T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
EKONOMETRİ PROGRAMI
YÜKSEK LİSANS TEZİ

TÜRKİYE HALKI VERİLERİNE DAYALI SİGARA KULLANIMINI ETKİLEYEN FAKTÖRLERİN BELİRLENMESİNDE ÇOK DEĞİŞKENLİ LOJİSTİK REGRESYON ANALİZ TEKNİKLERİ KULLANILARAK YAPILAN BİR ÇALIŞMA

Başak PASİN

Danışman
Prof. Dr. Levent ŞENYAY

İZMİR-2019

YÜKSEK LİSANS
TEZ ONAY SAYFASI

Üniversite : Dokuz Eylül Üniversitesi

Enstitü : Sosyal Bilimler Enstitüsü

Adı ve Soyadı : BAŞAK PASİN

Öğrenci No : 2013800200

Tez Başlığı : Türkiye Halkı Sağlık Verilerine Dayalı Çok Değişkenli ve Kategorik İstatistik Analizleri Kullanılarak Yapılan Bir Değerlendirme Çalışması

Savunma Tarihi : 17/07/2019

Danışmanı : Prof.Dr.Levent ŞENYAY

JÜRİ ÜYELERİ

Ünvanı, Adı, Soyadı	Üniversitesi	İmza
Prof.Dr.Levent ŞENYAY	- Dokuz Eylül Üniversitesi	.....
Prof.Dr.Cenk ÖZLER	- Dokuz Eylül Üniversitesi	.....
Prof.Dr.Şaban EREN	- Yaşar Üniversitesi	.....

BAŞAK PASİN tarafından hazırlanmış ve sunulmuş olan bu tez savunmada başarılı bulunarak oy birliği(X) / oy çokluğu( ) ile kabul edilmiştir.

Prof. Dr. Metin ARIKAN
Müdür

YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum "Türkiye Halkı Verilerine Dayalı Sigara Kullanımını Etkileyen Faktörlerin Belirlenmesinde Çok Değişkenli Lojistik Regresyon Analiz Teknikleri Kullanılarak Yapılan Bir Çalışma" adlı çalışmanın, tarafımdan, akademik kurallara ve etik değerlere uygun olarak yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

..../..../.....

Başak PASİN

İmza

ÖZET

Yüksek Lisans Tezi

Türkiye Halkı Verilerine Dayalı Sigara Kullanımını Etkileyen Faktörlerin Belirlenmesinde Çok Değişkenli Lojistik Regresyon Analiz Teknikleri Kullanılarak Yapılan Bir Çalışma

Başak PASİN

Dokuz Eylül Üniversitesi
Sosyal Bilimler Enstitüsü
Ekonometri Anabilim Dalı
Ekonometri Programı

Bu çalışmada, basit ve çoklu lojistik regresyon hakkında bilgi verilmiş, lojistik regresyon katsayılarının nasıl hesaplandığı anlatılmış, lojistik regresyon enter yöntemiyle tahminsel model bulunmuştur. Ortaya çıkan modelin verilere uyup uymadığına Hosmer-Lemeshow yöntemiyle bakılmıştır. Sınıflandırma tablosu oluşturulmuştur.

Bağımlı değişkenimiz olan sigara kullanma veya sigara kullanmama kategorik bir değişken olduğu için uygulamada ikili lojistik regresyon analizi kullanılmıştır. Elde edilen veriler SPSS 25'de analiz edilmiştir. Kişilerin sigara kullanımına etkili olan faktörleri ve bu faktörlerin etkilerini belirlemek amacıyla, daha önce yapılan çalışmalarda sigara kullanımı üzerinde etkili olarak belirlenen 11 faktör çalışmamızda bağımsız değişken olarak ele alınmıştır. Elde edilen analiz sonuçları ile mevcut çalışmaların sonuçları incelendiğinde, sigara kullanımı üzerinde etkisi olduğu düşünülen 11 faktörden 6'sı bu çalışmada etkili çıkmıştır. Bu 6 faktör kişilerin sigara kullanması ve kullanmamasında etkili faktörler olarak belirlenmiştir.

Anahtar Kelimeler: Lojistik Regresyon, SPSS, Sigara, ODDS.

ABSTRACT

Master's Thesis

A Study Related with Multivariate Logistic Regression Analysis Techniques on the Determination of Factors Affecting Smoking Using by People of Turkish Data

Başak PASİN

Dokuz Eylül University

Graduate School of Social Sciences

Department of Econometrics

Econometrics Program

In this study the simple and multiple logistic regression model have been defined. To find an estimated model with enter method, it has been explained how to estimate the regression coefficients of the models. Hosmer-Lemeshow test is used to to check whether the simplifier model fits the data or not. Classification table has been created.

Smoking or non-smoking was taken as dependent variable in this study. Since it is actually a categorical variable, we have been applied binary logistic regression analysis. Collected data has been analyzed in SPSS 25. The objective of this study is to identify and using binary logistic regression explore the factors affecting smoking or non-smoking. Accordingly, 11 factors (determinants) which had already been identified by researchers were taken as independent variable in our work. According to collected data and results of analysis, the number of 1 determinants considered to be important has been reduced to 6 in this study. These 6 factors have been determined as important for smoking or non-smoking..

Keywords: Logistic Regression, SPSS, Smoking ODDS.

TÜRKİYE HALKI VERİLERİNE DAYALI SİGARA KULLANIMINI ETKİLEYEN FAKTÖRLERİN BELİRLENMESİNDE ÇOK DEĞİŞKENLİ LOJİSTİK REGRESYON ANALİZ TEKNİKLERİ KULLANILARAK YAPILAN BİR ÇALIŞMA

İÇİNDEKİLER

İKİNCİ BÖLÜM
SİGARA KULLANIMI

ÜÇÜNCÜ BÖLÜM
KATEGORİK VERİ ANALİZİ - LOJİSTİK REGRESYON ANALİZİ

DÖRDÜNCÜ BÖLÜM

UYGULAMA

KISALTMALAR

CAPI Computer Assisted Personal Interviewing

CDC Centers for Disease Control and Preventio

ÇDİAT Çok Değişkenli İstatistiksel Analiz Tekniklerini

DSÖ Dünya Sağlık Örgütü

EKK En Küçük Karaler

KOAH Kronik Obstruktif Akciğer Hastalığı

TÜİK Türkiye İstatistik Kurumu

WHO World Health Organization

TABLOLAR LİSTESİ

ŞEKİLLER LİSTESİ

GİRİŞ

Sosyal bilimler alanında tutumların ölçülmesi gibi çalışmalarda çoğunlukla anket tekniğinden yararlanılmaktadır. Anket tekniği ile elde edilen veriler ise çoğunlukla kategorik yapıda olmakta ve bu nedenle çok değişkenli analiz tekniklerinin gerektirdiği varsayımlar sağlanamamaktadır. Böyle durumlarda varsayım gerektirmeyen daha esnek çok değişkenli analiz teknikleri tercih edilmektedir. Lojistik regresyon yöntemi de bu yöntemlerden biridir. Lojistik regresyon analizi; bağımlı değişkeni iki veya ikiden fazla şıkka sahip olan bir denklemde, bağımsız değişkenler ile bağımlı değişken arasındaki ilişkiyi ifade etmekte kullanılan bir yöntemdir.

Sigara içme durumu da tipik bir kategorik veri örneğidir. Sigara, vücuda zarar veren çok sayıda madde içermekte olup, birçok hastalığa neden olmaktadır. Kanser, kalp ve akciğer hastalıkları gibi birçok ölümcül hastalığa yol açmasına rağmen yıllardan beri yaygın olarak kullanılmaktadır. Sigaranın kısa sürede alışkanlık yapabilmesi, kolayca temin edilebilir olması, sadece sigara içenleri değil, çevrede bulunanların sağlığını da tehdit etmesi gibi nedenler, sigaranın halk sağlığı açısından önemini belirten nedenlerden bazılarıdır. Böylesine zararlı bir alışkanlıktan kişileri kurtarmak ve toplumda yaygınlaşmasını önlemek gerekmektedir. Bu nedenle kişileri sigara içmeye yönelten faktörlerin belirlenmesi önemlidir.

Bu çalışmada kullanılan bağımlı değişken kişilerin sigara kullanıp kullanmaması olup, iki şıklı kategorik bir değişkendir. Bu nedenle, kişilerin sigara kullanımını etkileyen faktörleri belirlemek amacıyla, çalışmada iki şıklı (binary) lojistik regresyon analizinden yararlanılmıştır.

Bu çalışmanın amacı, bireylerin sigara içmelerini etkileyen faktörleri belirlemek ve bu faktörlerin her birinin sigara içme üzerindeki etkisini ortaya koymaktır.

