer alabilir. Baęımlı deęişkenin kesikli olması durumunda uygulanan regresyon analizi yöntemine lojistik regresyon analizi denir. Lineer regresyon analizinin kullanılmadıęı durumlardan biri olan baęımlı deęişkenin iki ya da ikiden çok düzey içiren deęişken olduęu durumda, normallik varsayımı bozulmaktadır. Bu durumda lineer regresyon analizi uygulamak yerine alternatif olan lojistik regresyon uygulanmalıdır (Başarı, 1990; Atabey, 2010).

Lojistik regresyon analizinin amacı, istatistikte kullanılan dięer model kurma tekniklerinde olduęu gibi, en az deęişkeni kullanarak en iyi uyuma sahip olacak şekilde baęımlı ve baęımsız deęişkenler arasındaki ilişkiyi tanımlayabilen ve genel olarak kabul edilebilen bir model kurmaktır (Çoşkun vd, 2004).

Lojistik regresyon analizinde model parametreleri tahmin edilirken sıradan en küçük kareler yöntemi için gerekli olan varsayımlar sağlanamamaktadır. Bu nedenle lojistik regresyon modelinde parametre tahmini için ilgili varsayımlara gerek duymayan en çok olabilirlik (EÇO) yöntemi kullanılmaktadır. Ancak, en çok olabilirlik yöntemi kullanılarak lojistik regresyon modelinin parametrelerinin tahmini için ele alınan örnek hacminin büyük olması gerekmektedir (Ersan 2007). En çok olabilirlik tahminçileri, gözlem sayısının çok fazla olmadığı durumlarda aykırı deęer olarak adlandırılan ve verinin büyük çoğunluęundan önemli derecede farklılık gösteren gözlemlerin varlıęından oldukça etkilenmektedir. Doğrusal regresyon analizinde baęımlı deęişkenin alacaęı deęer kestirilirken, lojistik regresyonda baęımlı deęişkenin alabileceęi deęerlerin ortaya çıkabilme olasılıkları kestirilmektedir. Bu nedenle lojistik regresyon analizinde aykırı deęerler, modelden tahmin edilen başarı olasılıęı küçük olduęu halde gözlemlenen durumu başarılı olan ya da modelden elde edilen başarı olasılıęı yüksek olduęu halde gözlemlenen durumu başarısız olan gözlemler olarak tanımlanabilir(Yavuzkanat, 2011). Ayrıca, örnek hacminin küçük olması ya da veri setinin aykırı deęer içermesi durumunda genel lojistik regresyon yöntemleri geçerli ve güvenilir sonuçlar vermeyebilmektedir (Bilgin, 2012). Bu durumda, aykırı deęerlerin etkilerini en aza indirmeye çalışan robust lojistik regresyon yöntemleri alternatif tahmin yöntemleri olarak kullanılabilir. Bianco ve Yohai (1996) tarafından geliştirilen robust lojistik regresyon yöntemi literatürde yaygın olarak kullanılan yöntemlerden biridir.

Çoklu doğrusal regresyon analizinde olduğu gibi, lojistik regresyon analizinde de model kurulurken ele alınan bağımsız değişkenler arasında ilişki olmaması gerekmektedir. Bağımsız değişkenler arasında çoklu iç ilişki söz konusu olduğunda artık elde edilen tahmin ediciler minimum varyanslılık gibi önemli bir özelliği sağlayamamaktadır. Çoklu doğrusal regresyonda bağımsız değişkenler arasında çoklu iç ilişki varlığı durumunda ya ele alınan değişkenlerin indirgenmesi yoluna gidilmekte ya da yanlı regresyon yöntemlerine başvurulmaktadır. Ancak, çalışmadan önemli bir açıklayıcı değişkeni çıkarmak yerine yansızlıktan ödün vererek daha küçük varyanslı tahminler elde edilmesini sağlayan yanlı regresyon yöntemlerini kullanmak daha fazla kabul görmektedir. Bu amaçla literatürde yaygın olarak kullanılan yanlı regresyon yöntemlerinden biri Ridge regresyon yöntemidir. Çoklu doğrusal regresyon analizinde olduğu gibi, lojistik regresyon analizinde de Ridge regresyon yöntemi kullanılarak model parametreleri tahmin edilebilmektedir (Schaefer vd, 1984).

Bu tez çalışmasında, bağımlı değişkenin kategorik değişken olması durumunda uygulamada sıklıkla karşılaşılabilecek iki önemli problem göz önünde bulundurulmuştur. Bu durumlardan ilki bağımsız değişkenler arasında çoklu iç ilişki olması durumu, diğeri ise veri setinde aykırı değer bulunması durumudur. Her iki durumda da lojistik regresyon modelinin oluşturulabilmesi için parametre tahmin yöntemleri ele alınmıştır. Basitlik açısından bağımlı değişkenin iki düzeyli olduğu lojistik regresyon analizi göz önünde bulundurulmuştur. Ancak ele alınan yöntemler ikiden fazla değer alan bağımlı değişkenler içinde kullanılabilir. Çalışmanın ilk bölümünde, lojistik regresyon modeli diğer yöntemlerle karşılaştırmalı olarak ele alınmıştır. İkinci bölümde lojistik regresyon analizinin literatür taramasına yer verilmiştir. Üçüncü bölümde lojistik regresyon modeli, parametre tahmin yöntemi ve uyum iyiliği ölçütleri hakkında bilgi verilmiştir. Dördüncü bölümde lojistik ridge ve robust lojistik regresyon modelleri ve parametre tahmin yöntemleri anlatılmıştır. Son bölümde 132 ülkeden alınan alt endeks değerleriyle ülkelerin gelişmişlik düzeyini hesaplamak için lojistik regresyon, lojistik Ridge ve robust lojistik regresyon analizleri kullanılmıştır. İlgili modeller için elde edilen sonuçlar doğru sınıflandırma performansları bakımından karşılaştırılmıştır. Sonuç ve öneriler kısmında ise elde edilen sonuçlar değerlendirilmiş, veri setinde aykırı değer bulunması ya da çoklu iç ilişki olması durumunda kullanılabilecek alternatif lojistik regresyon analiz yöntemleri tartışılmıştır.

2. LİTERATÜR ARAŞTIRMASI

Lojistik regresyon analizi, son yıllarda tıpta özellikle epidemiolojide, ekonomi, ziraat, spor, biyoloji alanındaki uygulamalarda ve sınıflandırma sorununun olduğu her yerde sıklıkla kullanılmaktadır. Berkson (1944), biyolojik deneylerin analizinde lojistik regresyonu kullanmıştır ve bu literatürde ki ilk uygulamadır. Cox (1970), Berkson' nun önerdiği modeli geliştirmiştir. Özet gelişmeler ise ilk olarak Andersson (1979, 1983), tarafından verilmiştir (Bircan, 2004). Halperin ve ark. (1971) ve Finney (1971), lojistik regresyonun diğer analiz yöntemleriyle karşılaştırılması üzerine çalışmalar yapılmışlardır.

Aranda-Ordaz (1981) ve Johnson (1985), lojistik modelin uyum iyiliğiyle ilgili çalışmalar yapmışlardır. Pregibon (1981), iki gruplu lojistik modelde etkin, aykırı gözlemleri ve belirleme ölçülerini incelemiştir. Lee (1984), basit dönüşümlü deneme planları için lojistik modeller üzerinde çalışmıştır. Bonney (1987), lojistik regresyon modelinin geliştirilmesi için çalışmalar yapmıştır. Duffy (1990), lojistik regresyonda hata terimlerinin dağılışı ve parametre değerlerinin gerçek değerlere yaklaşımını incelemiştir.

O'Neil ve Barry (1995), trafik kazalarına ait verileri ele alarak yaş ve cinsiyet gibi bağımsız değişkenlerin kazalarda ölüm sebebi olabileceğini araştırmışlardır. Garside ve Glueck (1995), insanlarda sigara ve alkol kullanımı gibi risk faktörlerinin kalp hastalığı üzerindeki etkilerini incelemiştir. Kloiber ve ark. (1996), Peoples ve ark. (1991), Buescher ve ark. (1993)' nın kadınlarda düşük doğum ağırlığını etkileyen risk faktörlerini ve Sable ve Herman (1997), erken doğum ve düşük doğum ağırlığı arasındaki ilişkiyi incelemiştir (Bircan, 2004).

Heath ve ark. (2000), etnik kimlik farklılıklarının işgücü piyasasındaki etkilerini 1991 UK İşgücü Araştırması verilerini lojistik regresyonla analiz etmişlerdir. Akkaya ve Pazarlıoğlu (1998), lojistik regresyonu ekonomi alanında kullanmışlardır. Preiffer ve Gail (2001), ikiz kardeşlerin genetik özellikleriyle ilgili çalışma yapmışlardır. Önder (2001), lojistik regresyonda değişken seçimini ele almış ve koyunların süten kesilme ağırlıkları ile ilgilenmiştir. Hughes ve ark. (2004), hemşirelerin alkol ve bazı uyuşturucu madde içeren ilaçlara karşı bağımlılıklarını lojistik regresyonla incelemiştir. Tester ve ark. (2004), yollardaki hız kesici kasislerin yaralanan çocuk sayısını azaltma üzerindeki etkisini incelerken lojistik regresyondan yararlanmışlardır.

Bayram (2004), ikili ve çoklu lojistik regresyon analizinin teorik yapılarını ortaya koymak ve ücretsiz ve gelir karşılığı çalışan iş gücünü bağımsız değişkenlerle karşılaştırarak açıklamaya çalışmıştır. Bircan (2004), çocuklarda doğum ağırlıklarını etkileyen önemli risk faktörlerini belirlemek için tıp verileri kullanarak bu verilere lojistik regresyon analizi uygulanmıştır. Coşkun vd., (2004), diş hekimliği verileri kullanarak protez kullanımını etkileyen önemli risk faktörlerini belirlemek için lojistik regresyon analizi kullanılmıştır. Nargeleçekenler (2005), Bursa Emniyet Müdürlüğünden alınmış olan polis suç veri tabanı kullanılarak lojistik regresyon modeli tahmin edilmeye çalışılmıştır.

Oğuzlar (2005) çalışmasında lojistik regresyon analizi kullanarak ahlak, yankesicilik ve narkotik bürolarına ilişkin verilerden yararlanılarak suçluların profilinin belirlenmesinde lojistik regresyon analizi kullanılmıştır. Oğuz ve Budak (2005), Mersin ilindeki pilot geri kazanım programından elde edilen verilerden ailelerin geri kazanım programına katılımlarına etki eden faktörleri lojistik regresyon modeli ile analiz etmiştir. Ayhan (2006) sıralı lojistik regresyon tekniği kullanarak Türkiye'deki hemşirelerin iş bırakma niyeti ile sosyo-demografik özellikler, çalışma motivasyonu ve iş memnuniyeti maddeleri arasındaki ilişkiyi araştırmıştır.

Girginer ve Cankuş (2008), yaptıkları çalışmasında Eskişehir'de toplu taşıma araçlarından biri olan tranvayın kullanımında yolcu memnuniyetinin lojistik regresyon ile analiziyle ölçülmesini incelemişlerdir. Oktay vd., (2008), Erzurum ilinde yapılan ankette verilere lojistik regresyon modeli uygulanmış ve doğal gazın talep edilmesinde etkili olan faktörler tahmin etmişlerdir. Aktaş (2009), üniversite öğrencilerin sigara içme alışkanlıklarını etkileyen faktörleri belirlemek için demografik, davranış ve risk faktörüyle ilgili tahmin çalışmalarında sıkça kullanılan lojistik regresyon ve diskriminant denklemi belirlemiştir. Çokluk (2010), yaptığı çalışmasında belirli gruplara üye olma durumunu en iyi açıklayan bağımsız değişkenler kombinasyonunu belirlemeye yönelik ikili lojistik regresyon analizinin temel kavram ve süreçlerini açıklamayı amaçlamıştır.

Karagöz vd., (2010), iş tatminini etkileyen faktörleri belirleyebilmek için lojistik regresyon analizi kullanmışlardır ve iş tatmininde etkilendiği düşünülen faktörlerin çoğunun kriz ortamında etkisini kaybettiğini belirlemişlerdir. Baş ve Çakmak (2012), işletmelerde finansal başarısızlık öngörü çalışmasında kullanılan bağımsız değişkenlerden önemli olanları gri ilişkisel analiz yardımıyla belirlemişlerdir ve bağımsız değişkenleri

sınıflandırmak için lojistik regresyon kullanmışlardır. Şerbetçi ve Özçomak (2013), üniversite öğrencilerinin istatistik ve ekonometri derslerinde başarılarını etkileyen faktörlerin belirlenmesi ve elde edilen sonuçlarla bireyin akademik eğitim kalitesini geliştirmesine katkı sağlamak amacıyla yaptıkları bu çalışmalarında lojistik regresyon kullanmışlardır.

Akın ve Şentürk (2012), Avrupa Yaşam Kalitesi Anketi aracılığıyla ordinal lojistik regresyon analizi kullanılarak sosyo-demografik özelliklere göre mutluluk düzeyleri incelenmiş ve elde edilen sonuçlar ile daha önce yapılan çalışmalar karşılaştırılmıştır. Arı ve Önder (2013), yaptıkları çalışmada lojistik regresyon ve diğer farklı regresyon modellerini karşılaştırarak bu regresyon modellerinin hangi durumlarda kullanılabileceğini incelemişlerdir.

Arı (2013), yaptığı çalışmada sıralı lojistik regresyon modellerinde bağımlı değişken kategorilerini karşılaştırmak için farklı sıralı lojit modelleri kullanmış ve paralel doğrular varsayımının sağlanmadığı durumlarda Orantısız Oran Modeli (OOM), Orantısız Olmayan Oran Modeli (OOOM) ve Kısmi Orantısız Oran Modeli (KOOM)'nin çok terimli lojit modele göre tercih edilmesi gereken modeller olduğu sonucuna varmıştır.

3. LOJİSTİK REGRESYON MODELİ

Regresyon Analizi, incelenen değişkenler arasında belirlenen ilişkilerin kuramsal olarak gösterilmesini sağlamaktadır. Biri bağımlı, diğerleri bağımsız değişken/değişkenler olarak kabul edilen ikiden çok değişken arasındaki ilişkinin uygun matematiksel denklemlerle gösterilmesi önemlidir.

Genel doğrusal regresyon modelinin matris biçiminde gösterimi,

$$Y = X\beta + \varepsilon \quad (3.1)$$

biçimindedir. Bu modelde n , gözlem sayısını; p , bağımsız değişken sayısını göstermek üzere; Y , $(n \times 1)$ boyutlu bağımlı değişken değerlerinden oluşmuş gözlem vektörünü; X , $(n \times (p+1))$ boyutlu bağımsız değişken değerlerinden oluşmuş gözlem matrisi; β , $((p+1) \times 1)$ boyutlu bilinmeyen parametreler (katsayılar) vektörü; ε , $(n \times 1)$ boyutlu hata terimleri değerlerinden oluşmuş hata vektörünü göstermektedir.