Bu amaçla çalışmanın birinci bölümünde çok değişkenli istatistik analizin tanımı, kullanım alanları ve teknikleri konusunda bilgiler verilmiştir. İkinci bölümde sigaraya ilişkin genel bilgiler verilmiş, Dünya sağlık örgütünün (WHO) sigarayla ilgili rapor sonuçları paylaşılmıştır. Sigara kullanımının dünyada ve Türkiye'deki durumu açıklanmıştır. Sigaranın insan sağlığını nasıl etkilediğinden, zararlarından,

nikotin bağımlılığı tedavisinin gerekliliğinden bahsedilmiştir. Üçüncü bölümde kategorik veri analizi ile lojistik regresyon analizi ele alınmıştır. Bu kapsamda lojistik regresyon analizinin kullanım nedenleri, varsayımları, ikili lojistik regresyon analizi, analiz teknikleri, parametre tahmin yöntemleri, parametrelerin anlamlılık testleri vb. teorik konular ele alınmıştır. Son olarak dördüncü bölümde geniş bir anket verisine dayalı bir uygulamaya yer verilmiştir. Öncelikle değişkenlere ve sigara içme oranlarına ait frekans tabloları ele alınmış daha sonra belirlenmiş olan 11 değişken için her bir değişkene ait tek değişkenli basit lojistik regresyon uygulaması yapılmıştır. Sonrasında tek değişkenli basit lojistik regresyon analizi sonunda önemli bulunan değişkenlerle çok değişkenli lojistik uygulaması yapılarak bireylerin sigara kullanımını etkileyen faktörler belirlenmeye çalışılmıştır. Tahminlenen model için model uyum iyiliği ile sınıflandırma tablolarına yer verilmiştir.

BİRİNCİ BÖLÜM
ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

1.1. ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ ÖNEMİ VE LİRETARÜT TARAMASI

Tek değişkenli istatistiklerin kullanılması bazı durumlarda yetersiz ve eksik kalmaktadır. İstatistiksel araştırmalara konu olan güncel problemler ise genellikle tek değişkenle açıklanamayacak kadar karmaşıktır. Araştırmaya konu olan bir problemin çözümünde kuşkusuz problemi etkileyen ve birbirleriyle karşılıklı etkileşimleri olan birçok faktör vardır ve çözülecek problem bu birçok faktörü dikkate alarak incelenmektedir. Çok değişkenli analiz olayları daha basit bir yapıya getirmeyi incelenen değişkenleri sınıflandırmayı, değişkenler arası bağımlılık yapısını yok etmeyi ve boyut indirgemeyi hedefler. Amaç en az sayıda değişken ile incelenen olayları en iyi model ile ifade edebilmektir

Çok değişkenli analiz teknikleri, her birim için çok sayıda, birbirinden farklı, aynı zamanda birbiriyle ilişkili değişkenlerin ölçüldüğü ve bütün değişkenlerin eş zamanlı olarak dikkate alınıp birlikte incelendiği tekniklerdir.

Bir başka anlatımla, çok değişkenli istatistiksel analiz, incelenen olay ve çevresindeki çok sayıda içsel ve dışsal faktörleri dikkate alarak, problemi doğasındaki yapısına ilişkin bilgilere göre incelemek ve çözümlere ulaşmak için geliştirilmiş yöntemler bütünüdür (Özdamar,1999: 1).

Gatty (1966), Çok Değişkenli İstatistiksel Analiz Tekniklerini (ÇDİAT) "değişken grupları arasındaki karşılıklı ilişkileri ölçme ve açıklama imkanı veren tüm istatistik teknikler" şeklinde tanımlamıştır.

Sheth (1971)'a göre çok değişkenli analiz, örnek üzerinde ikiden fazla değişkeni eşanlı çözümleyen tüm istatistik tekniklere verilen addır.

Kachigan (1991)'a göre çok değişkenli istatistiksel analiz, bir nesne kümesi ile ölçülen iki veya daha dazla değişken özelliklerinin eş zamanlı araştırılması ile ilgilenen istatistiksel analiz yöntemidir.

Afifi ve Clark (1997)'a göre çok değişkenli analiz, çalışma konusu olan her bir kişi veya birim için çok sayıda değişkenin elde edildiği verilerin analizlerini açıklamak amacıyla kullanılan istatistik tekniklere verilen addır.

Timm (2002)'e göre ÇDİAT, gözlemlerin bağımlı ve bağımsız değişken kümeleri arasında ilişki kurmak amacıyla kullanılan analiz teknikleridir.

Harris (2001)' e göre çok değişkenli istatistik, değişken kümelenmesinin bir tahmin edici olarak dahil edildiği durumları ele almak amacıyla geçiştirilmiş, açıklayıcı tekniklerin bir sınıflandırmasıdır.

Hair *et* al.(1998)'a göre çok değişkenli analiz, çoklu değişkenlerin tek bir ilişki veya ilişki kümesi içerisindeki analizidir.

Shaw (2003)'e göre ise çok değişkenli istatistik, "iki veya daha fazla değişkeni aynı anda analiz etmeye yarayan yöntemleri tanımlamakta kullanılan teknikler" dir.

Çok değişkenli analiz, istatistiğin uygulamalarda kullanılan önemli bir koludur. Çok değişkenli analizde birbirleriyle ilişkili çok sayıda değişken söz konusudur. Kullanılan tekniklerle, çok sayıda değişkenin oluşturduğu sistemin yapısı belirlenir ve olabildiğince basit bir forma dönüştürülerek herhangi bir konuda doğru karar için gerekli bilgi çıkartılır.(Tatlıdil, 2002: 1)

Çok değişkenli istatistiksel yöntemler, incelenen değişken veya değişkenleri etkileyen faktörleri ve bu faktörlerin ilişkilerinin yapısını ortaya çıkararak çok boyutlu sistemi orijinal değişkenlerin doğrusal birleşimi olan az sayıda bileşen ile özetlemektedir. Bu sebeple, çok değişkenli istatistiksel yöntemler tek değişkenli yöntemlere göre daha kapsamlı veya daha karmaşık olabilmektedir.

Diğer yandan çok değişkenli analizlerde tek değişkenli analizlere göre parametre ve model tahminleri, hipotez testleri gibi daha çok analiz yapılabilmesi, çok değişkenli analizlerin daha karmaşık bir yapıya sahip olması sonucunu ortaya çıkarmaktadır.(Johnson ve Wichern, 2002: 3)

Çok değişkenli analiz teknikleri de istatistiğin diğer kollarında olduğu gibi, bilimsel çalışmaların sayısal sonuçlarının özetlenmesi, yorumlanması ve karar verilirken kullanılmasını sağlamaktadır. Bilimsel çalışmalarda ele alınan olaylar genellikle pek çok etkenin etkisi altına olmaktadır ve gözleme konu olan nesnelerin özellikleri de birbirleriyle ilişkilidir. Bu nedenle, uygulamalarda çok sayıda

değişkenle karşılaşılmaktadır. Yapılan çalışmaların geçerli ve güvenilir sonuçlar verebilmesi için, inceleme konusu olayları bütün yönleriyle değerlendirmek bir zorunluluk olmakla beraber bu zorunluluk sonucu araştırmacı çok değişkenli veri ve bunların analiziyle karşı karşıya kalır.(Tatlıdil, 2002: 1-3)

Çok değişkenli istatistik birden çok özelliğin analiziyle ilgilendiğinden, uygulamalarda değişik amaçlarla kullanılmaktadır. Bu amaçlar aşağıdaki gibi sıralanabilir.

a) **Basitleştirme ve Boyut İndirgeme:** Çok sayıda değişken tarafından belirlenen bir sistem, daha az sayıda gerçek ya da yapay değişkenle temsil edilip basitlik sağlanmaya çalışılır.

b) **Birimlerin Sınıflandırılması:**. Sınıflandırmanın diğer bir biçimi de gözlenen birimlerin önceden tanımlanmış sınıflardan hangilerine ait olduklarının belirlenmesidir. Bir başka deyişle gözlenen birimlerin değişik sınıflar oluşturup oluşturmadıkları belirlenmeye çalışılır. Bazı durumlarda ise birimler yerine değişkenlerin gruplandırılması söz konusu olmaktadır.

c) **Bağımlılık Yapısının İncelenmesi:** Değişkenlerin kovaryans ve korelâsyonlarından yararlanılarak bağımlılığın kaynaklarının ve sonuçlarının açıklanmasıdır. Örneğin; değişkenlerden bir ya da daha fazlasının bağımlı, diğerlerinin bağımsız olduğu regresyon analizinin amacı, değişkenler arasındaki bağımlılık yapısını ortaya çıkarmaktır.

d) **Hipotez Testleri ve Hipotez Oluşturma:** Çok değişkenli verilerde kullanılan hipotez testlerinden bazıları tek değişkenli teorinin çok değişkenli duruma genelleştirilmiş halidir. Bunlardan başka bazı çok değişkenli analiz yöntemleri ise hipotez testlerinde kullanılmaktadır. Bu yöntemlere, açıklanmaya çalışılan olay hakkında yeni model ve hipotezler ortaya çıkarmak amacıyla başvurulabilmektedir.

e) **Sıralama ve Ölçekleme:** Bazı uygulamaların asıl amacı birimlerin belli ölçülere göre sıralanmasıdır. Ölçekleme ise çok sayıda değişkenden yararlanılarak birimlerin daha az boyutla gösterilmesidir. Böylece, grafik gösterimlerinden de faydalanılarak birimlerin karşılaştırılması, yakınlık ya da uzaklıklarının belirlenmesi kolaylıkla yapılabilmektedir.

Çok değişkenli istatistik teknikleri Üretim, Hizmet, Bankacılık ve Finans, Turizm, Otomotiv, İşletmecilik, Ekonomi, Pazarlama, Eğitim, Tarım, Sosyoloji, Psikoloji, Şehircilik, Kimya, Fizik, Biyoloji, Meteoroloji, Jeoloji, Sağlık ve Doğa Bilimleri gibi hemen hemen her alanda uygulanmaktadır.