Lojistik regresyon modeli, bağımlı değişkenin ikili veya daha fazla değer aldığı ve bağımsız değişkenler ile arasındaki sebep-sonuç ilişkisini inceleyen bir analiz yöntemidir.

Doğrusal regresyon modelinde, bağımlı değişkenin değeri tahmin edilirken, lojistik regresyon analizinde bağımlı değişkenin alacağı değerlerden birinin gerçekleşme olasılığı tahmin edilmektedir (Bircan 2004).

3.1. Lojistik Regresyon ile Lineer Regresyonun İlişkisi

Lojistik regresyon ile doğrusal regresyon arasındaki en belirgin fark; lojistik regresyonda bağımlı değişkeninin düzeylerinin ikili ya da çoklu olmasıdır. Aralarındaki bu fark hem parametrik model seçimine hem de varsayımlara yansımaktadır. Fakat bu iki analiz arasında üç önemli fark vardır:

i) Doğrusal regresyonda kestirilecek bağımlı değişken sürekli ancak lojistik regresyonda bağımlı değişken kesikli değerler alır.

ii) Doğrusal regresyon analizinde bağımsız değişkenin çok değişkenli normal dağılıma sahip olma koşulu aranırken lojistik regresyonda böyle bir koşul aranmaz.

iii) Doğrusal regresyon analizinde bağımlı değişkenin değeri kestirilirken, lojistik regresyon analizinde bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı kestirilir(Bircan, 2004).

3.2. Lojistik Regresyon Modeli

Bağımlı değişkeni, sadece 0 ve 1 gibi iki mümkün değer alan bir regresyon modelini gözönünde bulunduralım. Bağımlı değişken sadece iki mümkün değer aldığından Bernoulli dağılımına sahip bir değişken olmaktadır. Bu durumda genel olarak regresyon modeli $E(\varepsilon_i) = 0$ olmak üzere,

$$y_i = x_i' \beta + \varepsilon_i \quad (3.2)$$

şeklinde yazılabilir. Burada $x_i' = [1, x_{i1}, x_{i2}, \dots, x_{ip}]$ bağımsız değişkenler matrisini, $\beta' = [\beta_0, \beta_1, \dots, \beta_p]$ bilinmeyen parametreler vektörünü ve y_i de Bernoulli rastgele değişkenini göstermektedir. Bu durumda bağımlı değişkenin alabileceği değerlere ilişkin olasılıklar Çizelge 3.1. de verildiği gibi ifade edilebilir.

Çizelge3.1. Bernoulli olasılık dağılımı

	Olasılık
0	$P(y_i=0) = 1-\pi_i$
1	$P(y_i=1) = \pi_i$

Çizelge 3.1 kullanıldığında y_i Bernoulli rassal değişkeninin beklenen değeri Eşitlik (3.3)'deki gibi elde edilebilir.

$$E(y_i) = 0.(1 - \pi_i) + 1.(\pi_i) = \pi_i \quad (3.3)$$

(3.2)' deki modelin beklenen değeri alındığında,

$$E(y_i) = E(x_i'\beta + \varepsilon_i) = E(x_i'\beta) + E(\varepsilon_i) = E(x_i'\beta) = x_i'\beta \quad (3.4)$$

$$E(y_i) = x_i'\beta = \pi_i \quad (3.5)$$

elde edilir. Bu ifade bağımlı değişken değerinin 1 olduğu durumda bu değeri alma olasılığını göstermektedir.

(3.2)' deki modele ilişkin varyans,

$$\sigma_{y_i}^2 = E(y_i^2) - [E(y_i)]^2 \quad (3.6)$$

şeklinde yazılabilir. Burada $[E(y_i)]^2$ ifadesinin elde edilmesi gerekmektedir.

$$E(y_i^2) = 0^2(1 - \pi_i) + 1^2 \pi_i = \pi_i \quad (3.7)$$

Eşitlik (3.7) de elde edilen ifade ve Eşitlik (3.5), Eşitlik (3.6)'da yerine yazılırsa

$$\sigma_{y_i}^2 = \pi_i - \pi_i^2 = \pi_i(1 - \pi_i) = E(y_i)(1 - E(y_i)) = (x_i'\beta)(1 - x_i'\beta) \quad (3.8)$$

elde edilir.

Hata teriminin beklenen değeri ve varyansı eldeedilmek istendiğinde eşitlik (3.5) yeniden düzenlenerek,

$$y_i = E(y_i) + \varepsilon_i = \pi_i + \varepsilon_i \quad (3.9)$$

şeklinde ifade edilebilir. Hata terimi ε_i bu eşitlikten çekilirse,

$$\varepsilon_i = y_i - \pi_i \quad (3.10)$$

olarak elde edilir. Hata terimi ε_i 'nin beklenen değeri,

$$E(\varepsilon_i) = E(y_i - \pi_i) = E(y_i) - E(\pi_i) = \pi_i - \pi_i = 0 \quad (3.11)$$

şeklinde elde edilir. Hata teriminin varyansı ise,

$$\text{var}(\varepsilon_i) = \text{var}(y_i - \pi_i) = \text{var}(y_i) = E(y_i)(1 - E(y_i)) \quad (3.12)$$

olarak belirlenir. Eşitlik (3.12)'den görüldüğü gibi hata teriminin varyansının, hata teriminin beklenen değerinin bir fonksiyonu olduğu görülmektedir. Bu nedenle modelde sabit

varyanslılık durumunun olmadığı görülmektedir. Bu durumda doğrusal regresyon analizi kullanılamamakta ve doğrusallaştırılarak sabit varyanslılık durumu sağlanabilmektedir (Arıcan, 2010).

Bağımlı değişkeninin alacağı değerleri (0,1) aralığına çekebilmek için dönüşüm yapılması gerekmektedir. $\pi(x)/[1-\pi(x)]$ dönüşümü bağımlı değişkenin sınırlarını (0,∞) aralığına taşımaktadır. Bağımlı değişkenin sınırlarının $(-\infty, +\infty)$ aralığında olması için ise $\pi(x)/[1-\pi(x)]$ oranının logaritmasının alınması gerekmektedir. Bu dönüşüm yardımıyla elde edilmiş olan yeni bağımlı değişken, bağımsız değişkenin doğrusal bir fonksiyonu olmaktadır. $\pi(x)$ 'i $(-\infty, +\infty)$ aralığında tanımlı hale getiren bu dönüşüm "logit dönüşüm" olarak adlandırılmaktadır (Atabey, 2010).

Logit dönüşümü,

$$\eta_i = x_i' \beta \quad (3.13)$$

$$\eta_i = \ln \left(\frac{\pi_i}{1-\pi_i} \right) \quad (3.14)$$

$$x_i' \beta = \ln \left(\frac{\pi_i}{1-\pi_i} \right) \quad (3.15)$$

(3.15) eşitliğin her iki tarafının üstel değeri alındığında,

$$\exp(x_i' \beta) = \exp \left[\ln \left(\frac{\pi_i}{1-\pi_i} \right) \right] \quad (3.16)$$

$$\exp(x_i' \beta) = \frac{\pi_i}{1-\pi_i} \quad (3.17)$$

$$\exp(x_i' \beta)(1-\pi_i) = \pi_i \quad (3.18)$$

elde edilir ve parantez içindeki ifadeler dağıtıldığında,

$$\exp(x_i' \beta) = \pi_i + \pi_i (\exp(x_i' \beta)) \quad (3.19)$$

$$\exp(x_i' \beta) = \pi_i (1 + \exp(x_i' \beta)) \quad (3.20)$$

yazılabilir. Buradan π_i değeri,

$$\pi_i = \frac{\exp(x_i \beta)}{1 + \exp(x_i \beta)} \quad (3.21)$$

olarak elde edilir. Başka bir ifade ile,

$$E(y_i) = \pi_i = \frac{e^{x_i \beta}}{1 + e^{x_i \beta}} = \frac{1}{1 + e^{-x_i \beta}} \quad (3.22)$$

şeklinde ifade edilebilir. Böylelikle lojistik regresyon modeli elde edilmiş olur.

3.3. Lojistik Regresyon Modelinde Parametre Tahmin Yöntemleri

Lojistik regresyon analizinde, bilinmeyen parametrelerin tahmininde en çok olabilirlik yöntemi kullanılmaktadır. Doğrusal regresyon analizinde kullanılan en küçük kareler yöntemi lojistik regresyon modelinde uygulandığında bağımlı değişkenin kategorik olmasından dolayı tutarlı sonuçlar vermemektedir. Parametre tahminlerinde yaygın olarak kullanılan diğer yöntem ise yeniden ağırlıklandırılmış iteratif en küçük kareler yöntemidir (Şahin, 1999).

3.3.1. En çok olabilirlik ve Newton-Raphson yöntemleri

1916 yılında Krogh tarafından kullanılan en çok olabilirlik yöntemi lojistik regresyon analizinde parametre tahmini için en sık kullanılan yöntemdir. EÇÖ yöntemi gözlemlenen veri kümesinin elde edilme olasılığını en büyük yapan bilinmeyen parametre değerlerinin tahminlerini elde etmeyi amaçlamaktadır. EÇÖ tahmin edicileri, olabilirlik fonksiyonu en büyük yapan parametre değerleri elde edecek şekilde belirlenmektedir.

Lojistik regresyon analizinde EÇÖ yöntemi uygulandığında olabilirlik fonksiyonu elde edilirken Y bağımlı değişkeninin 0,1 şeklinde iki değer alan Bernoulli dağılımına sahip olması sebebiyle, bağımsız değişkenin değeri verildiğinde Y bağımlı değişkenimizin 0' a eşit olma koşullu olasılığı $1 - \pi$ ifadesi, Y bağımlı değişkenin 1'e eşit olma koşullu olasılığı ise π ifadesi ile gösterilmektedir. Daha açık bir ifade ile $y_i = 0$ olduğunda gözlemlerin

olabilirlik fonksiyonuna katkısı $1 - \pi_i$, $y_i = 1$ olduğunda ise olabilirlik fonksiyona katkısı π_i kadardır.

Her bir gözlem Bernoulli dağılımından gelen bağımsız rassal değişkenler olarak ifade edildiğinden,

$$f(y_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad i=1,2,\dots,n \quad y_i=0,1 \quad (3.23)$$

şeklinde yazılabilir. Buradan rassal örnekleme ilişkin olabilirlik fonksiyonu Eşitlik (3.24)'deki gibi ifade edilir.

$$L(y, \beta) = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (3.24)$$

Olabilirlik fonksiyonu ile olabilirlik fonksiyonunun logaritmasını en büyük yapan parametre değerleri aynıdır. Olabilirlik fonksiyonunun logaritmasını kullanmak işlem kolaylığı sağladığından, logaritmik olabilirlik fonksiyonu daha çok tercih edilmektedir. Bu nedenle olabilirlik fonksiyonunun logaritmasını alınarak, log-olabilirlik fonksiyonu elde edilmektedir. Eşitlik (3.24)'de verilen olabilirlik fonksiyonundan hareketle log-olabilirlik fonksiyonu,

$$\ell(\beta) = \sum_{i=1}^n [y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)] \quad (3.25)$$

şeklinde elde edilir. Gerekli düzenlemeler yapıldığında,

$$= \sum_{i=1}^n y_i \ln\left(\frac{\pi_i}{1 - \pi_i}\right) + \sum_{i=1}^n \ln(1 - \pi_i) \quad (3.26)$$

ve

$$\pi_i = \frac{e^{x_i \beta}}{1 + e^{x_i \beta}} \text{ ve } 1 - \pi_i = \frac{1 + e^{x_i \beta} - e^{x_i \beta}}{1 + e^{x_i \beta}} = \frac{1}{1 + e^{x_i \beta}} = (1 + e^{x_i \beta})^{-1} \quad (3.27)$$

olduğundan log-olabilirlik fonksiyonu

$$\ell(\beta) = \sum_{i=1}^n y_i x_i' \beta - \sum_{i=1}^n \ln(1 + e^{x_i' \beta}) \quad (3.28)$$

şeklinde elde edilir. En çok olabilirlik tahmin edicilerini elde etmek için Eşitlik (3.28)' de verilen ifadenin ilgilenilen parametrelere göre türevlerinin alınması gerekmektedir. Lojistik regresyon modeli için β parametrelerine göre türev alındığında,

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^n \left(y_i - \frac{e^{x_i' \beta}}{1 + e^{x_i' \beta}} \right) x_i = \sum_{i=1}^n (y_i - \pi_i) x_i = 0 \quad (3.29)$$

elde edilir. Eşitlik (3.29)'da verilen denklemin β parametresi için çözümlenmesi gerekmektedir. Fakat elde edilen bu denklem türev işlemleriyle doğrudan çözümlenerek kapalı bir çözüm elde edilememektedir. Bu nedenle bu denklemin çözümü için iteratif yöntemler kullanılması gerekmektedir. β parametrelerinin tahmini için literatürde yaygın olarak kullanılan nümerik bir çözüm yöntemi olan Newton-Raphson yöntemi kullanılabilir. Bu amaçla, Eşitlik (3.29)'da verilen denklemin Taylor açılımını yazıldığında ilk olarak,

$$\frac{\partial \ell(\beta)}{\partial \beta'} = \sum_{i=1}^n \left(y_i - \frac{e^{x_i' \beta}}{1 + e^{x_i' \beta}} \right) x_i' \quad (3.30)$$

eşitliği ve bu eşitlikten hareketle,

$$\frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta'} = - \sum_{i=1}^n \frac{e^{x_i' \beta}}{(1 + e^{x_i' \beta})^2} x_i x_i' = - \sum_{i=1}^n \pi_i (1 - \pi_i) x_i x_i' \quad (3.31)$$

elde edilir. Burada,

$$x_i' \beta = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

$$X' = [x_1, x_2, \dots, x_n]_{(p+1) \times n}$$

$$Y' = [y_1, y_2, \dots, y_n]_{1 \times n}$$

$$P' = [\pi_1, \pi_2, \dots, \pi_n]_{1 \times n}$$

$$V = \text{diag}(\pi_i(1 - \pi_i))_{n \times n}$$

şeklinde ifade edilen vektör ve matrislerdir. Eşitlik (3.30) ve Eşitlik (3.31) yeniden düzenlendiğinde,

$$\frac{\partial \ell(\beta)}{\partial \beta} = X'Y - X'P \quad (3.32)$$

ve

$$\frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta'} = -X'VX \quad (3.33)$$

elde edilir. (3.32) ve (3.33) eşitlikleri (3.31) eşitliğinde yerine yazıldığında, lojistik regresyon modeli için EÇÖ tahmin edicileri,

$$\beta_{MLE} = \beta + (X'VX)^{-1} X'(Y - P) \quad (3.34)$$

olarak elde edilir. Bu denklem doğrudan çözülemediği için, gerçek kökü bulmak amacıyla Newton-Raphson yöntemi kullanılabilir. Bu denklemin iteratif çözümü için,

$$\beta^{(m+1)} = \beta^{(m)} - \left[\frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta'} \right]^{-1} \cdot \frac{\partial \ell(\beta)}{\partial \beta} \quad (3.35)$$

eşitliği elde edilir.

$$\beta^{(m+1)} = \beta^{(m)} + (X'V_m X)^{-1} X'(Y - P_m) \quad (3.36)$$

(3.36) eşitliğinde verilen $V_m = \text{diag}(\pi_{i,m}(1-\pi_{i,m}))$ ve $P_m' = (\pi_{1,m}, \pi_{2,m}, \dots, \pi_{n,m})$ önceki iterasyondan elde edilen köşegen matris ve sahte bağımlı olasılıklar vektörüdür. Bu eşitlik ortak paranteze alınırsa,

$$\beta^{(m+1)} = (X'V_m X)^{-1} X'V_m (X\beta^{(m)} + V_m^{-1}(Y - P_m)) \quad (3.37)$$

eşitliği yazılabilir. Burada $Q_a = X\beta^{(m)} + V_m^{-1}(Y - P_m)$ düzeltilmiş bağımlı değişken olarak tanımlandığında,

$$\beta^{(m+1)} = (X'V_m X)^{-1} X'V_m Q_a \quad (3.38)$$

elde edilir. β^m tahmin değerleri kullanılarak her iterasyonda $\pi_i(1-\pi_i)$ ağırlıkları güncellenmektedir. Newton-Raphson yöntemi ile $\beta^m, \beta^{(m+1)}$ arasındaki fark en aza indirgenecek bir noktaya gelinceye kadar iterasyona devam edilir (Arıcan, 2010). Son aşamada elde edilen $\beta^{(m+1)}$ tahmin değeri EÇÖ tahmin değeri olarak belirlenmektedir.

Lojistik modelin en çok olabilirlik kestirimlerini elde etmek için iterasyona başlarken eşitlik (3.27) deki başlangıç değerlerini belirlemek için çeşitli yaklaşımlar bulunmaktadır. Bunlardan en çok tercih edilen iki yaklaşım, diskriminant fonksiyonunun katsayılarını kullanmak veya grafiksel gösterimlerden gözle kestirimlerde bulunmaktır. Tüm iteratif çalışmalarda olduğu gibi, başlangıç değerlerinin doğruluğu iterasyon sayısı ve kestirimlerin doğruluğu üzerinde önemli etkiye sahip olmaktadır.

3.3.2.En küçük kareler yöntemi

Regresyon analizinde bilinmeyen parametreleri tahmininde en sık kullanılan yöntem en küçük kareler (EKK) yöntemidir. Bu yöntem bağımlı değişkenin koşullu ortalaması ile hata kareler toplamının minimizasyonuna dayanmaktadır. Böylelikle Y bağımlı değişkeninin koşullu dağılımının ortalaması hakkında bilgi elde edilmektedir.

$y = E(y/x) + \varepsilon$ şeklinde kurulan doğrusal regresyon modelinde gerekli varsayımlar sağlandığı durumda EKK yöntemi, bilinmeyen parametreleri istenilen özellikte tahmin edebilmektedir. Bu varsayımlar, hataların normal dağılıma sahip olması, hataların beklenen değerinin sıfır ve varyansının sabit olmasıdır. Regresyon analizinin varsayımlarının sağlanmadığı durumlarda EKK yöntemi ile yapılan tahminler etkin tahmin olma niteliği taşımamaktadır. Lojistik regresyon analizinde Y bağımlı değişkenin kesikli olması durumunda EKK yöntemi uygulanamamaktadır. Bu sebepten dolayı lojistik regresyon analizinde en küçük kareler yöntemi yerine en çok olabilirlik yöntemi kullanılmaktadır(Atabey, 2010).

3.3.3.Yeniden ağırlıklandırılmış en küçük kareler yöntemi

Parametre tahmininde kullanılan diğer yöntemlerden biri de yeniden ağırlıklandırılmış en küçük kareler yöntemidir. Bu yöntem, regresyon analizinde EKK yönteminin uygulanmamasına neden olan varsayımlarda varyansların sabit olmaması durumuyla karşılaşıldığı durumlarda doğru tahminler sağlamaktadır.

$j=1,2,\dots,J$ olmak üzere gruplandırılmış verilerde J grubun her birinde n_j denemeden r_j

başarı elde edildiğinde başarı oranı $P_j = \frac{r_j}{n_j}$ şeklinde ifade edilir. $\sigma_j^2 = Var(P_j) = P_j \left(\frac{1-P_j}{n_j} \right)$

şeklinde ifade edilen varyans değeri her biri binom dağılımlı olan gözlemler için

değişmektedir. Bu nedenle $lojit(r_j / n_j)$ 'nin bağımsız değişkenler üzerinde $w_j = \frac{n_j}{P_j(1-P_j)}$

ağırlığı ile ağırlıklandırılmasıyla regresyon analizi uygulanmaktadır (Atabey,2010). Burada

w_j ağırlık değerleri P_j 'nin bir fonksiyonudur. Bu durumda en küçük kareler yöntemi iteratif olarak uygulanmaktadır. Ağırlık değerleri her iterasyonda yeniden elde edilerek çözüme ulaşıncaya kadar iterasyona devam edilir.

3.4. Lojistik Sınıflandırma ve Katsayı Yorumları

Genellikle $P(y_i=1/x_i)$ değerini belirlemek amacıyla kullanılan lojistik regresyon modeli aynı zamanda bir sınıflandırma modeli olarak da kullanılabilir. Lojistik regresyon modelinde (3.31) eşitliği kullanılarak elde edilen değer 0,5 değerinden büyük olması durumunda $y_i=1$, bu değer 0,5 değerinden küçük olması durumunda ise $y_i=0$ olarak sınıflandırma yapılır (Tatlıdil, 1992).

İlgili regresyon katsayıları yorumlanırken, $\pi_i/(1-\pi_i)$ olarak adlandırılan odds oranı tahmininden yararlanılmaktadır (Agresti, 1996). Odds oranı değeri 1'e yakın değişkenler bağımlı değişkene önemli derecede etkide bulunan bağımsız değişkenler değildir. Bu değişkenlerin katsayıları anlamlı değil ise "önemli risk faktörü değildir" biçiminde yorumlanmaktadır. Odds oranı sıfıra yakın değerler alan regresyon katsayıları ise katsayı istatistiksel olarak anlamlı olmak koşulu ile bağımsız değişkenin önemli bir risk faktörü olduğunu fakat bağımlı değişkenin düşük değerler almasına neden olduğu negatif etkili bir

faktör olduğunu belirtir. 1" den büyük odds oranı değerleri için, regresyon katsayısı istatistiksel anlamlı olmak koşulu ile, bağımsız değişkenin önemli bir risk faktörü olduğu yorumu yapılmaktadır (Özdamar, 2004). Daha basit bir ifade ile, 1"den büyük bir odds oranının olayın gerçekleşmesinin olabilirliğinin arttığını, 1"den küçük bir odds oranı ise olayın gerçekleşmesinin olabilirliğinin azaldığını göstermektedir (Tüzüntürk, 2007).

3.5. Lojistik Regresyon Çeşitleri

3.5.1. İkili lojistik regresyon

Bağımlı değişkenin düzeyleri ikili cevap içeren regresyon modelidir. Bağımlı değişken kadın-erkek, başarılı- başarısız gibi değerler almaktadır. Bir ya da daha fazla bağımsız değişkenin ikili cevap içeren bağımlı değişken ile arasındaki ilişkiyi ortaya koymaya çalışan lojistik regresyon modelidir (Özdamar, 2004).

3.5.2. Ordinal lojistik regresyon

Bağımlı değişkenin sıralı ölçekle ölçülmüş olduğu durumlarda uygulanan bir regresyon yöntemidir. Sıralı ölçekli bağımlı değişken, en az üç kategoride gözlenen kötü, iyi, çok iyi gibi değerler içermektedir. Sıralı ölçekli veriler kodlanırken verileri doğal sıralama yapısına göre kodlamak gerekmektedir (Özdamar, 2004).

3.5.3. Nominal lojistik regresyon

Bağımlı değişkeninin isimsel olduğu durumlarda kullanılan regresyon yöntemidir. İsimsel ölçekli bağımlı değişkenin en az üç kategoride gözlenen değerler içermesi gerekmektedir ve bu kategoriler arasında doğal bir sıralama olması zorunlu değildir (Özdamar, 2004).

3.5.4. Koşullu ve sıralı lojistik regresyon

Koşullu ve sıralı lojistik regresyon analizinde bağımlı değişken olarak sıralı ikili kategorik değişken kullanılmaktadır. Sıralı ikili kategorik veri, birden fazla gruptan oluşan, her grupta iki veya daha fazla birimin olduğunu ve her grupta incelenen konu için en az bir tane olumlu cevabın olması gereken veri yapısıdır (Gürler, 1996).

3.6. Modelin Uyum İyiliğinin Değerlendirilmesi

Modelde bulunan bağımsız değişkenlerin modelin bağımlı değişkenini açıklamakta ne kadar etkin olduğunu göstermek için uyum iyiliğine bakılması gerekmektedir. Bu nedenle lojistik regresyon modelinde kurulan modelin bağımlı değişkeni açıklamada ne derecede etkin olduğunu göstermek için Hosmer-Lemeshow (H-L) testi, sınıflandırma tabloları ve yapay Cox&Snell ve Nagelkerke R2 değerleri kullanılmaktadır (Murat ve Işığışok, 2008).

3.6.1. Hosmer-Lemeshow (H-L) testi

Modelin uyum iyiliğini test etmek için David W. Hosmer ve Stanley Lemeshow tarafından geliştirilen Hosmer-Lemeshow (H-L) testi bağımlı değişkeni açıklayabilmek için oluşturulan en iyi modelin etkinliğini gösteren bir ölçüdür (Oğuzlar, 2005).

Bu testin amacı tahmin edilen olasılık değerlerini gruplandırmaktır. Uyum iyiliği test istatistiği \hat{C} , gözlenen ve beklenen frekanslardan oluşan ve grup sayısı olan g değeri ile $g \times 2$ tablosundan Pearson ki-kare istatistiği değeri olarak hesaplanır. Bu test istatistiği $(g-2)$ serbestlik derecesiyle ki-kare dağılımına sahiptir (Murat ve Işığışok, 2008).

3.6.2. Sınıflandırma tabloları

Sınıflandırma tabloları lojistik regresyonda uyum iyiliği testi için kullanılan yöntemlerden biridir. Sınıflandırma tablosunda bağımlı değişkenin gözlenen ve beklenen lojistik olasılıklardan türetilen "0" ve "1" değerleri yer almaktadır (Aksaraylı ve Saygın,

2011). Oluşturulan bu çapraz tablolar bağımlı değişkenin gözlenen gerçek değeri ile tahmin değerlerinin çaprazlanmasıyla oluşur. Sınıflandırma tablosu oluşturmak için bir (c) değeri belirlenir ve tahmin edilen oluşturulan c değer ile karşılaştırılarak uygun gruba atanması sağlanır. Tahmin edilen değer c değerini geçerse 1 grubuna, geçmez ise 0 grubuna atanır. Burada c değeri genellikle başka bir kriter yoksa 0.5 olarak alınmaktadır (Murat ve Işığışok, 2008).

Oluşturulan tabloda sınıflandırma yüzdelerinin yüksek olması, kurulan modelin sınıflandırmada oldukça başarılı olduğunu göstermektedir. Ayrıca bu durum ele alınan değişkenlerle bağımlı değişkenin uyumun iyi olduğunu göstermektedir (Aksaraylı ve Saygın, 2011).

3.6.3. Yapay Cox&Snell ve Nagelkerke R^2 değerleri

Doğrusal regresyon analizinde belirlilik katsayısı değeri olan R^2 değeri modeldeki bağımsız değişkenlerin bağımlı değişken üzerindeki etkisini ölçmektedir. Bu değer 0 ve 1 arasında değer almaktadır. R^2 değeri 1 'e ne kadar yakın olursa modelin o kadar uyumlu olduğunu belirtilmektedir. Belirlilik katsayısının bağımlı değişkenin kategorik olduğu durumda 1'e yakın değer alması sadece tüm kestirilen olasılıkların 0 veya 1'e eşit olması durumunda mümkün olabilmektedir (Özarıcı, 1996).

Klasik regresyon modellerinde hesaplanan R^2 değeri yerine lojistik regresyon modellerinde yapay R^2 değeri olan Cox&Snell ve Nagelkerke R^2 değerleri hesaplanabilmektedir (Mutlu, 2013).

Bu bölüme kadar lojistik regresyon modelinde varsayım bozulmalarının olmadığı durumda izlenecek klasik yaklaşımlara yer verilmiştir. Ancak, lojistik regresyon analizinde de regresyon analizinde olduğu gibi bağımsız değişkenler arasında çoklu iç ilişki söz konusu olabilmektedir. Ayrıca veri setinde aykırı değer bulunması modelin varsayımlarının sağlanmasına engel olmaktadır. Bu nedenle ilerleyen bölümde, lojistik regresyon analizinde bağımsız değişkenler arasında çoklu iç ilişki olması yada veri setinde aykırı değer bulunması durumunda kullanılacak alternatif lojistik regresyon yöntemleri açıklanacaktır.

4. LOJİSTİK REGRESYONDA ALTERNATİF TAHMİN YÖNTEMLERİ

Regresyon modellerinde bağımsız değişkenler arasında çoklu iç ilişki olması ve modeli oluşturacağımız veride aykırı değer olması karşılaşılabilecek sorunlar arasındadır. Regresyon modellerinde karşılaştığımız çoklu iç ilişki sorunu lojistik regresyon modelinde de ortaya çıkmaktadır. Regresyon modelinde bağımsız değişkenlerin ilişkisiz olduğu kabul edilir. Ancak bağımsız değişkenler her zaman regresyon modelinde ilişkisiz olmayabilir. Bu durumda çoklu iç ilişki sorunu ortaya çıkmaktadır (Montgomery ve Peck, 2001). Çoklu iç ilişkinin varlığı durumunda başta tahmin edicilerin minimum varyanslı olma özelliğini kaybetmesi gibi önemli sorunlar meydana gelmektedir. Çoklu iç ilişki durumunda regresyon katsayıları belirsiz ve bu katsayıların standart hataları sonsuz olmaktadır. Ayrıca regresyon katsayılarının varyans ve kovaryansları artmakta, modelin R^2 değeri oldukça yüksek çıkmaktadır. Bütün bu sonuçların aksine bağımsız değişkenlerden hiçbiri veya çok azı kısmi istatistiksel olarak anlamlı bulunmaktadır. Ayrıca, ilgili bağımsız değişkenlerin bağımlı değişkenle olan ilişkilerinin yönü kuramsal ve ampirik beklentilerle çelişebilmektedir.

Çoklu iç ilişki sorunun ortaya çıkmasında birçok neden olabilir. Bunlar arasında modelin belirlenmesi, uygulanan veri toplama metodu, aşırı tanımlanmış model ve modeldeki veya kitledeki zorluklar çoklu iç ilişki sorunun ortaya çıkma sebeplerindendir (Montgomery ve Peck, 2001).