Değişik uygulama alanlarına yönelik belli başlı önemli çok değişkenli analiz uygulama örnekleri sıralanacak olursa regresyon modelinin kullanımına ilişkin ilk çalışmalar Berkson (1944) tarafından yapılmış ve model Finney (1972) tarafından biyolojik deneylerde probit analize bir alternatif olarak önerilmiştir. (Freeman, 1987: 238) Daha önce lojistik regresyon analiziyle (Bergh, Baigi, Mansson, Mattsson, & Marklund,2007; Chen, Rosenheck, Greenberg, & Seibyl, 2007) veya probit model ile (Huberman, Iyengar, & Jiang, 2007) emeklilik sistemi üzerine de pek çok çalışma yapılmıştır (Bergh, 2007: 25-33).

Benzer şekilde Amerika'da Robin Fisher ve Jennifer Campbell (2002) tarafından sağlık sigortasıyla ilgili olarak sigortası olmayanların yaş, cinsiyet ve ekonomik durumlarına göre sınıflandırılmasında lojistik regresyon modelinden yararlanılmıştır (Temple, 2004: 13-23).

Finans sektöründe ise Lojit analizi ilk kez A.B.D.'de Ohlson (1980) tarafından kullanılmış olup çalışmada, iflastan üç yıl öncesine kadar gidilerek bir iflas tahmin modeli oluşturulmuştur. Türkiye'de ise finans alanında ilk kez 1991 yılında Ramazan Aktaş tarafından mali başarısızlığı belirleyici model kurulmuştur. Bu çalışmada çoklu regresyon analizi, doğrusal ve kuadratik diskriminant analizi, lojit ve probit modeller kullanılmıştır.

Hamarat (1998) Türkiye'de bulunan illeri sağlık, nüfus, ve bazı ekonomik ve kültürel göstergelerine göre kümeleme analizi yöntemi ile gruplandırmıştır. Sonucunda iller 8 alt kümeye ayrılmış olup illerin gruplara doğru atanıp atanmadıklarının saptanması aşamasında ise Kümeleme Analizi sonuçlarına Ayırma

(Diskriminant) Analizi uygulamış ve illerin başarılı bir şekilde kümelendiği sonucunu elde etmiştir (Hamarat, 1998: 1).

Huang, Soutar, Brown (2001) üretim alanında yeni ürün geliştirme ile ilgili bir çalışma yapmış olup çalışmada faktör analizi ve kümeleme analizi kullanmışlardır (Huang, Soutar ve Brown, 2001: 53-59).

Literatürde son yıllarda doğa bilimlerinde çok değişkenli istatistiksel tekniklerin geliştirilmesinde ve uygulanmasında büyük bir artış görülmektedir. Doğa bilimi açısından bakıldığında topluluk çalışmalarında yer alan karmaşık etkileşimleri açıklamak büyük önem taşımakta olup Williams (1976), Orloci (1978), Whittaker (1978a; 1978b), Gauch (1982), Legendre ve Legendre (1983), Pielou (1984)'un çalışmaları, bu alandaki öncü çalışmalar olmuştur.

Akçagöz, Özkan ve Kızılay (2005) tarım alanında, Antalya ili Merkez, Manavgat ve Serik ilçelerine bağlı köylerde 2003 yılı üretim dönemi için 143 çiftçi ile anket çalışması yapmış olup elde edilen verilere faktör analizi uygulanmıştır (Akçagöz ve diğerleri 2005: 63-71).

Yine faktör analiziyle ilgili olarak sosyoloji alanında Prof.Dr Hüseyin Tatlıdil'in 2002 tarihli "Uygulamalı Çok Değişkenli İstatistiksel Analiz" kitabında 26 Avrupa ülkesi'nin istihdam verileri üzerine yapılmış bir analiz çalışması bulunmaktadır.

Otomotiv sektöründe Cankurt ve Miran (2010), Aydın yöresinde çiftçilerin traktör satın alma eğilimleri ve traktör satın alırken göz önüne aldıkları faktörleri araştırmış olup çalışmada lojistik regresyon analizi uygulanmıştır (Cankurt ve Miran, 2010: 43-51).

Jeoloji alanında Sağlam (2013) Çok Değişkenli İstatistiksel Yöntemler ile Toprak Özelliklerinin Gruplandırılması adlı bir çalışma yapmış olup çalışmada faktör analizi ve kümeleme analizi kullanmıştır (Sağlam, 2013: 7).

Turizm sektöründe Baloğlu ve McClearly (1999) ABD'de bir anket çalışması yapmış olup çalışmada ANOVA ve MANOVA kullanmışlardır. Yine turizm sektöründe Choi ve Chu (2000) Hong Kong'da müşteri memnuniyet çalışması yapmış olup çalışmadan faktör analizi ve çoklu regresyon analizi kullanmışlardır.

Bankacılık sektöründe ise Chen (1999) Tayvan'da "Rekabet Avantajı" adlı bir çalışma yapmış olup çalışmada faktör analizi kullanmıştır.

Son olarak pazarlama sektöründe ise; Withey ve Panitz (1995) ABD'de satın alma davranışları üzerine bir çalışma yapmış olup çalışmada ANOVA ve diskriminant analizi kullanmıştır. Yine pazarlama alanında Walsh, Thurau ve Mitchell (2001) Almanya'da "Pazar Bölümleme" adlı bir çalışma yapmış olup çalışmada faktör ve kümeleme analizi kullanmıştır.

1.2. ÇOK DEĞİŞKENLİ VE KATEGORİK VERİ ANALİZ TEKNİKLERİ

Literatürde yoğun olarak kullanılan çok değişkenli istatistiksel analiz teknikleri şunlardır;

Hotelling T^2 Testi

Çok değişkenli normal değişken varsayımına dayandırılan iki ve daha çok değişkenli, tek örnek ve iki örnek hipotezlerinin test edilmesinde kullanılan çok değişkenli istatistik yöntemidir.

Çok Değişkenli Varyans Analizi (*MANOVA*)

İki ve ya daha fazla değişkenli normal dağılım gösteren veri setlerine göre kurulmuş hipotezlerin test edilmesinde kullanılmaktadır. Grup sayısı ikiden fazla olduğu durumlarda Hotelling t^2 testinin alternatifidir.

Çok Değişkenli Kovaryans Analizi (*MANCOVA*)

İki ve ya daha fazla değişkenli normal dağılım gösteren veri setlerine ilişkin kurulmuş hipotezlerin, ortak değişkenlerin de yer aldığı veri yapısıyla test edilmesinde kullanılır. Verilerin normal dağıldığını kabul eder fakat eş parametrik olmayan teste sahip değildir (Chatfield ve Collins, 1980).

Kümeleme Analizi (*Cluster Analysis*)

Yapıları hakkında kesin bilgilerin bulunmadığı bir veri seti içindeki grupları ve/veya değişkenleri, birbirine benzer ve sayısı belirlenmemiş alt kümelere ayırma yöntemidir. Amaç; gruplandırılmış verileri benzerliklerine göre sınıflandırmaktır. Segmentasyon analizi ve taksonomi analizi olarak da adlandırılan kümeleme analizi ile kendi aralarında heterojen ve kendi içinde birbiri ile homojen kümler oluşturulur. Kümeleme analizi aşamalı ve aşamalı olmayan (küme sayısı belli) olmak üzere 2 şekilde yapılmaktadır.

Ayırma Analizi (*Discriminant Analysis*)

Veri setindeki değişkenlerin, sayısı önceden belirlenmiş iki ve daha fazla sayıda gruplara ayrılmasını sağlayan, değişkenlerin özellikleri itibariyle en uygun gruba ayrılmalarını sağlayacak fonksiyonları türeten bir yöntemdir Ayırma analizinin, grupları birbirinden ayırmayı sağlayan fonksiyonları bulmak ve hesaplanan fonksiyonlar aracılığıyla yeni gözlenen bir birimi sınıflama hatası minimum olacak biçimde g tane gruptan herhangi birine atamak amacıdır.

Temel Bileşenler Analizi *(Principal Component Analysis*)

Birbiriyle ilişkili çok sayıda değişken içeren veri matrislerinden, birbiriyle bağımsız ve daha az sayıda yeni veri yapıları elde etmek amacıyla kullanılan bir yöntemdir. Temel bileşenler analizinde amaç aralarında yüksek korelasyon bulunan verilerden daha az sayıda ve aralarında korelâsyon yapısı bulunmayan değişkenler oluşturmaktır. Değişkenler arasındaki bağımlılık yapısının yok edilmesi ve/veya boyut indirgemesi temel bileşenler analizinin temel hareket noktasıdır.

Faktör Analizi (*Factor Analysis*)

Veri seti içerisindeki birbiriyle ilişkili çok sayıda değişkeni, bir araya getirerek yeni daha az sayıda anlamlı ve birbirinden bağımsız faktörler haline

getirmek amacıyla kullanılan bir yöntemdir. Temel bileşenler analizinde olduğu gibi boyut indirgeme ve bağımlılık yapısının yok edilmesi temel hareket noktasıdır.