Çoklu iç ilişkinin varlığını belirlemede çeşitli yöntemler bulunmaktadır. Geliştirilen bu yöntemler arasında korelasyon matrisi, varyans şişirme faktörü, $X'X$ matrisinin öz değerlerinin incelenmesiyle çoklu iç ilişki probleminin varlığı tespit edilebilmektedir. Bu problemin üstesinden gelebilmek için çeşitli yanlı regresyon tahmin yöntemleri önerilmiştir. Lojistik regresyon modelinde çoklu bağıntı problemi olduğu durumda en çok olabilirlik yöntemine alternatif olarak önerilen yöntemlerden biri de ridge regresyondur.

Regresyon modelinde karşılaşılan diğer bir sorun ise veri setinde aykırı değer/değerlerin olmasıdır. Veri setinden uzak olan değere aykırı değer denir. Oluşturulan regresyon modelinde aykırı değerlerin olması durumunda lojistik regresyonda bağımlı değişkenin 0-1 değerlerini alma olasılıkları modellediğinden modelde elde edilen başarı olasılığı küçük olduğu halde gözlenen durum başarılı, elde edilen başarı olasılığı yüksek olduğu halde gözlenen durumun başarısız olarak belirlenmesine neden olmaktadır (Yavuzkanat, 2011). Aykırı değer problemini en aza indirmek için bu gözlemlerden fazla

etkilenmeyen robust yöntemlere başvurulmalıdır. Robust lojistik regresyon yöntemi kullanılarak modelin tahmini gerçekleştirilebilir.

Çalışmanın bu kısmında yukarıda açıklanan bu iki önemli probleme çözüm olabilecek yöntemler ele alınacaktır. Öncelikle bağımsız değişkenler arasında çoklu iç ilişki varlığı durumu üzerinde durularak, yanlı regresyon modeli açıklanacaktır. Daha sonra aykırı değerlerin varlığı durumunda alternatif bir yöntem olan robust lojistik regresyon yöntemi kısaca açıklanacaktır.

4.1. Lojistik Ridge Regresyon

Hoerl ve Kennard (1970) tarafından önerilen ridge regresyon çoklu iç ilişki sorunu ile karşılaşıldığında kullanılan yanlı regresyon yöntemlerinden birisidir. Yanlı regresyon tahmin edicileri arasında ridge tahmin edicisi en yaygın olarak kullanılan tahmin edicidir.

Çoklu iç ilişki sorunu ile karşılaşıldığında yapılan analizler tahmin edicinin varyans değerini olduğundan büyük olarak belirlemekte ve regresyon katsayıları gerçek değerlerinden farklı çıkmaktadır. Bu duruma çözüm olarak Hoerl ve Kennard, en küçük kareler tahmin edicisinin $X'X$ matrisinin köşegen elemanlarına bir k sabiti ekleyerek ridge tahmin edicisini önermişlerdir. Ridge tahmin edicisi yansızlıktan ödün verdiği için ele alınan parametreleri yanlı olarak tahmin etmektedir. Fakat tahmin ediciler artık daha küçük varyansa sahip olmaktadır. Çoklu doğrusal regresyon için Hoerl ve Kennard (1970) tarafından önerilen Ridge tahmin edicisi aşağıdaki şekilde ifade edilmiştir.

$$\hat{\beta}_R = (X'X + kI)^{-1} X' y \quad (4.1)$$

Eşitlik (4.1) k ($0 < k < 1$) yanlılık sabiti veya daraltma parametresi olarak da adlandırılmaktadır ve k sabiti seçimi Ridge tahmin edicisinin performansını etkilemektedir (Hoerl ve Kennard, 1970).

Bağımlı değişkenin iki düzeyli olduğu durumda (4.1) de verilen tahmin edici kullanılamaz. Bu durumda kullanılacak lojistik ridge regresyon tahmin edicisi aşağıdaki gibi belirlenebilir. Bağımlı değişken iki düzeyli ve Bernoulli dağılımına sahip olduğundan

$$f(y_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad i=1,2,\dots,n \quad y_i=0,1 \quad (4.2)$$

yazılabilir. Buradan en çok olabilirlik fonksiyonu,

$$L(y, \beta) = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (4.3)$$

elde edilir. En çok olabilirlik fonksiyonu doğrusal olmadığından ve logaritmik fonksiyonla maksimum oldukları nokta değişmediğinden, log-olabilirlik fonksiyonu,

$$\ell(\beta) = \sum_{i=1}^n [y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)] \quad (4.4)$$

$$= \sum_{i=1}^n y_i \ln\left(\frac{\pi_i}{1 - \pi_i}\right) + \sum_{i=1}^n \ln(1 - \pi_i) \quad (4.5)$$

olarak yazılabilir.

$$\pi_i = \frac{e^{x_i \beta}}{1 + e^{x_i \beta}} \quad 1 - \pi_i = \frac{1 + e^{x_i \beta} - e^{x_i \beta}}{1 + e^{x_i \beta}} = \frac{1}{1 + e^{x_i \beta}} = (1 + e^{x_i \beta})^{-1}$$

olduğundan

$$= \sum_{i=1}^n y_i x_i \beta - \sum_{i=1}^n \ln(1 + e^{x_i \beta}) \quad (4.6)$$

elde edilir.

Dufy ve Santer (1989) lojistik regresyon modeli için en çok olabilirlik yöntemini kullanarak ridge lojistik regresyon tahmin edicisini önermişlerdir. Bu tahmin edicinin Schaefer ve ark. (1984) tarafından sunulan Ridge lojistik tahmin edicisi ile asimtotik olarak denk olduğunu göstermişlerdir (Arıcan, 2010).

$$\ell(\beta) = \sum_{i=1}^n y_i (\ln\left(\frac{\pi_i}{1 - \pi_i}\right)) - \sum_{i=1}^n \ln(1 + e^{x_i \beta}) \quad (4.7)$$

Dufy ve Santer (1989) bu log-olabilirlik fonksiyonundan minimum yanlılık üretecek şekilde $\|\beta\|^2$ 'yi (yani β 'nın normunu) d gibi bir sabit sayı ile çarparak çıkarmışlardır.

$$\ell^\lambda(\beta) = \ell(\beta) - d\|\beta\|^2 \quad (4.8)$$

$$= \sum_{i=1}^n \left[y_i(x_i'\beta) - \ln(1 + e^{x_i'\beta}) \right] - d\beta'\beta \quad (4.9)$$

Buradan

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^n \left(y_i x_i' - \frac{x_i' e^{x_i \beta}}{1 + e^{x_i \beta}} \right) = \sum_{i=1}^n \left(y_i - \frac{e^{x_i \beta}}{1 + e^{x_i \beta}} \right) x_i' \quad (4.10)$$

$$= \sum_{i=1}^n (y_i - \pi_i) x_i' \frac{\partial \ell^\lambda(\beta)}{\partial \beta} = \frac{\partial \ell(\beta)}{\partial \beta} - \frac{\partial (d\beta'\beta)}{\partial \beta} \quad (4.11)$$

$$= \sum_{i=1}^n (y_i - \pi_i) x_i' - 2d\beta \quad (4.12)$$

$$= \sum_{i=1}^n \left(x_i' y_i - x_i' \frac{e^{x_i \beta}}{1 + e^{x_i \beta}} \right) - 2d\beta \quad (4.13)$$

$$\frac{\partial \ell^\lambda(\beta)}{\partial (\beta)} = \sum_{i=1}^n x_i' (y_i - \pi_i) - 2d\beta \quad (4.14)$$

(4.15) eşitliğinin β ' ya göre ikinci kısmi türevini alırsak

$$\frac{\partial^2 \ell^\lambda(\beta)}{\partial \beta^2} = \left[\sum_{i=1}^n x_i' y_i - x_i' \frac{e^{x_i \beta}}{1 + e^{x_i \beta}} \right]' - (2d\beta)' \quad (4.15)$$

$$= - \sum_{i=1}^n \left[\frac{x_i' e^{x_i \beta}}{(1 + e^{x_i \beta})^2} x_i - 2dI \right] \quad (4.16)$$

burada I birim matrisi olmak üzere,

$$= - \sum_{i=1}^n x_i' \pi_i (1 - \pi_i) x_i - 2dI \quad (4.17)$$

bu ifade

$$V(\beta) = \text{diag}(\pi_i(1-\pi_i)) \quad (4.18)$$

olduğundan matris formuyla yazılırsa

$$-\frac{\partial^2 \ell^\lambda(\beta)}{\partial \beta^2} = X'V(\beta)X + 2dI \quad (4.19)$$

elde edilir.

$V(\beta) \rightarrow \beta$

$\ell^\lambda(\beta) = \ell(\beta) - d\|\beta\|^2$ 'nin birinci dereceden β 'ya göre kısmi türevinin, Taylor serisi ile β parametresinin gerçek bir değeri olan β_0 etrafında birinci dereceden açılımı

$$\left. \frac{\partial \ell^\lambda(\beta_{rid})}{\partial \beta} \right|_{\beta_{rid}} = \left. \frac{\partial \ell^\lambda(\beta)}{\partial \beta} \right|_{\beta_0} - (\beta_{rid} - \beta_0)' \frac{\partial^2 \ell^\lambda(\beta)}{\partial \beta^2} + (\|\beta_{rid} - \beta_0\| \quad (4.20)$$

$$= \left. \frac{\partial \ell^\lambda(\beta)}{\partial \beta} \right|_{\beta_0} - (\beta_{rid} - \beta_0)' (X'V(\beta)X + 2dI) \quad (4.21)$$

$$\left. \frac{\partial \ell^\lambda(\beta)}{\partial \beta} \right|_{\beta_0} = U(\beta_0) \quad (4.22)$$

şeklinde yazılabilir.

$$U^\lambda(\beta_0) = U(\beta_0) - 2d\beta_0 \quad (4.23)$$

olmak üzere

$$\left. \frac{\partial \ell^\lambda(\beta_{rid})}{\partial \beta} \right|_{\hat{\beta}_{rid}} = [U(\beta_0) - 2d\beta_0] - (\beta_{rid} - \beta_0)' [X'V(\beta_0)X + 2dI] \quad (4.24)$$

elde edilir. Burada

$$U(\beta_0) = \sum_{i=1}^n x_i' \left(y_i - \frac{e^{x_i \beta_0}}{1 + e^{x_i \beta_0}} \right) \quad (4.25)$$

dır.

$$\left. \frac{\partial \ell^{\lambda}(\beta_{rid})}{\partial \beta} \right|_{\hat{\beta}_{rid}} = \mathbf{0} \text{ alınarak } \hat{\beta}_{rid} \text{ için birinci sıram yaklaşımı yapılır. (4.24) eşitliği}$$

sıfıra eşitlendikten sonra

$$0 = [U(\beta_0) - 2d\beta_0] - (\hat{\beta}_{rid} - \beta_0)' [X'V(\beta_0)X + 2dI] \quad (4.26)$$

elde edilir.

Sağdan $[X'V(\beta_0)X + 2dI]^{-1}$ ile çarpıldıktan sonra eşitliğin her iki tarafının transpozu alınırsa

$$U(\beta_0) - 2d\beta_0 = \hat{\beta}_{rid}' [X'V(\beta_0)X + 2dI] - \beta_0' [X'V(\beta_0)X + 2dI] \quad (4.27)$$

$$[X'V(\beta_0)X + 2dI]^{-1} \cdot [U(\beta_0) - 2d\beta_0] = \hat{\beta}_{rid} - \beta_0 \quad (4.28)$$

yazılabilir ve

$$\hat{\beta}_{rid} = \beta_0 + [X'V(\beta_0)X + 2dI]^{-1} [U(\beta_0) - 2d\beta_0] \quad (4.29)$$

olarak elde edilir. Buradan

$$\hat{\beta}_{rid} = [X'V(\beta_0)X + 2dI]^{-1} \cdot [U(\beta_0) + (X'V(\beta_0)X + 2dI)\beta_0 - 2d\beta_0] \quad (4.30)$$

$$= [X'V(\beta_0)X + 2dI]^{-1} [U(\beta_0) + [X'V(\beta_0)X\beta_0 + 2d\beta_0 - 2d\beta_0]] \quad (4.31)$$

$$\hat{\beta}_{rid} = [X'V(\beta_0)X + 2dI]^{-1} [U(\beta_0) + (X'V(\beta_0)X)\beta_0] \quad (4.32)$$

birinci sıra tahmin edicisi bulunmuş olur. Bu ifadeye $d=0$ alınarak Eşitlik (4.32) düzenlenirse

$\hat{\beta}_{MLE}$ elde edilir.

$$\hat{\beta}_{MLE} = [X'U(\beta_0)X]^{-1} [U(\beta_0) + X'U(\beta_0)X\beta_0] \quad (4.33)$$

$$= [X'U(\beta_0)X]^{-1} U(\beta_0) + \beta_0 \quad (4.34)$$

β_0 eşitliğin diğer tarafına geçirildiğinde,

$$\hat{\beta}_{MLE} - \beta_0 = [X'U(\beta_0)X]^{-1} U(\beta_0) \quad (4.35)$$

yazılabilir. Eşitlik (4.35) de verilen eşitliğin her iki tarafı $[X'U(\beta_0)X]$ ile çarpılırsa

$$U(\beta_0) = [X'U(\beta_0)X] \cdot [\hat{\beta}_{MLE} - \beta_0] \quad (4.36)$$

elde edilir. (4.32) eşitliğinde $U(\beta_0)$ yerine (4.36) eşitliği yerine yazılırsa

$$\hat{\beta}_{rid} = [X'U(\beta_0)X + 2dI]^{-1} [U(\beta_0) + X'U(\beta_0)X(\beta_0)] \quad (4.37)$$

$$= [X'U(\beta_0)X + 2dI]^{-1} [X'U(\beta_0)X(\hat{\beta}_{MLE} - \beta_0) + X'U(\beta_0)X(\beta_0)] \quad (4.37)$$

$$= [X'U(\beta_0)X + 2dI]^{-1} [X'U(\beta_0)X\hat{\beta}_{MLE} - X'U(\beta_0)X(\beta_0) + X'U(\beta_0)X(\beta_0)] \quad (4.38)$$

$$= [X'U(\beta_0)X + 2dI]^{-1} [X'U(\beta_0)X\hat{\beta}_{MLE}] \quad (4.40)$$

elde edilir. $2d=k$ olmak üzere Ridge tahmin edicisi

$$\hat{\beta}_{ridge} = [X'U(\beta_0)X + kI]^{-1} [X'U(\beta_0)X\hat{\beta}_{MLE}] \quad (4.41)$$

elde edilmiş olur.

4.2. Robust Lojistik Regresyon Yöntemi

İstatistiksel olarak yapılan çalışmalarda incelenen bazı gözlemlerin diğer gözlemlere göre çok büyük ya da çok küçük değerler aldığı durumlar olabilmektedir. Bu tür gözlemler aykırı değer olarak tanımlanmaktadır. Eğer bir gözlem veri setinden silindiğinde regresyon parametre kestiriminde önemli bir farklılığa sebep oluyor ise bu gözleme etkin gözlem denilmektedir. Etkin gözlemler veri setindeki diğer gözlemlerle karşılaştırıldığında regresyon denklemini bireysel olarak veya topluca etkileyen gözlemlerdir. Regresyon analizinde gözlemler aykırı olarak nitelenebilir ancak söz konusu gözlemler regresyon katsayıları ya da model üstünde anlamlı bir değişikliğe yol açmıyorsa, bu gözlemlerin etkili gözlem olduğu söylenemez (Yavuzkanat, 2011). Aykırı değerler etkin gözlem olmak zorunda değildir ve etkin gözlemlerde aykırı değer olmak zorunda değildir.

Aykırı değerlerin varlığında kurulan lojistik regresyonda her bir gözlemin ayrıntılı olarak incelenmesi ve aykırı değerlerin belirlenmesi gerekir. Tek bir aykırı değer kurulan model ya da parametre tahminleri üzerinde önemli etkiye sahiptir. Aykırı değerlerin olması durumunda modelin varsayımları sağlanamamaktadır. Model biçimi doğru oluşturulamamakta ve tahmin edilen parametre değerleri gerçeği yansıtmamaktadır. Ayrıca EÇO yöntemi ile elde edilen parametre kestirimleri, ideal veri setleri için iyi sonuçlar verirken aykırı değerler bulunması durumunda oldukça hassastır (Hadi ve Imon, 2008). Aykırı değerlerin bulunması durumunda modeli doğru tahmin edebilmek için bu değerlerin etkisini en aza indirmeye çalışan robust yöntemlere başvurulmaktadır.

Lojistik regresyonda bağımlı değişkenin 0,1 değerini alma olasılığı hesaplanırken aykırı değerlerin bulunması halinde tahmin edilen başarı olasılığı küçük olduğu durumda gözlenen durum başarılı, başarı olasılığı yüksek olduğu durumda gözlemler başarısız olarak değerlendirilebilmektedir. Bu durumda öncelikle varsa aykırı değerler belirlenerek, bu durum için gereken robust yaklaşımlar kullanılmaktadır. Lojistik regresyonda analizinde de aykırı değerlerin tespiti için çeşitli yaklaşımlar geliştirilmiştir. Bunlar arasında standartlaştırılmış artıklar, Pearson artıklar, standart Pearson artıklar yöntemleri yaygın olarak kullanılan yöntemlerdir (Yavuzkanat, 2011). Bu yaklaşımlardan herhangi biri yardımıyla lojistik regresyon analizinde artık değerlerin hangi gözlemler olduğu tespit

edilebilmektedir. Saptanan aykırı değerlerin veriden çıkartılması söz konusu olabilir. Ancak bu durum veri kaybına neden olmakta ve küçük örneklem durumunda ortaya çıkabilecek model belirleme hataları ile karşılaşılma olasılığını arttırmaktadır. Bunun yerine aykırı değerlerin etkisini azaltan robust yöntemler kullanılarak lojistik regresyon modeli oluşturulabilir.

Çalışmanın bu kısmında veri setinde aykırı değer bulunması durumunda kullanılacak alternatif bir robust Lojistik regresyon yöntemi ele alınmıştır. Literatürde en çok kullanılan robust lojistik regresyon tahmin edicisi, Bianco ve Yohai (1996) tarafından geliştirilen ve bir M sınıfı tahmin edici olan Bianco-Yohai Tahmin edicisidir. Bu tahmin edici, Eşitlik (4.42)'de tanımlandığı gibidir. Burada amaç fonksiyonu,

$$\beta_n = \arg \min \sum_{i=1}^n \phi(Y_i, \pi(x_i' \beta)) \quad (4.42)$$

ve

$$\phi(Y_i, \pi(x_i' \beta)) = \tilde{n}(d_i(\beta)) + \tilde{n}_0(\pi(x_i' \beta)) \quad (4.43)$$

şeklinde ifade edilmektedir. $d_i(\beta)$, β 'ların varyanslarını göstermekte, $\tilde{n}(t)$ sınırlandırılmış fonksiyon,

$$\tilde{n}(0) = 0 \text{ ve } \tilde{n}'(t) = (1-t)I(0 < t < 1) \quad (4.44)$$

ve dengeleyici fonksiyon

$$\tilde{n}_0(p) = \tilde{n}_1(p) + \tilde{n}_1(1-p) \quad (4.45)$$

olmak üzere, eşitsizlikte \tilde{n}_1 , \tilde{n}' ya bağlı olarak

$$\tilde{n}_1(p) = \int_0^p \tilde{n}'(-\ln t) dt, \quad p \in (0,1) \quad (4.46)$$

şeklinde elde edilmektedir (Hobza vd., 2012). BY tahmin edicisi yukarıda da görüldüğü gibi iteratif bir yöntemdir.

5.UYGULAMA

5.1. Giriş

Ülkelerin gelişmişlik düzeyleri, her ülkeyi ilgilendiren en önemli konuların başında gelmektedir. Kısa süre öncesine kadar, ülkelerin gelişmesinde ekonomik göstergeler daha önemli iken, 1990 yılından bu yana yayınlanan insani gelişme raporları (İGR) daha etkili hale gelmiştir. Bu raporlar dünya ülkelerine yol gösterici olmakta ve gelişmişlik konusunda öncelikleri kısmen de olsa değiştirme yolunda önemli adımlar atılmasına sebep olmaktadır. Raporların içeriğinde, ülkeleri karşılaştırmalı olarak inceleyen güncelleştirilmiş insani gelişme göstergelerini (İGG) yer almaktadır. İGG, temeli insan olan bakış açısı ile ülkelerin gelişmişlik düzeylerini ölçen bir endekstir. Ekonomik büyüme ve üretime dayalı olan göstergelerden farklı olarak, bu endeks ile refah seviyesi ülke bazında ölçülmek yerine, ülkelerde yaşayan insanlar açısından ölçülmektedir. Böylelikle, ülkelerin gelişmişlik düzeyi belirlenirken temeli insan olan bir bakış açısı meydana getirilmektedir.

İnsani gelişme endeksi (İGE), üç göstergeye dayanmaktadır: Doğumda yaşam beklentisi olarak ölçülen ömür; yetişkinlerin okur-yazarlık ve ilk, orta ve yüksek eğitimde okullaşma oranı ile birlikte ölçülen eğitim düzeyi ve satın alma gücü paritesine göre kişi başına gayri safi milli gelir ile ölçülen yaşam standardıdır (Human Development Report, 2010). Son yıllardaki çalışmalar ile bu endeksin hesaplanmasında bazı değişikliklere gidilmiştir.

2000’li yıllarda, İnsani Gelişmişliğin eksik yanı olarak sosyal göstergelerin de sürece eklenmesi ihtiyacı doğmuş ve bu nedenle sosyal gelişmişlik endeksi yaratılmıştır. Bu tarihten sonra hesaplanan sosyal gelişmişlik endeksi (SGE) ise aşağıda ifade edilen üç göstergeye dayanmaktadır: Temel insani ihtiyaçlar, refahın temelleri ve fırsatlar. Bu üç alt endeksin ortalaması alınarak SGE’si hesaplanmaktadır. Temel insani ihtiyaçlar, bir ülkede insanların varlıklarını sürdürebilmeleri için gerekli olan beslenme, barınma, suya erişim ve güvenlik gibi temel ihtiyaçların ne kadar karşılandığını ölçmektedir. Refahın temelleri alt endeksi ise bireylerin ve toplumların refahlarını arttırmalarını sağlayan temel eğitim, bilgiye ve iletişim kanallarına erişim, sağlıklı yaşam ve çevresel sürdürülebilirlik değerlerini barındırmaktadır.

5.2. Sosyal Gelişmişlik Endeksi

Sosyal Gelişmişlik Endeksi, daha fazla birey odaklı bir gelişmişlik ölçütünden yola çıkarak dünyada kalkınma ile ilgili tartışmayı şekillendirmek ve bir ülkede yaşayan vatandaşların hayat kalitesini ilgilendiren kararları daha bilgili ve sağlıklı bir şekilde almalarına katkı vermeyi amaçlamayan bir gelişmişlik endeksidir.

SGE, bir ülkenin vatandaşlarının temel ihtiyaçlarını karşılama, onların hayat kalitelerini sürdürüp iyileştirebilecekleri ortamı oluşturma ve bireylerin potansiyellerini tam olarak değerlendirebilecekleri koşulları sağlama kapasitesini ölçmektedir. SGE hesaplanırken, bir ülkenin milli geliri ile birlikte temel insani ihtiyaçları, refah ihtiyaçları, temel tıp, su, barınak ihtiyaçları ve bilgi, iletişim ihtiyaçları ile kişisel hak ve özgürlükler, eğitim ihtiyaçları adı verilen alt endeks değerleri hesaba katılmaktadır. Sadece ekonomik göstergeler ile eğitim durumlarını göz önüne alan diğer endekslerden bu yönüyle farklıdır. Bu endeks, bir bütün olarak bir ülkede yaşayan vatandaşların gelişmişlik düzeyini hesaplamaya çalışmaktadır. Lars Osberg (2001), Needs and Wants: What is Social Progress and How should it be measured adlı çalışmasında sosyal gelişmişliğin tanımını yapmış ve endeksin nasıl oluşturulması gerektiğine dair ipuçları vermiştir. Konya Valiliği (2011), Temel Bileşenler Analizi ile Konya Karaman Sosyal Gelişmişlik endeksine ait bir çalışma yapmıştır. Porter ve ark. (2013), son yıllarda tüm dünyanın kabul ettiği Sosyal Gelişmişlik Endeksini oluşturmuşlardır. Ayrıca oluşturdukları bu endeks değerleri her yıl düzenli olarak rapor halinde sunulmaktadır. Ali Sökmen'in (2014) yılındaki Sosyal Gelişmişlik Türkiye için ne ifade ediyor çalışmasında bu endeksi ayrıntısıyla tanıtılmaya çalışılmıştır.

Bu çalışmada ülkelerin gelişmişlik düzeyini hesaplamada alt endeks değerleri hesaba katılarak Lojistik Regresyon analizi uygulanmıştır. Çalışmanın başında veri seti incelendiğinde açıklayıcı değişkenler arasında çoklu iç ilişki ve aykırı değer tespit edilmiştir. Model varsayımları sağlanmadığında R^2 ye bağlı performans ölçüleri doğru bir karşılaştırma ölçütü olmamaktadır. Çalışmada, lojistik, lojistik Ridge ve robust lojistik regresyon modellerinin doğru sınıflandırma performanslarını karşılaştırmak amacıyla inceleme konusu olan veri seti öncelikle iki kısma ayrılmıştır. 132 ülkeden seçilen 80 ülkenin alt endeks değerleri eğitim verisini, geriye kalan 52 ülkenin alt endeks değerleri test verisini oluşturmaktadır. Eğitim verisinin boyutu ilgili regresyon modellerinin kurulmasına izin verecek şekilde tespit edilmiştir. Daha az boyuttaki bir veri seti model kurulmasına olanak

sağlamamıştır. Geriye kalan 52 ülkeye ait veri ise test verisi olarak belirlenmiştir. Lojistik regresyon analizi için modelin genel olarak anlamlılığını ve doğru sınıflandırma oranı SPSS 18 paket programı ile belirlenmiştir. Aynı sonuçlar, R paket programıyla da elde edilebilmektedir. Veri setinde çoklu iç ilişki ve aykırı değer problemi belirlendiği için, çoklu iç ilişkinin giderilmesi için yanlı regresyon yöntemine, aykırı değerler tespit edildiğinden dolayı robust lojistik regresyon yöntemine başvurulmuştur. Her iki yaklaşım için analizler R paket programında yazılan kodlar ile yapılmıştır.

İlk aşamada bağımlı ve bağımsız değişkenler aşağıdaki gibi belirlenerek, analiz bu değişkenler ile sürdürülmüştür.

Bağımlı değişken

$$y = \begin{cases} 0, \text{Sosyal Gelişmişlik durumu yüksek (endex değeri 0,5'den büyük)} \\ 1, \text{Sosyal Gelişmişlik durumu düşük (endex değeri 0,5'den küçük)} \end{cases}$$

Bağımsız değişkenler

$x_1 = \text{MG: Milli Gelir}$

$x_2 = \text{Tİİ: Temel İnsan İhtiyaçları}$

$x_3 = \text{Rİ: Refah İhtiyaçları}$

$x_4 = \text{FR: Fırsat}$

$x_5 = \text{TTİ: Temel Tıp İhtiyaçları}$

$x_6 = \text{SİH: Su İhtiyaçları}$

$x_7 = \text{BİH: Barınak İhtiyaçları,}$

$x_8 = \text{Gİ: Güvenlik İhtiyaçları}$

$x_9 = \text{Bİ: Bilgi İhtiyaçları}$

$x_{10} = \text{BİE: Bilgi ve İletişime Erişim}$

$x_{11} = \text{Sİ: Sağlık İhtiyaçları}$

$x_{12} = \text{ES: Ekosistem Sürdürülebilirliği}$

$x_{13} = \text{KH: Kişisel Haklar}$

x_{14} = KÖ: Kişisel Özgürlük

x_{15} = TOL: Tolerans

x_{16} = İE: İleri Eğitim

olarak seçilmiştir.

Eğitim verisi olarak belirlenen 80 ülkeden alınan alt endeks değerleri için tanımlayıcı istatistik bilgileri Çizelge 5.1. 'de verilmiştir. Çizelge 5.1. incelendiğinde Y bağımlı değişkeninin çoğunlukla 0 değerini aldığı, bir başka ifade ile veri setinde çoğunlukla sosyal gelişmişlik endeks değeri yüksek ülkelerin yer aldığı görülmektedir.

Çizelge 5.1. Tanımlayıcı istatistikler

Değişkenler	Minimum	1. Kartil	Medyan	Ortalama	3. Kartil	Maksimum
y	0.000	0.000	0.000	0.01	0.000	1.000
MG	0	2759	4482	8262	8447	46547
Tİİ	25.94	65.52	67.80	68.76	75.96	94.87
Rİ	42.42	57.51	61.22	63.39	69.01	89.78
FR	24.30	38.30	42.80	47.03	51.15	88.01
TTİ	27.60	84.95	88.00	85.46	92.53	98.78
SİH	5.00	57.75	67.00	76.67	96.00	100.00
BİH	6.00	57.75	60.00	60.55	71.25	91.00
Gİ	28.00	54.50	59.00	59.33	64.00	94.00
Bİ	28.00	86.00	88.00	85.26	93.00	99.49
BİE	13.00	44.75	50.50	54.71	62.25	98.00
Sİ	50.00	69.00	72.00	71.34	75.00	82.78
ES	8.00	28.75	36.50	42.35	58.50	86.13
KH	2.00	21.00	30.00	39.80	51.75	98.80
KÖ	34.00	60.00	62.00	64.72	68.25	94.00
TOL	21.00	36.00	39.00	44.67	53.00	88.00
İE	5.00	31.00	34.00	37.28	43.25	82.00

Çizelge 5.1. de yer alan açıklayıcı değişkenler arasında çoklu iç ilişkinin varlığının tespiti için öncelikle eğitim verisi üzerinden korelasyon matrisi oluşturulmuştur. Korelasyon değerlerine ilişkin sonuçlar Çizelge 5.2.'de verilmiştir.