Çok Boyutlu Ölçekleme Analizi (*Multi-Dimensional Scaling*)

Çok boyutlu ölçekleme analizi, nesneler arasındaki ilişkilerin bilinmediği, fakat birbirlerine olan uzaklıklarının hesaplanabildiği durumlarda, uzaklıklardan yararlanılarak, nesneler arasındaki ilişkileri ortaya koymaya yardımcı olan bir yöntemdir.

Kanonik Korelasyon Analizi (*Canonic Correlation Analysis*)

Kanonik korelasyon analizi, çok sayıda değişkenden oluşan iki ya da daha çok değişken seti (küme) arasındaki ilişkinin derecesini ortaya koyan çok değişkenli istatistik analiz tekniklerinden biridir.

Çok Değişkenli Regresyon Analizi (*Multivariate Regression Analysis*)

Çok değişkenli regresyon analizi, iki bağımlı ve bir ya da daha fazla bağımsız değişken arasındaki sebep-sonuç ilişkilerinin ortaya çıkarılmasını amaçlayan çok değişkenli analiz yöntemidir

Uyum Analizi (*Correspondence Analysis*)

Uyum analizi kontenjans tablosu durumuna getirilmiş kategorik verilerin sıra ve sütunlarının birlikte değişimlerini, daha az boyutlu bir uzayda grafiksel olarak göstermeyi amaçlayan çok değişkenli analiz yöntemidir. Yöntem, değişkenlerin ve bu değişkenlerin alt sınıflarının birbiri ile değişimlerini koordinat ekseninde görüntüleyerek çözümlenmesini amaçlamaktadır (Özdamar, 1999: 461).

Yapısal Eşitlik Modeli

Yapısal eşitlik modeli çok değişkenli istatistiksel tekniklerin birleşiminden oluşan ve çok kuvvetli bir analiz tekniğidir. Yapısal eşitlik modellemesi, regresyon gibi istatistiksel tekniklere kıyasla, birçok bağımlı ve bağımsız değişkenler arasındaki ilişkilerin modellenmesi ile karmaşık bir araştırma problemini tek bir süreçte, sistematik ve kapsamlı bir şekilde ele almayı sağlamaktadır. Bağımlı değişken ile bağımsız değişkenler arasındaki doğrudan ilişkilerin yanı sıra dolaylı ilişkilerin varlığının söz konusu olduğu çok basamaklı bir modelde, regresyon analiziyle doğrudan etkiler tespit edilebilirken, değişkenlerin dolaylı etkileri göz ardı edilmektedir. Dolayısıyla doğrusal regresyon gibi yöntemlerde bağımlı ve bağımsız değişkenler arasındaki bağlantıların sadece tek bir düzeyde ele alınması, yapısal eşitlik modellemesinde ise, her bir ilişki düzeyinin eş anlı olarak değerlendirilmesi yöntemler arasındaki farklardandır. Yapısal eşitlik modellemesi özellikle psikoloji, pazarlama vb. bilimlerde değişkenler arasındaki ilişkilerin değerlendirilmesinde ve modellerin testinde kullanılmaktadır.

Doğrusal Olasılık Modelleri (*Logit Analysis*)

Birden fazla bağımsız değişken, tek bir bağımlı değişkeni tahmin etmek için kullanılmasına yönelik bir model tipidir. Uygulanan teknik bakımından çok değişkenli regresyon analizine benzer olup, doğrusal olasılık modellerinin çok değişkenli regresyon modellerinden ayıran fark ise, bağımlı değişkenin ayırma analizinde olduğu gibi ölçülemez olmasıdır.

Probit Regresyon Modeli (*Probit Regression Models*)

Bağımlı değişkeni; “evet/hayır” gibi yanıtlardan oluşan ve “0” ya da “1” şeklinde kodlanan kategorik modelleri tahmin etmek için kullanılan yaklaşımlardan biridir. Lojistik regresyon analizine alternatif olarak kullanılmaktadır (Albayrak ve ark, 2005: 301). Probit modeller, iki değer alabilen bağımlı değişken ile birçok

bağımsız değişken arasındaki ilişkiyi tahmin ederek hangi bağımsız değişkenlerin bağımlı değişken üzerinde tahmin edici gücü olduğunu gösterirler.

Lojistik Regresyon Analizi (*Logistic Regression Analysis*)

Regresyon yöntemleri bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi açıklayan, bağımlı değişkenin kategorik olduğu durumlarda kullanılan yöntemdir. Lojistik regresyon analizinde hedef, en az sayıda değişkeni kullanarak en iyi uyuma sahip olacak şekilde, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi gösteren bir model kurmaktır. Bu yöntemde diğer yöntemlerde olduğu gibi bağımlı değişkenin normal dağılması, sürekli olması varsayımları yoktur. Bağımlı değişken kategorik değişkendir ve kategorik değişkenler kesikli dağılım gösterdiğinden bağımlı değişken Bernoulli veya binom dağılır. Dolayısıyla beklenen değer ve varyanslar eşit değildir. Bu da gözlemlerin farklı dağılıma sahip olduğunu belirtir.

İKİNCİ BÖLÜM
SİGARA KULLANIMI

2.1. TÜTÜNÜN TARİHÇESİ

Anavatanı tam olarak bilinmese de tütünün tarihçesinin 4000 yıl öncesine kadar gittiği bilinmektedir. Tütünün Avrupa ile tanışması ise Amerika kıtasını keşfeden Kristof Kolomb tarafından yerlilerin çiğnediği tütünü Avrupa'ya taşımasıyla gerçekleşmiştir. Özellikle Kırım Savaşı, Birinci ve İkinci Dünya Savaşları tütün alışkanlığının yayılmasında büyük rol oynamıştır.

Tütünün Osmanlı'da kullanımı ise 1600 yılları başlarında Genovalı ve Venedikli denizciler tarafından I□stanbul'a getirilmesiyle başlamıştır. Tütünün Osmanlı'ya gelmesiyle beraber özellikle ahşap evlerin yoğun olduğu İstanbul'da yangın olaylarında artışlar olmuştur. IV. Murat döneminde bu yangınlar nedeniyle sigaraya karşı ağır yasaklamalar getirilmiştir fakat devlet büyüklerinin de yaygın olarak kullanmasından dolayı bu yasaklamalar tütün kullanımını önlemede etkili olmamıştır.

Osmanlı devleti tütün üretimine ilk olarak Trakya'da bulunan Yenice, Kavala ve İskeçe kentlerinde başlamıştır. Osmanlılar döneminde tütünün kağıda sarılıp yakılarak içilmesi ve 1880 yılında yapılan sigara sarma makinesinin icadı sigara içme alışkanlığının şekillenmesine yol açmıştır.

Tütün üretimine başlandıktan sonra 1884 yılında Osmanlı Devleti ilk sigara fabrikası olan İstanbul Cibali sigara fabrikasını kurmuştur. Daha sonra Adana ve Samsun gibi illerde de sigara fabrikası kurulmuştur (Özkul ve Sarı, 2008: 1).

Cumhuriyetin ilan edilmesinden sonra tütünün satın alınması, işletilmesi, sigara imal edilip satılması ve tütün maddesi ile alakalı diğer tüm işlemlerin hükümetçe yürütülmesi kararı verilmiştir.

1986 yılından sonra ise sigara üretimi konusunda devlet tekeli kaldırmış ve üretimler için özel firmalara da izin vermiştir.

2.2. NİKOTİN BAĞIMLILIĞI

Bağımlılık kişinin madde kullanımı üzerindeki kontrolünü kaybetmesi olarak tanımlanır. Dünya Sağlık Örgütü (DSÖ) madde bağımlılığını 'kullanılan bir psikoaktif maddeye kişinin daha önceden değer verdiği diğer uğraşlardan ve nesnelerden belirgin olarak daha yüksek bir öncelik tanıma davranışı' şeklinde tanımlar. Zamanla madde kullanımı kişinin kendisine ve çevresine zarar verici bir davranış haline gelir. Sigara kullanmak veya sigara dumanına maruz kalmak zamanla bireylerde fiziksel ve psişik bağımlılık oluşmasına neden olur. Tütünde esas bağımlılık yapan madde nikotindir. Sigara diğer bağımlıklara göre daha çok alışkanlık yapıcı daha az zevk verici bir bağımlılık türü olarak kabul edilmektedir.

2.2.1. Dünyada Sigara Bağımlılığı

Dünyada sigara kullanımı hızla yaygınlaşmış olup yetişkinlerin yaklaşık üçte birinin sigara içtiği ve kadın nüfusunda da sigara içme oranının giderek arttığı bilinmektedir. Dünya Sağlık Örgütü (DSÖ) verilerine göre dünyada 15 yaş üstü 1,2 milyardan fazla insan sigara içmektedir ve Dünya genelinde nüfusun %43'u sigara kullanmaktadır. Ayrıca her yıl dünyada yaklaşık 6 milyon sigara kullanıcısının yaşamını kaybettiği ve bu kayıpların 5 milyondan fazlasının ise doğrudan sigaradan kaynaklandığı görülmektedir (WHO, 2008: 48).

Birinci Dünya Savaşı sonrası sigaranın kullanımı artmış olup, 60 lı yıllarda kullanımı zirve yapmıştır. Daha sonra gelişmiş ülkelerde sigara kullanımında önemli derecede bir düşüş gözlenmiştir.