Çizelge 5.2 Korelasyon matrisi

	MG	Tii	Ri	FR	Tii	SiH	BiH	Gi	Bi	BiE	Si	ES	KH	KÖ	TOL	İE
MG	1	,755**	,791**	,823**	,539**	,710**	,682**	,713**	,511**	,839**	,422**	,515**	,671**	,730**	,771**	,785**
Tii	,755**	1	,828**	,756**	,902**	,923**	,908**	,805**	,865**	,811**	,298**	,379**	,558**	,749**	,631**	,842**
Ri	,791**	,828**	1	,877**	,686**	,794**	,839**	,647**	,719**	,921**	,469**	,768**	,781**	,855**	,753**	,757**
FR	,823**	,756**	,877**	1	,576**	,717**	,738**	,661**	,620**	,888**	,444**	,580**	,915**	,874**	,848**	,819**
Tii	,539**	,902**	,686**	,576**	1	,779**	,834**	,630**	,876**	,654**	,195*	,216*	,360**	,630**	,449**	,739**
SiH	,710**	,923**	,794**	,717**	,779**	1	,832**	,617**	,798**	,776**	,214*	,430**	,574**	,676**	,595**	,773**
BiH	,682**	,908**	,839**	,738**	,834**	,832**	1	,638**	,847**	,774**	,378**	,433**	,554**	,758**	,621**	,775**
Gi	,713**	,805**	,647**	,661**	,630**	,617**	,638**	1	,573**	,679**	,296**	,298**	,496**	,616**	,578**	,703**
Bi	,511**	,865**	,719**	,620**	,876**	,798**	,847**	,573**	1	,655**	,152	,226**	,417**	,671**	,455**	,752**
BiE	,839**	,811**	,921**	,888**	,654**	,776**	,774**	,679**	,655**	1	,334**	,596**	,817**	,804**	,771**	,779**
Si	,422**	,298**	,469**	,444**	,195*	,214*	,378**	,296**	,152	,334**	1	,331**	,409**	,481**	,504**	,194*
ES	,515**	,379**	,768**	,580**	,216*	,430**	,433**	,298**	,226**	,596**	,331**	1	,608**	,561**	,515**	,358**
KH	,671**	,558**	,781**	,915**	,360**	,574**	,554**	,496**	,417**	,817**	,409**	,608**	1	,741**	,741**	,598**
KÖ	,730**	,749**	,855**	,874**	,630**	,676**	,758**	,616**	,671**	,804**	,481**	,561**	,741**	1	,726**	,746**
TOL	,771**	,631**	,753**	,848**	,449**	,595**	,621**	,578**	,455**	,771**	,504**	,515**	,741**	,726**	1	,631**
İE	,785**	,842**	,757**	,819**	,739**	,773**	,775**	,703**	,752**	,779**	,194*	,358**	,598**	,746**	,631**	1

** .Korelasyon 0.01 anlam düzeyinde anlamlı. * .Korelasyon 0.05 anlam düzeyinde anlamlı

Çizelge 5.2. incelendiğinde, bağımsız değişkenler arasındaki korelasyon değerlerinin büyük kısmının 0,75 ten büyük olduğu görülmektedir. Bu çoklu iç ilişkinin varlığının bir göstergesi olarak nitelendirilmektedir. Ayrıca modelin R^2 değerinin yüksek olması da bir çoklu iç ilişki göstergesi olarak değerlendirilmektedir. Bu nedenle lojistik regresyon modeli için R^2 değerleri hesaplanmış ve Çizelge 5.3.'te verilmiştir.

Çizelge 5.3. Bağımsız değişkenlerin bağımlı değişkeni açıklama oranları

Model Özeti			
Adım	-2 Loglikelihood değeri	Cox&Snell R Square değeri	Nagelkerke R Square değeri
1	0,000	0,478	1,000

Çizelge 5.3. incelendiğinde modeldeki bağımsız değişkenlerin, bağımlı değişkeni açıklama oranı; Cox&Snell'e göre %48, Nagelkerke'ye göre %100'dür. Yüksek korelasyon ve yüksek R^2 değerleri tek başına çoklu iç ilişki tespiti için yeteri kadar güçlü göstergeler değildir. Bu nedenle modelin anlamlılık sınaması yapılmış ve modele ilişkin sonuçlar Çizelge 5.4. ve Çizelge 5.5.'te verilmiştir.

Çizelge 5.4. Hosmer-Lemeshow genel anlamlılık testi

Hosmer-Lemeshow testi			
Adım	χ^2 değeri	Serbestlik derecesi	Olasılık değeri
1	0,000	8	1,000

Hosmer-Lemeshow testi " H_0 : Model uygundur" ve " H_1 : Model uygun değildir" hipotezleri altında modelin uyum iyiliğinin değerlendirilmesi amacıyla yapılmaktadır. Çizelge 5.4.' de görüldüğü gibi Hosmer-Lemeshow test istatistiği değeri \hat{C} , 0,000 olarak hesaplanmıştır. Modelin hesaplanmış olan \hat{C} test istatistiği değeri, %99 güvenirlilik seviyesinde 8 serbestlik derecesi için χ^2 test istatistiğinin tablo değeri olan $\chi_{0,01;8}^2$ ile karşılaştırıldığında H_0 hipotezi kabul edilir. Böylece, modelin genel olarak anlamlı olduğuna belirlenmiştir.

Lojistik regresyon analizi kullanılarak kurulan modele ilişkin regresyon katsayıları ve analiz sonuçları Çizelge 5.5.'de verilmiştir.

Çizelge 5.5. Lojistik regresyon modeli analiz sonuçları

Değişkenler	Katsayılar	Standart hata	z değeri	Pr (> z)
Sabit	115.8	1112000	0	1
MG	0.0008359	23.03	0	1
Tİİ	1.453	53120	0	1
Rİ	-3.145	101100	0	1
FR	-0.6958	33880	0	1
TTİ	-0.6154	19820	0	1
SİH	-0.3068	16570	0	1
BİH	-0.5722	11740	0	1
Gİ	-0.3829	20040	0	1
Bİ	-0.6138	24800	0	1
BİE	0.4425	22870	0	1
Sİ	1.062	28510	0	1
ES	0.698	28760	0	1
KH	0.2038	14910	0	1
PF	0.2788	11620	0	1
TOL	-0.05762	10260	0	1
İE	0.4615	14010	0	1

Çizelge 5.5. te, katsayıların tek tek anlamlılık sınamasına bakıldığında ise, tüm katsayıların anlamsız olduğu görülmektedir. Lojistik regresyon modeli için, model genel olarak anlamlı bulunmasına rağmen katsayıların tek tek anlamlılığına bakıldığında anlamsız olduğu Çizelge 5.5.'te görülmektedir. Modelin genel olarak anlamlı olması ancak katsayıların anlamsız bulunması bir çoklu iç ilişki göstergesi olduğundan, tüm bu analiz sonuçlarından çoklu iç

ilişki probleminin varlığından söz edilebilir. Çoklu iç ilişki probleminde, tahmincilerin standart hataları olduğundan yüksek çıkmaktadır.

Çizelge 5.6. Modelin doğru sınıflandırma tablosu

Adım	Gözlenen	Tahmin edilen		
		y		Doğru sınıflandırma yüzdesi
		0,00	1,00	
1	0,00	72	0	100,0
	1,00	0	8	100,0
Genel yüzde				100,0

Çizelge 5.6.'da model için kesim değeri (c) 0,5 olarak belirlenebilmekte ve tahmin edilen değerler, kesim değerleri ile karşılaştırılarak uygun kategoriye atama yapılabilmektedir. Tahmin edilen değer eğer 0,5 değerinden büyük ise 1 kategorisine, küçük ise 0 kategorisine atama yapılmaktadır. Böylelikle modelin genel olarak doğru sınıflandırma oranı tespit edilebilmektedir. Ancak kurulan bu model çoklu iç ilişki varlığından dolayı sınıflandırma amaçlı kullanılamamaktadır. Çünkü modelin tüm katsayıları istatistiksel olarak anlamsız bulunmuştur.

İncelenen veri setinde çoklu iç ilişki sorununun olması nedeniyle bu probleminin giderilmesi için öncelikle lojistik ridge regresyon analizi uygulanmıştır.

5.3. Lojistik Ridge Regresyon Model Sonuçları

Lojistik regresyon modelinde, çoklu iç ilişki probleminin giderilmesi için yanlı tahmin edicilere başvurulur. Sözü edilen yanlı tahmin yöntemlerinden bir tanesi lojistik ridge regresyon modeli bu problemi gidermek için kullanılmıştır.

Lojistik ridge regresyon analizi kullanılarak ülkelerin alt endeks değerlerine ait verilerin analiz sonuçları Çizelge 5.7.'de verilmiştir. Ridge parametresinin seçimi için literatürde birçok yaklaşım vardır. Ancak çalışmada k'nın seçimi R paket programı

tarafından otomatik olarak belirlenmiştir. Ridge parametresi 0.1430933 olarak hesaplanmış ve varyans değeri 2.298 olarak bulunmuştur.

Çizelge 5.7. Lojistik ridge regresyon analiz sonuçları

Değişkenler	Katsayılar	Standart hata	z değeri	Pr (> z)
Sabit	6.466	-	-	-
MG	-0.000002662	0.2184	-1.107	0.268254
Tİİ	-0.01537	0.3557	-5.564	2.63e-08 ***
Rİ	-0.01561	0.3301	-4.333	1.47e-05 ***
FR	-0.007696	0.3249	-3.013	0.002587 **
TTİ	-0.01989	0.5609	-4.173	3.00e-05 ***
SİH	-0.007853	0.4693	-2.988	0.002806 **
BİH	-0.01371	0.4284	-5.06	4.19e-07 ***
Gİ	-0.008589	0.4062	-2.805	0.005033 **
Bİ	-0.01992	0.5214	-4.973	6.59e-07 ***
BİE	-0.00711	0.3321	-3.365	0.000765 ***
Sİ	-0.002845	0.4766	-0.352	0.725068
ES	0.0001313	0.5227	0.039	0.96914
KH	-0.001583	0.4383	-0.763	0.445561
PF	-0.01232	0.4144	-2.955	0.003124 **
TOL	-0.004002	0.3809	-1.387	0.165331
İE	-0.009896	0.3704	-3.901	9.58e-05 ***

p < 0.001***, p < 0.01**, p < 0.05*, p < 0.1

Buna göre, Lojistik ridge regresyon denklemi;

$$\hat{y} = 6.466 + (-0.01537)x_2 + (-0.01561)x_3 + (-0.007696)x_4 + (-0.01989)x_5 + (-0.007853)x_6 + (-0.01371)x_7 + (-0.008589)x_8 + (-0.01992)x_9 + (-0.00711)x_{10} + (-0.01232)x_{14} + (-0.009896)x_{16}$$

olarak yazılabilir. Çizelge 5.7. incelendiğinde ülkelerin milli gelirinin(x_1), sağlık ihtiyaçlarının(x_{11}), ekosistem sürdürülebilirliğinin(x_{12}), kişisel özgürlüğün(x_{13}) ve tolerans(x_{15}) değişkenlerinin sosyal gelişmişlik endeksi üzerine modele etkisi anlamsız bulunurken temel insan ihtiyaçları (x_2), refah ihtiyaçları(x_3), fırsat(x_4), temel tıp ihtiyaçları(x_5), su ihtiyaçları(x_6), barınak ihtiyaçları(x_7), güvenlik ihtiyaçları(x_8), bilgi ihtiyaçları(x_9), bilgi

ve iletişime erişim(x_{10}), kişisel özgürlük(x_{14}), ileri eğitim(x_{16}) değişkenlerinin sosyal gelişmişlik endeksi üzerine etkileri anlamlı bulunmuştur.

5.4. Robust Lojistik Regresyon Modeli Sonuçları

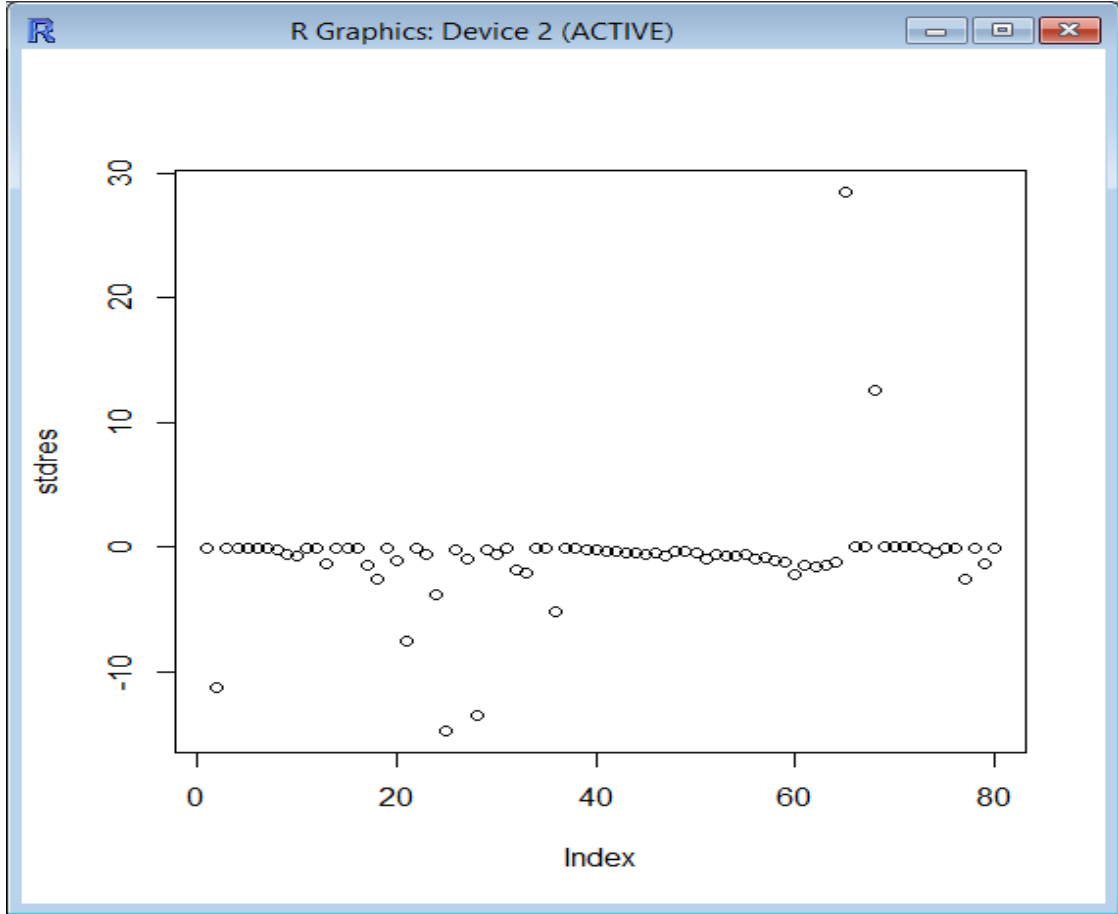
Lojistik regresyon analizinde de çoklu doğrusal regresyonda olduğu gibi aykırı değerler olabilir. Bu durumda model varsayımları sağlanamadığından elde edilen tahmin değerleri gerçeği yansıtmamaktadır. Bu çalışmada literatürde yaygın olarak kullanılan Bianco-Yohai tahmincisi ele alınmış ve uygulama verileri için tahminler elde edilmiştir.