Dünya Sağlık Örgütü (DSÖ), sigara ile bağlantılı hastalıklar sebebiyle 1950 ile 2000 yılları arasında yaklaşık 60 milyon insanın öldüğünü ve bunun II. Dünya Savaşı sebebiyle meydana gelen ölümlerden daha fazla olduğunu bildirmiştir. Sigaranın ABD'deki ölümlerin ise %29'sinden sorumlu olduğu belirtilmektedir. Geçtiğimiz yüzyılın sonunda yaşları 35-69 yaş arasında olan tüm insanların ölümlerinin %30'unun, 69 yaş üstü insanların ölümlerinin ise %14'ünün sigara içimine bağlı geliştiği tahmin edilmektedir. Yapılan çalışmalarda sigara kullanımının beklenen yaşam süresini bütün yaş gruplarında 16 yıl, 35-69 yaş grubunda ise 22 yıl

kadar kısalttığı belirlenmiştir. Japonya'da yapılan bir çalışmada ise, 45 yaş ve üzerindeki nüfusun tıbbi harcamalarının %4'ünün sigara kaynaklı olduğu belirlenmiştir.

Dünya Sağlık Örgütü'nün (DSÖ) yaptığı çalışmalar doğrultusunda nüfusunun %30'u ile Şili, %27'si ile Macaristan, %26'sı ile Estonya ve Çin listenin başını oluştururken Türkiye ise %23,8'lük sigara kullanım oranıyla dünyada 11. sırada yer almaktadır. Bu verilere göre Türkiye' de erkeklerin %37,3'u nün ve kadınların ise %10,7'sinin sigara kullandığı saptanmıştır (WHO, 2015).

2.2.2. Türkiye'de Sigara Bağımlılığı

Sağlık Bakanlığının 1988 yılında yaptığı bir araştırmada 15 yaşından büyük bireylerde sigara içme oranı %43.6 iken, 1997 yılında bu oranın %50'ye yaklaştığı bildirilmiştir. Sigara içme yaygınlığı 1988'de kadınlarda %24, erkeklerde ise %63 olarak saptanmıştır.

2012 yılında yetişkinlerin tütün ürünlerini kullanımları hakkında bilgi toplamak ve karar alıcılara, araştırmacılara veri kaynağı olmak amacıyla Türkiye İstatistik Kurumu (TÜİK) ve Hastalıkları Önleme ve Kontrol Vakfı (Centers for Disease Control and Prevention, CDC Foundation) "Küresel Yetişkin Tütün Araştırması" adlı bir çalışma gerçekleştirmiştir. Yapılan bu araştırmanın birincisi 2008 yılında yürütülmeye başlanmıştır.

Araştırma sonuçlarına göre; Türkiye genelinde 15 ve üstü yaştaki bireylerin 2008 yılında %31,3'ü her gün veya ara sıra tütün ve tütün mamullerini kullanırken 2012 yılında bu oran yaklaşık olarak %27'ye düşmüştür. Tütün ve tütün mamulü kullananların oranı 2012'de 2008'e göre erkeklerde 6,5 puan, kadınlarda ise 2,1 puan düşmüştür. 25-34 ve 35-44 yaş arasındaki bireyler ise her gün veya ara sıra tütün ürünleri kullandığını belirtmişlerdir. Tütün ve tütün mamulü kullanan 15 ve üstü yaştaki bireylerden tütün ve tütün mamulü kullanımını 12 ay içerisinde bırakmayı planlayanların oranı 2008 yılı için %27,8 iken, 2012 yılı için %35,4 olarak elde edilmiştir. Bu oran kadınlar için 2008 yılında %40,8 iken, 2012 yılında %44,9'a yükselmiştir. Aynı oran erkekler için sırasıyla %40,5 ve %41,8'dir. 2008 yılından bu

yana bırakmayı planlayanların sayısı da 2008 yılında yapılan araştırmaya kıyasla %6,6 oranında artmıştır (TÜİK, 2012).

Türkiye 19 Temmuz 2009 tarihinde başlattığı bir çalışma ile halka açık alanlarda sigara içilmesini yasaklayan bir karar alarak, bu uygulama ile toplum içinde sigara kullanımının azaltılması amaçlanmıştır.

TÜI□K'in 2016 yılı verilerine göre Türkiye'de 15 yaş üzeri nüfusun % 26,5'i her gün tütün ürünü kullanmaktadır. Bu oranın 2014 yılında 27,3 olduğu 2010 yılında ise 25,4 olduğu görülmektedir. Türkiye' de tütün kullanımı oldukça yaygın olmakla beraber bu durum büyük bir halk sağlığı sorunudur. Ülkemizde nikotin bağımlılığına sahip olan kişilerin çok büyük bir kısmı ilk sigara deneyimlerini 20 yaşının altında yaşamaktadır. Yapılan araştırmalardan yola çıkarak Türkiye' de sigara içenlerin sayısı giderek artmakta yaş ortalaması ise giderek düşmektedir (Özcebe, 2008: 9-10).

2014'te ise ülkemizde 17 milyon civarında sigara tiryakisi olduğu bilinmektedir. Kadınlarda sigara içme oranı % 28, erkeklerde ise % 57 olarak saptanmıştır. Yaklaşık olarak yılda 100.000 yani her 6 dakikada 1 kişi sigaradan kaynaklı hastalıklara bağlı olarak yaşamını yitirmektedir.

2.3. SİGARA KULLANIMININ SAĞLIĞA ETKİLERİ

Sigara bağımlılığının fiziksel ve psikolojik bağımlılık olmak üzere 2 yönü vardır. Sigara içinde bulunan nikotin maddesi dolayısıyla fiziksel bağımlılık yapmaktadır. Psikolojik bağımlılık ise kişiye göre değişir. Kendine güvensiz, sorunlardan kaçan kişiler psikolojik bağımlılığa daha eğilimlidir. Sigara kullanımı vücudun bütün sistemlerine zarar vermekle beraber en çok zarar gören organlardan biri akciğerdir.

Sigara kullanan kişi sayısının fazla olduğu ülkelerde görülen akciğer kanserinin %90'nın nedeni sigara temellidir (Alberg ve Samet, 2003: 123). Yaşamı boyunca sigara kullanan kişilerin %15'inde akciğer kanseri görülmektedir (Dubey ve Charles, 2008: 941-46). Akciğer kanseri oluşma riski sigara içme süresi ve günlük içilen sigara miktarına göre değişmektedir (Özlu□, 2010: 30). Fakat hayatı boyunca

hiç sigara kullanmamış kişilerin %10'unda da akciğer kanseri görülebilmektedir (Dubey ve Charles, 2008: 941-46).

Sigara kullanımı bazı etkiler ile kronik obstruktif akciğer hastalığı (KOAH) gelişimine neden olur. Dünya genelinde KOAH hastalığı olan kişilerin yaklaşık %95'nin sigara kullananlar ve 20 yıldan fazla bir süre boyunca her gün sigara içmiş oldukları tahmin edilmektedir. (Özyardımcı, 2002: 30-440).

Ayrıca hamilelik döneminde olan bir annenin sigara içmesi sonucunda düşük doğum tartılı bebek dünyaya getirme olasılığı artar. Aynı zamanda erişkinlik döneminde sigara kullanan bireylerin, akciğer gelişimi yavaşlar ve bu da KOAH gelişimine neden olur (Aytemur ve diğerleri, 2010: 44-513).

Sigara içen kişilerde meydana gelen tüm bu ölümcül risklerin yanı sıra sigara içen insanların sigara içmeyen insanlara göre vücut direnci daha zayıftır. Sigara içen bireylerin, bırakan veya hiç içmemiş bireylere göre akut ve kronik hastalığa daha çok yakalandıkları görülmekte ve gündelik hayat aktivitelerinden daha fazla uzak kaldıkları, daha fazla yatarak gün geçirdikleri ve daha fazla okul veya iş devamsızlığı yaptıkları bilinmektedir. Sigara kullanımına erken yaşta başlayıp uzun süre sigara içmeye devam eden kişilerin yarısının ölüm nedeni sigara kullanımından kaynaklanmaktadır ve bu kişilerin orta yaşlarda hayatlarını kaybettikleri görülmektedir. Sigara kullanmayan kişilerin yaşam süreleri sigara kullanan kişilere göre 20-25 yıl daha uzundur (Karlıkaya ve diğerleri, 2006: 51-64).

Sigara doğrudan ölüme sebep olmasa bile arteroskleroz, KOAH (kronik obstrüktif akciğer hastalığı) ve akciğer kanseri gibi birçok kronik hastalığın tetikleyicisidir. Ayrıca sigara kullanımı kalp-damar ve beyin-damar hastalıkları yaşanmasının ise başlıca nedenlerinden biridir. Sigaranın sebep olduğu hastalıkların birçoğu ise ölümle sonuçlanabilmektedir. (Karlıkaya ve diğerleri, 2006: 51-64).

ÜÇÜNCÜ BÖLÜM
KATEGORİK VERİ ANALİZİ – LOJİSTİK VERİ ANALİZİ

3.1. KATEGORİK VERİ ANALİZİ

Değişkenler nitel (kalitatif) ve nicel (kantitatif) olmak üzere iki gruba ayrılmaktadır. Sosyal bilimlerde daha sıklıkla kullanılan nitel değişkenler, sözcüklerle ifade edilirken, nicel değişkenler ise sayılarla ifade edilirler. Kişilerin adları, cinsiyetleri ve oturdukları şehirler gibi değişkenler nitel değişkenlere; boy uzunlukları ve kiloları ise nicel değişkenlere örnek olarak gösterilebilir.

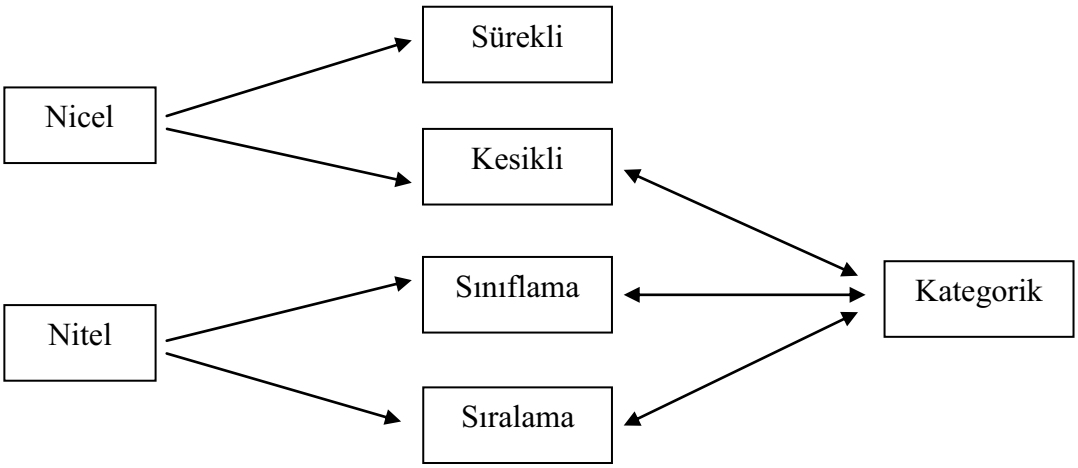
Nicel değişkenler de kendi aralarında sürekli ve kesikli olmak üzere ikiye ayrılırlar. Sürekli değişkenlerde, iki değişken değeri arasına sonsuz çoklukta değer alabilirken, kesikli değişkenlerde değişken sadece belirli değerler alabilmektedir. Ailedeki çocuk sayısı kesikli bir değişken iken, ailenin toplam geliri sürekli bir değişkendir.

Nitel değişkenler ise gözlemlerin sınıflara (kategorilere) ayrılarak açıklandığı veri çeşididir. Kategoriler isimseldir. Örneğin cinsiyet (kadın/erkek), sağlık durumu (iyi-orta-kötü) gibi.

Nitel değişkeler kendi aralarında sırasız nitel ve sıralı nitel olmak üzere iki gruba ayrılırlar. Sırasız nitel veri tiplerinde verilerin sınıflandırılmasında bir sıralama mümkün değildir. Örneğin cinsiyet (kadın – erkek) gibi kategorilerde bir sıralama mümkün değildir. Sıralı nitel veri tipinde ise; kategoriler sıralar halinde karşımıza çıkar örneğin gelir durumu (düşük – orta – yüksek), seklinde olabilir. Görüldüğü gibi kategoriler sıralanmış bir halde karşımıza çıkmıştır.

Bu aşamada kategorik değişkenler, nitel değişkenler ve kesikli nicel değişkenleri kapsamaktadır. Bu ise Şekil 1.'deki gibi özetlenebilir.

Şekil 1: Kategorik Veri Tanımı



3.1.1. Ölçek Türleri

a) **Sınıflayıcı (Nominal) Ölçek:** Belirlenmiş olan numaralar veya simgeler yoluyla yalnızca incelenmekte olan nesnelerin tanımlanmasını sağlayan, en basit ölçek türüdür. Burada verilen numara ve simgeler sadece bir isim işlevi görür. Örneğin; cinsiyet için 1:Kadın 2:Erkek şeklinde bir sınıflandırma yapılabilir.

b) **Sıralayıcı (Ordinal) Ölçek:** Sınıflayıcı ölçekte olduğu gibi değerlendirilecek olan nesnelerin benzer olanları yine aynı numara ya da simgelerle ifade edilebilir fakat değerlendirilen birimleri gruplandırmanın yanında birbirlerine göre küçüklük büyüklük ilişkisi kurarak sıralama yapma özelliği vardır. Örneğin memnuniyet düzeyini ölçen likert ölçek kullanılmış bir ankette; 1-Hiç memnun değil 2-Memnun değil 3-Kısmen memnun 4-Memnun 5-Çok memnun biçiminde yapılan bir sıralama memnuniyet düzeyini ifade eder.

c) **Aralıklı (Interval) Ölçek:** Nesneler arasındaki uzaklıklar ölçülebilir ve yorumlanabilir. Bu ölçek sayılar arasındaki eşit aralıkları kabulü üzerinde kurulmuştur. Örneğin, 5 ile 10 arasındaki aralık, 45 ile 50 arasındaki aralığa eşittir.

d) Oransal (Ratio) Ölçek: Oranlı ölçekte ise sıfır değeri yokluk ifade etmektedir. Ayrıca ölçümler arasında oransallık vardır. Örneğin nesnelerin birine 1, birine 3, birine 6 değerleri verilmişse, 6 ile ifade edilenin 1 ile ifade edilenden altı kat, 3 ile ifade edilenden iki kat daha büyük olduğu söylenebilir.

3.1.2. Kategorik Verilerin Analizinde Kullanılan Olasılık Dağılımları

a) Binom Dağılımı

"Başarılı" ve "başarısız" şeklinde tanımlanmış iki mümkün sonucu olan iadeli örnekleme metodu ile uygulanan bir deneyde her denemenin "başarılı" olma olasılığının p olduğu durumda, n deneme sonucunda başarı sayısı olan x'in dağılımına Binom dağılımı denir. Dağılımın olasılık fonksiyonu aşağıdaki gibidir:

$$p(x) = \frac{n!}{x!(n-x)!}\ p^x\ (1-p)^{n-x} \quad x=0,1,2,\ldots\ldots,n$$

Burada her deneme birbirinden bağımsızdır yani bir denemenin sonucu diğer bir denemeyi etkilemez. Bu denemeler "Bernoulli Denemeleri" olarak bilinir. Dağılımın beklenen değeri ve varyansı aşağıdaki gibi hesaplanmaktadır:

$E(x)= np$

$Var(x)=np(1-p)$

b) Poisson Dağılımı

Bir olayın belirli bir zaman(t) periyodunda, *x* kez meydana gelme olasılığının hesaplanmasını sağlayan bir olasılık dağılımıdır. Bu durumda olasılık fonksiyonu aşağıdaki gibidir:

$$p(\mathrm{x}) = \frac{e^{-\lambda}\,\lambda^{x}}{x!} \qquad \mathrm{x}=0,1,2,\ldots..,\mathrm{n} \quad (\mathrm{e}=2{,}718)$$

Burada λ , bir olayın t zaman periyodunda ortalama meydana gelme sayısıdır. Poisson dağılımına uygun bir değişkenin beklenen değeri $E(x)$ ve varyansı $V(x)$ λ 'ya eşittir.

c) Multinominal Dağılım (Çok terimli dağılım)

Multinominal dağılım ikiden fazla mümkün sonuçlu deneyler için uygulanan kesikli bir dağılımdır. Binom dağılımının genelleştirilmiş şeklidir, dolayısı ile binom dağılımı varsayımları burada da geçerliliğini korumaktadır.

n , i. mümkün sonucun kaç kez meydana geldiğini göstersin (i = 1,2,..., k). pi her denemede i. olayın meydana gelme olasılığıdır ve $\sum_{i=1}^{k} pi=1$'dir. Bu durumda $(n_1,n_2,....,n_k)$ $(n,p_1,.....,p_k)$ parametreli bir Multinominal dağılım gösterir.

$$(n_1,n_2,....,n_k) \sim Mult(n,p_1,.....,p_k)$$

Dağılımın olasılık fonksiyonu $xi \geq 0$, $x_1+.....+x_k=n$ koşulları altında

$$p(n_1 = x_1,.....,n_k = x_k) = \frac{n!}{x1!.....xk!}\; p_1^{x1}.....p_k^{xk}$$ fonksiyonu ile gösterilir.

Genel olarak her bir mümkün sonuç Binom dağılımı gösterir.

$$n_i \sim Bin\,(n,\, p_i)$$

Dolayısıyla ni 'lerin beklenen değerleri ve varyansları Binom dağılımındaki gibi hesaplanır.

$$E(n_i) = np_i$$

$$Var(n_i) = np_i\,(1-p_i)$$

3.1.3. Kotenjans Tabloları

Çapraz tablo ve ya olumsallık tablosu olarak da bilinen kontenjans tabloları genellikle kategorik değişkenler arasındaki ilişkilerin ortaya çıkarılmasında kullanılan ve doğal sayıların matris formunda düzenlendiği tablolardır. Bu tablolarda yer alan değerler değişken kategorilerinin veri kümesinde kaç kez tekrarlandığını gösteren frekanslar (tekrarlanan değerler) yani sayım verileridir. Kontenjans tabloları içerdikleri değişken adedine göre adlandırılırlar. Tablo tek değişkene ilişkin frekansları içeriyorsa tek-yönlü, iki değişkene ilişkin frekansları içeriyorsa iki-yönlü, üç değişkene ilişkin frekansları içeriyorsa üç-yönlü, daha fazla değişkene ilişkin frekansları gösteriyorsa çok yönlü kontenjans tabloları olarak adlandırılmaktadır.