Öncelikle veri seti içindeki aykırı değerler standartlaştırılmış artıklara dayalı olarak elde edilmiştir. Oluşturulan lojistik regresyon modelinde kullanılan gözlemler incelendiğinde aykırı değerlere rastlanmıştır. Hesaplanan artık değerleri $[-2,+2]$ aralığı dışında ise aykırı değer olarak ifade edilmektedir. Bulunan aykırı değerler Çizelge 5.8.'de verilmiştir. Çizelge 5.8.'den görüldüğü gibi eğitim verisi içerisindeki 8 gözlem diğer gözlem değerlerine uzak olduğundan aykırı değer olarak belirlenmiştir. Bu değerlere ait grafik ise Şekil 5.1.'de gösterilmiştir.

Çizelge 5.8. Aykırı değerler

Gözlem no	Aykırı değerler
2	-11.242
21	-7.55768
24	-3.72822
25	-14.6624
28	-13.4364
36	-5.16138
65	28.45183
68	12.54902

Şekil 5.1. de yatay ekseninde gözlemler dikey ekseninde ise standartlaştırılmış artık değerleri yer almaktadır. Şekil 5.1. incelendiğinde standartlaştırılmış artıkların 8 tanesinin, ortalaması 0 olan artık değerlerinde oldukça uzakta yer aldığı görülmektedir.



Şekil 5.1. Aykırı değerler için grafik gösterimi

Bianco ve Yohai (1996) tarafından geliştirilen robust lojistik regresyon analizi kullanılarak ülkelerin alt endeks değerlerine ait verilerin analiz sonuçları Çizelge 5.9.'de verilmiştir.

Çizelge 5.9. Robust lojistik regresyon analiz sonuçları

Değişkenler	Katsayılar	Standart hata	z değeri	Pr (> z)
Sabit	115.8	3.744	30.927	< 2e-16 ***
MG	0.0008359	0.00006495	12.869	< 2e-16 ***
Tİİ	1.453	0.2007	7.243	4.38e-13 ***
Rİ	-3.145	0.3454	-9.107	< 2e-16 ***
FR	-0.6958	0.1025	-6.791	1.12e-11 ***
TTİ	-0.6154	0.08837	-6.964	3.31e-12 ***
SİH	-0.3068	0.06088	-5.039	4.69e-07 ***
BİH	-0.5722	0.0422	-13.557	< 2e-16 ***
Gİ	-0.3829	0.07289	-5.253	1.50e-07 ***
Bİ	-0.6138	0.08044	-7.63	2.35e-14 ***
BİE	0.4425	0.07738	5.719	1.07e-08 ***
Sİ	1.062	0.0989	10.736	< 2e-16 ***
ES	0.698	0.09468	7.373	1.67e-13 ***
KH	0.2038	0.04944	4.122	3.75e-05 ***
PF	0.2788	0.04446	6.271	3.59e-10 ***
TOL	-0.05762	0.03108	-1.854	0.0637 .
İE	0.4615	0.04007	11.517	< 2e-16 ***

p < 0.001***, p < 0.01**, p < 0.05*, p < 0.1

Buna göre, robust lojistik regresyon denklemi;

$$\hat{y} = 115.8 + 0.0008359x_1 + 1.453x_2 + (-3.145)x_3 + (-0.6958)x_4 + (-0.6154)x_5 + (-0.3068)x_6 + (-0.5722)x_7 + (-0.3829)x_8 + (-0.6138)x_9 + 0.4425x_{10} + 1.062x_{11} + 0.698x_{12} + 0.2038x_{13} + 0.2788x_{14} + 0.4615x_{16}$$

şeklindedir.

Robust Lojistik Regresyon modeli katsayılarının anlamlılıkları incelendiğinde modele katılan milli gelirin (x_1), temel insan ihtiyaçları (x_2), refah ihtiyaçları (x_3), fırsat (x_4), temel tıp ihtiyaçları (x_5), su ihtiyaçları (x_6), barınak ihtiyaçları (x_7), güvenlik ihtiyaçları (x_8), bilgi ihtiyaçları (x_9), bilgi ve iletişime erişim (x_{10}), sağlık ihtiyaçlarının (x_{11}), ekosistem

sürdürebilirliğinin(x_{12}), kişisel özgürlüğün(x_{13}), kişisel özgürlük(x_{14}) ve ileri eğitim(x_{16}) değişkenlerinin sosyal gelişmişlik endeksi üzerine etkileri anlamlı, tolerans(x_{15}) değişkeninin sosyal gelişmişlik endeksi üzerine etkisi anlamsız çıkmıştır.

5.5. Yöntemlerin Karşılaştırılması

Bu bölüme kadar ele alınan eğitim verisi için lojistik, lojistik ridge ve robust lojistik regresyon yöntemleri uygulanmıştır. Elde edilen bu modellerden hangisinin kullanılacağına belirlenmesi için modellerin doğru sınıflandırma oranları hesaplanmıştır ve Çizelge 5.10.'da verilmiştir.

Çizelge 5.10. Doğru sınıflandırma oranları

Yöntem	Eğitim verisi	Test verisi
Lojistik Regresyon	% 100	-
Lojistik Ridge Regresyon	% 100	% 94
Robust Lojistik Regresyon	% 100	% 100

132 ülkeden oluşan veri seti için belirlenen 80 ülkenin alt endeks değerleri eğitim verisini, geriye kalan 52 ülkenin alt endeks değerleri test verisini oluşturmaktadır. Her bir model için eğitim verisinden oluşturulan regresyon denklemlerine test verisi değerleri girilerek y değerleri hesaplanmaktadır. Hesaplanan y değerleri logit dönüşümü ile 0-1 değer aralığına dönüştürülmektedir. Lojistik regresyonda belirtilen sınıflandırma kuralla göre doğru sınıflandırma oranları hesaplanmaktadır. Kullanılan yöntemlerde lojistik regresyon modelinin genel olarak model anlamlı fakat modelin katsayılarının anlamsız olduğunu Çizelge 5.6.'da gösterilmiştir. Bu nedenle model test verisi için kullanılamamaktadır. Test verisi için kullanılan lojistik ridge regresyon modelinin doğru sınıflandırma oranının %94 olduğunu ve robust lojistik regresyon modeli için doğru sınıflandırma oranının %100 olduğunu görülmektedir. Diğer bir anlatımla ele alınan veri seti için robust lojistik regresyon modeli, lojistik ve lojistik ridge regresyon modelinden daha başarılı tahminler vermektedir.

6. SONUÇ VE ÖNERİLER

Bu tez çalışmasında, lojistik regresyon analizinde karşılaşılabilecek iki önemli sorun ele alınmıştır. Bunlardan birincisi bağımsız değişkenler arasında çoklu iç ilişki olma durumu, diğeri ise veri setinde model varsayımlarını bozan aykırı değerlerin var olması durumudur. Her iki durumda da klasik lojistik regresyon analizinin performansı düşmektedir. Bu nedenle, ilk olarak bağımsız değişkenler arasında çoklu iç ilişki olması durumunda kullanılabilir yanlı bir regresyon yöntemi olan lojistik ridge regresyon yöntemi açıklanmıştır. Daha sonra veri setinde aykırı değer olması durumunda lojistik regresyon yöntemine alternatif olabilecek robust lojistik regresyon yöntemi incelenmiştir.

Bu tez çalışmasında da bağımlı değişken iki kategoriden oluştuğundan ikili lojistik regresyon modeli kullanılmıştır. Tez çalışmasının ilk bölümünde lojistik regresyon modelinin literatürde diğere yöntemlerle hangi noktalarda benzerlik ve farklılık gösterdiği vurgulanmıştır. İkinci bölümde lojistik regresyon modelinin tarihçesinden ve yapılan çalışmalardan bahsedilmiştir. Üçüncü bölümde lojistik regresyon modeli, tahmin yöntemleri ve değerlendirme kriterleri üzerinde durulmuştur. Lojistik regresyon analizinin doğrusal regresyon analizi arasındaki ilişki üzerine durulmuştur ve tercih edilme nedenleri açıklanmıştır. Dördüncü bölümde lojistik regresyon analizi kullanılırken karşılaşılabilecek iki önemli probleme çözüm önerileri sunulmuştur. Çoklu iç ilişki sorununun çözümü için Lojistik Ridge regresyon, aykırı değer sorununun çözümü için robust lojistik regresyon yöntemi tanıtılmıştır. Lojistik Ridge regresyon modelinde parametre tahmini için EÇO yöntemi kullanılmıştır. Ayrıca veri setinde aykırı değer bulunması durumunda ise robust lojistik regresyon modellerinden olan Bianco-Yohai tahmin edicisi anlatılmıştır. Çalışmanın beşinci bölümünde, anlatılan regresyon yöntemleri sosyal gelişmişlik endeksi verileri kullanılarak performansları bakımından karşılaştırılmıştır.

Çalışmanın uygulama bölümünde 132 ülkeden alınan alt endeks değerleri ile vatandaşların gelişmişlik düzeyini hesaplanmıştır. 132 ülke arasından seçilen 80 ülkenin alt endeks değerleri eğitim verisini, geriye kalan 52 ülkenin alt endeks değerleri test verisini oluşturmaktadır. Oluşturulan lojistik regresyon modelinde ele alınan bağımsız değişkenlerin tamamının sosyal gelişmişlik endeksi üzerine etkileri anlamsız çıkmıştır. Modelin genel olarak uyumuna bakıldığında model anlamlıdır. Eğitim verisi için hesaplanan korelasyon

matrisi incelediğinde çoklu iç ilişki probleminin varlığı belirlenmiştir. Çoklu iç ilişki probleminin üstesinden gelebilmek için lojistik Ridge regresyon modeli kullanılmıştır ve modelin parametre tahmini gerçekleştirilmiştir. Elde edilen modelde bağımsız değişkenlerin çoğunluğunun sosyal gelişmişlik endeksi üzerine etkileri anlamlı çıkmıştır.

Ayrıca uygulamada kullanılan veri seti için aykırı değer tespiti yapılmıştır. Eğitim verisinde 8 tane aykırı değer tespit edilmiştir. Bu aykırı değerlerin etkisini modelde en aza indirmek için robust lojistik regresyon modeli uygulanmıştır. Bianco-Yohai tahmin edicisi kullanılarak oluşturulan modelde parametre tahminleri yapılmış ve tek bir değişken dışında geriye kalan tüm bağımsız değişkenlerin sosyal gelişmişlik endeksine etkileri anlamlı çıkmıştır.

Tüm bu analizler sonucunda incelenen veri seti için, kullanılan yöntemlerden hangisinin daha iyi olduğuna doğru sınıflandırma tablosuyla karar verilmiştir. Çoklu iç ilişki problemi ortaya çıktığından lojistik regresyonla yapılan tahminler gerçeği yansıtmayacağı bilinmektedir. Aykırı değerlerinde veri setinde bulunması alternatif regresyon modellerinin kullanılmasını gerektirmiştir. Sınıflandırma tablosuna bakıldığında lojistik Ridge regresyonun doğru sınıflandırma oranının %94 ve robust lojistik regresyon ise %100 olduğu belirlenmiştir.

Lojistik regresyon analizinde çoklu iç ilişki probleminin varlığı ve aykırı değer/değerlerin bulunması durumunda parametre tahminleri gerçeği yansıtmamaktadır. Bu durumlarda alternatif regresyon modelleri kullanılmalıdır. Bu çalışmada incelenen veri seti için en iyi performansı robust lojistik regresyon analizi vermesine rağmen, bu durum robust lojistik regresyon analizinin çoklu iç ilişki probleminin çözümüne alternatif olduğu anlamına gelmemektedir. Sadece incelenen veri setinde aykırı değerler çoklu iç ilişkinin şiddetini arttıran nitelikte olabilmektedir. Bu çalışmanın sonucunda lojistik regresyon analizi uygulanırken çoklu iç ilişki varlığı durumunda lojistik Ridge regresyon, aykırı değerlerin varlığı durumunda robust lojistik regresyon analizinin kullanılması önerilmektedir.

KAYNAKLAR DİZİNİ

- Akay, K.U., 2014, A graphical evaluation of logistic ridge estimator in mixture experiments, *Journal of Applied Statistics*, 41, 6, 1217-1232.
- Akkaya, S., Ve Pazarlıoğlu, M. V., 1998, *Ekonometri*, Erkan Matbaacılık, İzmir.
- Akın, H.B., Şentürk, E., 2012, Bireylerin Mutluluk Düzeylerinin Ordinal Lojistik Regresyon Analizi İle İncelenmesi, 10, 37, s.83-93.
- Aksaraylı, M. ve Saygın, Ö., 2011, Algılanan hizmet kalitesi ve lojistik regresyon analizi ile hizmet tercihinin etkisinin belirlenmesi, *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 13, 1, 21-37.
- Aktaş, C., 2009, Lojistik Regresyon Analizi: Öğrencilerin Sigara İçme Alışkanlığı Üzerine Bir Uygulama, *Erciyes Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 26, 107-121.
- Albayrak, A.S., 2005, Çoklu Doğrusal Bağlantı Halinde En küçük Kareler Tekniğinin Alternatifi Yanlı Tahmin Teknikleri Ve Bir Uygulama, *ZKÜ Sosyal Bilimler Dergisi*, 1, 1, 105-126.
- Anderson, J.A., 1983, Robust Inference Using Logistic Models, *Bulletion of International Statistical Institute*, 48, 35-53.
- Anderson, J.A., Richardson, S.C., 1979, Logistic discrimination and bias correction in maximum likelihood estimation. *Technometrics*, 21, p.71-78.
- Aranda-Ordaz, F. J., 1981, On Two Families of Transformations to Additivity for Binary Response Data, *Biometrika*, 68, 357–363.
- Arı, A., Önder, H., 2013, Farklı Veri Yapılarında Kullanılabilecek Regresyon Yöntemleri, *Anadolu Tarım Bilimleri Dergisi*, 28, 3, s.168-174.
- Arı, E., 2013, Sıralı lojistik regresyonda paralel doğrular varsayımı ve çözümleme yaklaşımları, *Doktora Tezi*, Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü, Eskişehir, 100 s.
- Arıcan, E., 2010, Nitel yanıt değişkenine sahip regresyon modellerinde tahmin yöntemleri, *Yüksek Lisans tezi*, ÇU, 105 s.