Genel olarak tablolardaki satır değişkeni R (Row), sütun değişkeni C (Column) ile gösterilir. Kontenjans tablolarındaki frekanslar (f_{ij}) ile gösterilir.

i : satır değişkeninin kategorisi

j : sütun değişkeninin kategorisi

R değişkeninin kategorileri R1, R2,…..,RM; C değişkeninin kategorileri C, C2, ……CN olsun.

f_{ij}: (Ri ve Cj) kategorilerinde ki birim sayısının gösterir

Tablo 1: İki-Yönlü (M×N) Boyutlu Kontenjans Tablosunun Genel Gösterimi

Değişkenler	**C Değişkeni**				
R Değişkeni	C1	C2	.	.	CN
R1	f_{11}	f_{12}	.	.	f_{1n}
R2	f_{21}	f_{22}	.	.	f_{2n}
.	.	.	.	.	.
	.	.	.	.	.
RM	f_{m1}	f_{m2}			f_{mn}

İkiden çok kategorik değişkenden oluşan çok-yönlü bir kontenjans tablosunda değişkenlerden birincisi (X_1) n_1, ikincisi (X_2) n_2, n.si (X_n) n_m kategorili ise tablonun

boyutu ($n_1 \times n_2 \times \ldots \times n_m$) olacaktır. Kontenjans tablolarındaki frekanslar (f_{ijz}) ile gösterilir.

i :X_1 değişkeninin kategorisini,

j :X_2 değişkeninin kategorisini

z :X_m değişkeninin kategorisini belirtir.

3.1.4. Göreli (Relatif) Risk, Odds Değeri ve Odds Oranı

a) Göreli (Relatif) Risk

Sadece iki kategoriye sahip olan değişkenler arasındaki oranların eşitliği istatistiksel olarak test edilebilir. En basit iki yönlü kontenjans tablosuna örnek olacak şekilde, Tablo 2 ele alınırsa,

Tablo 2: İki Yönlü Kontenjans Tablosu

	Evet	**Hayır**	**Toplam**
Kadın	f_{11}	f_{12}	f_{1+}
Erkek	f_{21}	f_{22}	f_{2+}
Toplam	f_{+1}	f_{+2}	f_{++}

Tablo 2de bir soruya verilen cevapların cinsiyetlere göre dağılımları yer almaktadır. Soruya verilen "Evet" cevabının i. satırdaki oranı p_i (başarı), "Hayır" cevabının oranı da 1- p_i (başarısızlık) olarak nitelendirip, cinsiyete göre başarı ve başarısızlık olasılıkları test edilebilir.

Örneklemden bulunan her bir satıra ait başarı olasılıkları,

$$p_1 = \frac{f11}{f1+}$$

$$p_2 = \frac{f21}{f2+}$$

Bu durumda cinsiyete göre başarı olasılıklarının farkı test edilebilir. p_1-p_2 oranlar farkı -1 ile +1 değerleri arasında kalacaktır. Eğer başarı olasılıkları arasında fark yok ise yani $p_1=p_2$ ise satır ve sütün arasında bağımsızlık olduğunu yani örneğimiz için, soruya verilen "Evet" cevabının cinsiyete göre değişiklik göstermediği söylenebilir.

p_1-p_2' nın tahminlenmiş standart hatası:

$$\widehat{\sigma}(p_1\text{-}p_2)=\sqrt{\frac{p1(1-p1)}{f1+}+\frac{p2(1-p2)}{f2+}}$$

Örneklem büyüklükleri arttıkça standart hata azalır ve böylece $\pi_1-\pi_2$ ' ın tahmini düzelir.

π_1-π_2 için büyük örneklem %100(1-α) güven aralığı:

$$(p_1\text{-}p_2) \pm Z_{\alpha/2}\widehat{\sigma}(p_1\text{-}p_2)$$

Başarıların oranlarının farkı, $\pi_1-\pi_2$ iki sıranın ana karşılaştırmasıdır. Başarısızlıkların karşılaştırılması da aynı şekildedir.

Başarı olasılıkları test edilirken, oranların 0'a ya da 1'e yakın olması durumunda oranların farkı yoluyla yapılan hipotez testi, yanıltıcı sonuçlar verebilir. İlaç kullandığında yan etkiye maruz kalan hastaların oranına bağlı olarak iki ilacın karşılaştırılmasını dikkate aldığımızda 0.010 ve 0.001 arasındaki fark, 0.410 ve 0.401 arasındaki farka eşittir. Fakat 0.010 ve 0.001 arasındaki fark, 0.0410 ve 0.401 arasındaki farktan daha çok dikkate değer olabilir Bu şekildeki durumlarda oranların oranı aynı zamanda açıklayıcı niteliktedir (Agresti, 1996: 21).

Bu nedenle, oranların farkı (π_1-π_2) yerine oranların oranı $\left(\frac{\pi1}{\pi2}\right)$ daha iyi bir gösterge olacaktır. Bu değere "göreli (relatif) risk" adı verilir (Altaş ve Z. Yıldırım, 2003 : 214).

Göreli risk Riske maruz kalanlarda sonucun ortaya çıkma olasılığının maruz kalmayanlarla karşılaştırmalı olasılığıdır. Örneğin, "sigara içenlerde KOAH olma olasılığı sigara içmeyenlere göre 15 kat daha fazladır" şeklinde tanımladığımızda

sonucun ortaya çıkma olasılığının sigara içmeyenlere göre kaç kat arttığı ifade edilmektedir.

b) Odds Değeri

Bir olayın ortaya çıkma olasılığının (π), ortaya çıkmama olasılığına (1-π) oranına "*Odds"* denir. Odds değeri negatif olmayan değerler alır ve olayın ortaya çıkma ve çıkmama olasılıkları eşit ise odds değeri 1'e eşit olur.

$$\text{Odds} = \frac{\pi}{1-\pi}$$

İlgilenilen olayın, ortaya çıkma olasılığı, ortaya çıkmama olasılığından büyük ise, odds değeri 1'den büyük değer alır. Yani, odds değerinin 1'den büyük olması için, olayın ortaya çıkma ihtimalinin 0.5'ten büyük olması gerekir. π olayın ortaya çıkma ihtimali 0 değerine yaklaşıyorsa olaya ait odd değeri 0'a, 1 değerine yaklaşıyorsa $+\infty$ 'a yaklaşan değer alır.

Odds değeri, değişkenler arası bağımlılık yapısını her değişken kümesi için açıklayan bir kavramdır. Örneğin, 30-39 yaş grubu arasında sigara içen bir kişinin KOAH olma olasılığı 0.6 ise KOAH hastalığına ait odds değeri, $\frac{0.6}{1-0.6}$ = 1.5'tir. Bunun anlamı, 30-39 yaş grubu arasında KOAH olma olasılığı sigara içmek ile 1.5 kat artar ya da 30-39 yaş grubu arasında sigara içenlerin KOAH olma olasılığı, sigara içmeyenlere göre %50 daha fazladır.

c) Odds Oranı

İki odds değerinin birbirine oranı, "odds oranı"nı verir.

$$\theta = \frac{\pi 1/(1-\pi 1)}{\pi 2/(1-\pi 2)}$$

Tablo 3: İki Yönlü Kontenjans Tablosu için Olasılık Değerleri

	Evet	**Hayır**
Kadın	p_{11}	p_{12}
Erkek	p_{21}	p_{22}

Kadınlara ve erkeklere ait Evet odds'unu bulunacak olursa;

$$\text{odds}_E = \frac{p11}{p12}$$

$$\text{odds}_K = \frac{p21}{p22}$$

İki odds değerinin oranından;

$$\theta = \frac{\text{oddsE}}{\text{oddsK}} = \frac{p11/p12}{p21/p22}$$

odds oranı elde edilir. Odds oranı, çapraz çarpım oranı olarak da adlandırılır.

Odds oranı (θ), negatif olmayan bir tamsayıya eşittir. Değişkenler birbirinden bağımsız ise ($\pi_1=\pi_2$) θ =1'dir.Bu değişkenlerin birbirinden bağımsız olduğunun karşılığıdır. Karşılaştırma için bir ölçü olarak kullanılır. Odds oranı hem artış hem de azalış yönünde 1'den uzaklaştıkça değişkenler arası ilişki kuvvetlenir. $1<\theta<\infty$ ise paydaki değişken için başarı odds'u, paydadaki değişkenin başarı odds'undan yüksektir. $0<\theta<1$ durumunda ise, paydaki değişkene ait odds değeri, paydadaki değişkene ait odds değerinden düşüktür.

R ×C kotenjans tablosunda satır değişkeni için,

$$\binom{R}{2} = \frac{R\,(R-1)}{2}$$

sayıda farklı kategori çifti, sütün değişkeni için ise;

$$\binom{C}{2}=\frac{C\,(C-1)}{2}$$

sayıda farklı kategori çifti elde edileceğinden, R×C kontenjans tablosu için elde edilebilecek farklı odds oranı sayısı,

$$\binom{R}{2}.\binom{C}{2}$$

olur.