KAYNAKLAR DİZİNİ (devam ediyor)

- Ariffin, S.B., Midi, 2014, The effect of high leverage points on the logistic ridge regression estimator having multicollinearity, AIP Conference Proceeding, 1602, 1105-1111.
- Asar, Y., Karabrahimoğlu, A., Genç, A., 2015, Modified Ridge Regression Parameters: A Comparative Monte Carlo Study, Hacettepe Journal of Mathematics and Statistics, 43, 5, 827-841.
- Atabey, Ö., 2010, Lojistik regresyon modeli ve geriye doğru eliminasyon yöntemiyle değişken seçiminin hipertansiyon riski üzerine uygulamasında Bootstrap yöntemi, Yüksek Lisans tezi, GÜ, 122 s.
- Ayhan, S., 2006, Sıralı lojistik regresyon analiziyle Türkiye'deki hemşirelerin iş bırakma niyetini etkileyen faktörlerin belirlenmesi, Yüksek Lisans Tezi, Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü, Eskişehir, 66 s.
- Bas, M., Çakmak, Z., 2012, Gri İlişkisel Analiz ve Lojistik Regresyon Analizi ile İşletmelerde Finansal Başarısızlığın Belirlenmesi ve Bir Uygulama, Anadolu Üniversitesi Sosyal Bilimler Dergisi, 12, 3, s.63-81.
- Başarır, G., 1990, Çok Değişkenli Verilerde Ayırsama Sorunu ve Lojistik Regresyon Analizi, Doktora tezi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü Ankara, 145 s.
- Bayram, N., 2004, Multinomial Lojistik Regresyon Analizinin İstihdamdaki İşgücüne Uygulanması, İstanbul Üniversitesi İktisat Fakültesi Mecmuası, 54, 2, 61-76.
- Berkson, J., 1944, Application of the Logistic Function to Bio-Assay. Journal of the American Statistical Association, 9, 357-365.
- Bianco, A.M., Yohai V.J., 1996, Robust Statistics, Data Analysis, and Computer Intensive Methods, Springer New York, 109, 17-34.
- Bilgin, M., 2012, Türetilmiş İkili Heterojen Veri Yapılarında Genel, Sağlam Ve Kesin Lojistik Regresyon Yöntemlerinin Karşılaştırılması, Yüksek Lisans, ESOGÜ, 59 s.
- Bircan, H., 2004. Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama, Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 2, 185-208.
- Bondell, H.D., 2008, A characteristic function approach to the biased sampling model, with application to robust logistic regression, Journal of Statistical Planning and Inference, 138, 3, 742-755.

KAYNAKLAR DİZİNİ (devam ediyor)

- Bonney, G. E., 1987, Logistic Regression For Dependent Binary Observations, *Biometrics*, 43, 951-973.
- Buescher, P. A., Larson, L. C., Nelson, M. D., Lenihan, A. J., 1993, Prenatal WIC Participation Can Reduce Low Birth Weight and Newborn Medical Costs: A Cost Benefit Analysis of Wic Participation in North Carolina. *Journal of the American Dietetic Association*, 93, 163-166.
- Bulut, Y.M., 2011, Çoklu İç İlişki Durumunda Kısmi En Küçük Kareler Regresyonu ve Alternatif Yöntemlerle Karşılaştırılması, Yüksek Lisans, ESOGÜ, 79 s.
- Cessie, S.L., Van Houwelingen, J.C., 1992, Ridge Estimators in Logistic Regression, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41, 1, 191-201.
- Cox, D. R., 1970, The Analysis of Binary Data, Methuen, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 21, 2, 113-120.
- Coskun, S., Kartal, M., Coskun, A., Bircan, H., 2004, Lojistik Regresyon Analizinin İncelenmesi ve Diş Hekimliğinde Bir Uygulama, *Cumhuriyet Üniversitesi Diş Hekimliği Fakültesi Dergisi*, 7, 1, 41-50.
- Çokluk, Ö., 2010, Lojistik Regresyon Analizi: Kavram ve Uygulama Kuram ve Uygulamada Eğitim Bilimleri / Educational Sciences: Theory & Practice, 10, 3, 1357-1407.
- Çolak, E., 2002, Koşullu ve Sınırlandırılmış Lojistik Regresyon Yöntemlerinin Karşılaştırılması ve Bir Uygulama, Yüksek Lisans, ESOGÜ, 55 s.
- Duffy, D. E., Santner, T. J., 1989, On the Small Sample Properties of Norm-Restricted Maximum Likelihood Estimators for Logistic Regression Models, *Communs Statist, Theory Meth.*, 18, 959-980.
- Erkoç, A., 2011, Lineer Olmayan Regresyon Modellerinde Çoklu İç İlişki Problemi, Yüksek Lisans, MÜ, 45 s.
- Finney, D. J., 1971. *Probit Analysis*, Cambridge University Press, p.272.

KAYNAKLAR DİZİNİ (devam ediyor)

- Garside, P. S., Glueck, C. J., 1995, The Important Role of Modifiable Dietary and Behaviour Characteristic in the Causation and Prevention of Coronary Heart Disease Hospitalization and Mortality. *Journal of American College of Nutrition*, 14, p.71-79.
- Girginer, N., Cankuş, B., 2008, Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 15, 1, s.181-193.
- Gürler, Ü., 1996, Bivariate estimation with right-truncated data, *Journal of the American Statistical Association*, 91, 1152-1165.
- Hadi A.S., Imon R.A.H.M., 2008, Identification of Multiple Outliers in Logistic Regression, *Communications in Statistics—Theory and Methods*, 37, 1697–1709.
- Halpering, M., Blackwelder, W. L., Verter, J. I., 1971, Estimation of the Multivariate Logistic Risk Function: A Comparison of the Discriminant Function and Maximum Likelihood Approaches. *Journal of the Chronic Disease*, 24, 125-158.
- Heath, A., McMahon, D., Roberts, J., 2000, Ethnic Differences in the Labour Market: A Comparison of the Samples of Anonymized Records and Labour Force Survey. *Journal of the Royal Statistical Society, Series A*, 163, 3, 341-361.
- Hobza, T., Pardo, L., Vajda, I., 2012, Robust median estimator for generalized linear models with binary responses, *Kybernetika*, 48, 4, 768-794.
- Hoerl, A.E., Kennard, R.W., 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12, 1, p.55-67.
- Hughes, R.G., Rogers, A.E., 2004, First, Do No Harm: Are You Tired? *American Journal of Nursing*, 104, 3 p.36-8.
- Johnson, W., 1985, Influence Measures for Logistic Regression. Another Point of View, *Biometrika*, 72, 1, 59–65.
- Karacasu, M., Ergül, B., Altın Yavuz. A., 2013, Estimating the causes of traffic accidents using logistic regression and discriminant analysis, *International Journal of Injury Control and Safety Promotion*, 21, 4, 305-313.
- Karagöz, Y., Kınır, S., Yıldız M.S., 2010, İş Tatminini Etkileyen Faktörlerin Kriz Ortamındaki Etkisinin Lojistik Regresyon Analizi İle Belirlenmesi, *SÜ İİBF Sosyal ve Ekonomik Araştırmalar Dergisi*, 13, s.341-362.

KAYNAKLAR DİZİNİ (devam ediyor)

- Kloiber, L. L., Winn, N. J., Shaffer, S. G., Hassanain, R. S., 1996, Late Hyponatremia in Very Low Birth Weight Infants: Incidence and Associated Risk Factors. *Journal of the American Dietetic Association*, 96, 880-884.
- Klugman, J., 2010, Human Development Report 2010 Team, Research Report, 238 p.
- Konya Valiliği, 2011, Konya Karaman İlçeleri Sosyal Gelişmişlik Endeksi, Yayın no: 2, s. 1-46.
- Krogh, A., 1916, *The Respiratory Exchange of Animals and Man*. Longmans, Green & Co, London
- Lecessie, S., Van Houwelingen J. C., 1992, Ridge Estimators in Logistic Regression. *Applied Statistics*, 41, 1, 191-201.
- Lee, A. H., Silvapulle, M. J., 1988, Ridge Estimation in Logistic Regression. *Communications in Statistics, part B-Simulations and Computations* 4, 1231-1257.
- Mansson, K., Shukur G., 2013, On Ridge Parameters in Logistic Regression, *Communications in Statistics - Theory and Methods*, 40, 18, 3366-3381.
- Menard, S., 2002 *Applied Logistic Regression Analysis, Second Edition*, London: Sage Publications, 12-14.
- Montgomery, D.C., Peck, A.E., 2001, Vining, G.G., *Introduction to the Linear Regression Analysis*, John Wiley & Sons, s.642.
- Muniz, G., Golam Kibria, B.M., 2009, On Some Ridge Regression Estimators: An Empirical Comparisons, *Communications in Statistics - Simulation and Computation*, 38, 3, 621-630.
- Murat, D. and Işığışık, E., 2008, Expectations regarding the economic and political situation in the 2007 election period : the case of Bursa, *Uludağ Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 27, 2, 1-24.
- Mutlu, M., 2013, Açık İşletme Kömür Madenciliğinde Lojistik Regresyon Analizi İle İş Kazalarının Değerlendirilmesi, *Yüksek Lisans, ESOGÜ*, 109 s.

KAYNAKLAR DİZİNİ (devam ediyor)

- Myers, R. H., 1990 Classical and Modern Regression with Applications, Massachusetts: PWS-Kent Publishing Company, p.448.
- Nargeleçekenler, M., 2005, Suç Veri Tabanının Lojistik Regresyon Analizi İle Tahmini: Bursa Örneği, Polis Bilimleri Dergisi, 7, 2, 31-50.
- Neter, J., W. Wasserman, M. Kunter 1990, Applied Linear Statistical Models, CRC Press, p.1184.
- Ogoke, U.P., Nduka E.C., Nja M.E., 2013, A New Logistic Ridge Regression Estimator Using Exponentiated Response Function, Journal of Statistical and Econometric Methods, 2, 4, 161-171.
- Oktay, E., Özer, H, Özçomak, M.S., 2008, Konutlardaki Doğal Gaz Sisteminin Tercih Edilmesinde Etkili Olan Faktörlerin Araştırılması: Erzurum Örneği Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 12, 2, s.341-353.
- O'Neil, T. J., Barry, S. C., 1995, Truncated Logistic Regression, Biometrics, 51, 533-541.
- Oğuz, B., ve Budak, F., 2005, Evsel Katı Atıkların Geri Kazanımı Programlarına Katılımı Etkileyen Faktörlerin Belirlenmesi, 6. Ulusal Çevre Mühendisliği Kongresi, İstanbul.
- Oğuzlar, A., 2005, Lojistik regresyon analizi yardımıyla suçlu profilinin belirlenmesi, Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, 19, 21-37.
- Orhunbilge, N., 2000,Uygulamalı Regresyon ve Korelasyon Analizi, Avcıol Basım Yayın, İstanbul, s.240-251.
- Orsberg, L., 2001, Needs and Wants - What is Social Progress and How Should it be Measured?, 23-41.
- Önder, H., 2001, Lojistik Regresyonlarda Değişken Seçimi, Yüksek Lisans Tezi, Çukurova Üniversitesi, Adana, 63s.
- Özarıcı, Ö., 1996, Farklı not sistemlerinde öğrencinin başarılı olma olasılığının probit regresyon analiziyle değerlendirilmesi, Yüksek Lisans Tezi, Eskişehir Osmangazi Üniversitesi, Eskişehir, 63s.
- Özçomak, M.S., Oktay, E., Özer, H., 2005, EKEV Akademi Dergisi,

KAYNAKLAR DİZİNİ (devam ediyor)

- Özdamar, K., 2004, Paket Programlar ile İstatistiksel Veri Analizi 1, Kaan Kitapevi, Eskişehir
- Özdamar, K., 2004, Paket Programlar ile İstatistiksel Veri Analizi 2, Kaan Kitapevi, Eskişehir.
- Özdemir, H., 2010, Lojistik Regresyon İle Kredi Sorgulama ve Bir Uygulama, Yüksek Lisans, YTÜ, 108 s.
- Peoples, M. D., Siegel, E., Suchindran, C. M., Origasa, H., Ware, A., Barakat, A., 1991, Characteristics of Maternal Employment During Pregnancy: Effects on Low Birthweight, American Journal of Public Health, 81, 007-1012.
- Pregibon, D., 1981, Logistic Regression Diagnostics, The Annals of Statistics, 9, 4, 705-724.
- Pfeiffer, R.M., Gail, M.H., 2001, Inference for Covariates that Accounts for Ascertainment and Random Genetic Effects in Family Studies, Biometrika, 88, p.933-948.
- Porter, M.E., Stern, S., Loria, R.A., 2013, Social Progress Index 2013, 1-53 s.
- Sable, M. R., Herman, A. A., 1997, The Relationship Between Prenatal Health Behaviour Advice and Low Birthweight. Public Health Reports. 112, 332-339.
- Schaefer, R. L., Roi, L. D., Wolfe, R. A., 1984, A Ridge Logistic Estimator, Communications in Statistics Theory and Methods, 13, 99-113.
- Sökmen, A., 2014, Sosyal Gelişme Endeksi Türkiye İçin Ne İfade Ediyor?, Araştırma Raporu, 1-11 s.
- Şahin, M., 1999, Lojistik regresyon ve biyolojik alanlarda kullanımı, Yüksek Lisans Tezi, Kahramanmaraş Sütçü İmam Üniversitesi, Kahramanmaraş, 88 s.
- Şerbetçi, A., Özçomak, M.S., 2013, Sıralı Lojistik Regresyon Analizi ile İstatistik ve Ekonometri Derslerinde Başarıyı Etkileyen Faktörlerin Belirlenmesi Atatürk Üniversitesi İktisadi ve İdari Bilimler Fakültesi Öğrencileri Üzerine Uygulama, Kahramanmaraş Sütçü İmam Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 3, 1, s.110-89.
- Tatlıdil, H., 1992, Uygulamalı Çok Değişkenli İstatistiksel Analiz, Cem Ofset, Ankara, s.292-293.

KAYNAKLAR DİZİNİ (devam ediyor)

- Tester, J.M., Rutherford, G.V., Wald, Z., Rutherford, M.V., 2004, A Matched Case–Control Study Evaluating the Effectiveness of Speed Humps in Reducing Child Pedestrian Injuries, *American Journal of Public Health*, 94, 4, p. 646-650.
- Tüzüntürk, S., 2007, Ekonometri bölümü mezunlarının çalışma hayatına girişi: deneysel bir alan araştırması, 8. Türkiye Ekonometri ve İstatistik Kongresi,
- Urgan, N.N., 2007, Lojistik Regresyonda Bazı Yanlı Kestiricilerin İncelenmesi, Doktora Tezi, DÜ, 70 s.
- Ürük, E., 2007, İstatistiksel Uygulamalarda Lojistik Regresyon Analizi, Yüksek Lisans Tezi, MÜ, 83 s.
- Yavuzkanat, M.S., 2011, Lojistik Regresyonda Çoklu Aykırı Gözlemlerin Belirlenmesi ve Etkililiklerinin İncelenmesi, Yüksek Lisans Tezi, GÜ, 76 s.