θ odds oranlarının örnekleme dağılımı, sağa çarpık bir asimetri verir. Bu durumdan kaynaklanan yorum güçlükleri, odds oranları yerine doğal logaritmaları (lnθ) yorumlanarak ortadan kaldırılır. θ =1 durumunda, lnθ =0, 0 <θ <1 durumunda lnθ < 0, 1< θ < ∞ durumunda ise lnθ > 0 olacaktır (Altaş ve Z. Yıldırım, 2003 : 215).

d) Odds Oranı ve Göreli Risk Arasındaki İlişki

Göreli risk $\frac{\pi 1}{\pi 2}$ ve odds oranı θ = $\frac{\pi 1/(1-\pi 1)}{\pi 2/(1-\pi 2)}$ şeklinde tanımlanır. Odds oranı ve göreli risk için verilen formüllerden aşağıdaki sonuç elde edilir (Agresti, 1996):

Odds oranı = Göreli Risk x $\frac{(1-\pi 2)}{(1-\pi 1)}$

Odds oranı = Göreli Risk x $\frac{(1-p2)}{(1-p1)}$

Örneğin odds oranının 1.5 olması *p*'in p_2'nin 1.5 katı olduğu anlamına gelmemektedir. Bu ancak bir göreli risk yorumudur. Çünkü odds ölçümü olasılıklardan daha çok oranlarla ilgilidir. θ = 1.5 odds oranı değeri; olasılık (odds) değeri $p_1/(1\text{-}p_1)$'in, olasılık (odds) değeri $p_2/(1\text{-}p_2)$'nin 1.5 katı olduğunu göstermektedir. İlgilenilen sonucun olasılığı πi (başarı oranları) her iki grup içinde sıfıra yakın olduğunda odds oranı ve göreli risk büyüklüğü birbirine benzer. Birbirine

benzer olduğu durumlarda odds oranı 1.5; p_1'in p_2'nin 1.5 katı olduğu anlamına gelebilmektedir.

Odds oranı ile göreli risk arasındaki bu ilişki oldukça yararlıdır. Çünkü bazı veri setleri için göreli riski hesaplamak mümkün olmayabilir. Böyle durumlarda odds oranı hesaplanarak göreli riskin yaklaşık değeri olarak kullanılır. Her bir π_i değeri küçük olduğunda sağlanan bu benzerlikten dolayı, olgu-kontrol incelemelerinde olduğu gibi göreli risk dolaysız olarak tahmin edilemediğinde odds oranı göreli riskin yaklaşık bir ölçümünü elde edilmesine de olanak sağlar.

3.2. LOJİSTİK REGRESYON ANALİZİ

Regresyon yöntemleri bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi açıklayan yöntemlerdir. Nicel veriler kullanılarak yapılan analizlerde parametrik testler kullanılırken nitel verilerde ise parametrik olmayan testler kullanılmaktadır. Nitel veri analizlerinde kullanabilecek yöntemlerden bir tanesi de lojistik regresyon analizidir.

Lojistik regresyon analizi çok değişkenli regresyon analizine oldukça benzemekle beraber aralarında önemli farklılıklar vardır. Çok değişkenli regresyon analizinde bağımlı değişkenin normal dağılım gösterdiği, bağımsız değişkenlerin arasında çoklu doğrusal bağımlılık olmadığı, hata terimlerinin sıfır ortalamalı ve varyansının normal dağıldığı ve gözlemler arasında otokorelasyon bulunmadığı varsayılmaktadır. Bu varsayımların ihlali durumunda çok değişkenli regresyon analizi uygulanamaz (Johnson, 1988:207). Lojistik regresyon analizinin diğer yöntemlerde olduğu gibi bağımlı değişkenin normal dağılması, sürekli olması varsayımları gerektirmeyen yapısı sayesinde en çok tercih edilen analiz yöntemlerinden birisidir.

Lojistik regresyon analizi bağımlı değişkenin kategorik olarak, ikili, üçlü veya çoklu kategorilerde gözlendiği durumlarda bağımsız değişkenlerle sebep-sonuç ilişkisini belirlemede yararlanılan bir yöntemdir. Lojistik regresyon analizi uygulanırken bağımlı değişkenin niteliğine dikkat edilmektedir. Bağımlı değişkenin iki kategorili olması durumunda iki kategorili (binary), ikiden daha fazla sırasız kategorili olması durumunda çok kategorili (multinominal), ikiden daha fazla sıralı

kategorili olması durumunda sıralı (ordinal) lojistik regresyon modeli kullanılmaktadır. Lojistik regresyonun kullanım amacı en az değişkenle en iyi model uyumuna sahip olacak şekilde, bağımsız değişkenlerin bağımlı değişkenleri tatmin edilir düzeyde açıklayabildiği, istatistiksel olarak anlamlı en uygun modeli kurmaktır. Amaçlarından birisi sınıflandırma, diğeri ise bağımlı ve bağımsız değişkenler arasındaki ilişkileri araştırmak olan lojistik regresyon analizinde, bağımlı değişken kategorik kesikli değerler almaktadır.

3.2.1. Lojistik Regresyon Analizin Varsayımları

Lojistik regresyon bağımsız değişkenlerin dağılımına ilişkin herhangi bir varsayım sağlanmasını gerektirmemekte olup, yine de analizin sağlıklı sonuçlar vermesi için bazı noktalara dikkat etmek gerekmektedir.

Lojistik regresyon analizinin 6 varsayımı vardır. Bu varsayımlar şunlardır.(Baydemir, 2014: 35-36.)

i) Gelecek tahmini yapabilmek için değişkenleri matematiksel kalıba oturtmak en çok karşılaşılan sorunlardan birisidir. Bilgi karmaşasına yol açacak kadar gereksiz değişken modeli açıklamayı güçleştirebilir ve zaman kaybına yol açabilir. Bu yüzden modelin matematiksel kalıbı doğru belirlenmesi gerekmektedir.

ii) Bağımsız değişkenler arasında çoklu doğrusal bağlantı yoktur. Bağımsız değişkenler arasında ilişki olması tahmin sonucunu ciddi biçimde etkiler. Eğer bağımsız değişkenler arasında ilişki varsa lojistik regresyon denklemine ait katsayıların standart hataları büyür. Çoklu doğrusal bağlantı, katsayıların tahmininde yanlılığa sebep olmasa da katsayıların standart hataları arttığı için güvenilirliği etkiler.

iii) Bağımsız değişkenlerle bağımlı değişkenin olasılık değeri arasındaki ilişki doğrusal değildir. Lojistik regresyon analizi doğrusallık varsayımını gerektirmez. Fakat lojistik regresyon denkleminde logik dönüşüm yapılarak aradaki bağıntı lineer hale getirilir. Bu logit dönüşümü doğrusallaştırma işlemini barındırdığı için doğrusallık şartı arar. Bu sebeple varsayımın gerçekleşmemesi bağımlı ve bağımsız değişken arasındaki ilişkilerin açıklanma gücünü düşürür.

iv) Lojistik regresyon modelinin hata terimleri arasında otokorelasyon yoktur. Diğer bir deyişle, hata terimleri serisel olarak ilişkili değildir.

v) Gözlem sayısı, tahmin edilecek parametre sayısından fazla olmalıdır.

vi) Bağımlı değişken değerleri kendi aralarında ilişkili değildir ve bu değerler 0 ile 1 aralığındadır.

3.2.2. Lojistik Regresyon İle Doğrusal (Lineer) Regresyon Arasındaki İlişki

Doğrusal regresyonda, bağımlı değişken sürekli ve normal dağılıma sahip olmalıdır ve bağımsız değişkenler normal dağılıma sahip bir kitleden çekilmelidir. Hata terimleri 0 ortalama ve sabit (δ^2) varyanslı normal dağılıma uygunluk göstermelidir. Hata terimleri arasında ilişki yoktur. Hata terimleri birbirinden bağımsızdır.

Lojistik regresyonda, bağımlı değişkenin sürekli olma zorunluluğu yoktur ve bağımlı değişken kesikli değerler alabilir veya sürekli bir değişken kesikli hale getirilebilir. Bağımlı değişkenin normal dağılım göstermesi şart olmamakla beraber bağımlı değişkenin binom dağılımı gibi kesikli dağılım gösterdiği durumlarda lojistik regresyon analizi kullanılabilir. Bağımlı değişkenin iki sonuçlu olduğu durumlarda doğrusal regresyon modelindeki sabit varyanslılık ve normallik gibi varsayımlar bozulmakta ve hata terimlerinin normal dağılıma değil binom dağılımına sahip olması sebebiyle gözlem varyansları da eşit olmamaktadır. Böyle durumlarda lineer regresyon analizi yerine lojistik regresyon analizi kullanılabilmektedir (İyit, 2003).

Lojistik regresyonu doğrusal regresyondan ayıran en önemli özelliklerden biri de lojistik regresyonda bağımlı değişkenin kategorik değişken olmasıdır. Ayrıca doğrusal regresyon analizinde bağımlı değişkenin değeri tahmin edilir iken, lojistik regresyonda bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı tahmin edilir.

3.2.3. Lojistik Regresyon Analizi İle Diğer Çok Değişkenli Yöntemler Arasındaki İlişki

Çok değişkenli analiz yöntemlerinden biri olan lojistik regresyon analizinin diğer çok değişkenli analiz yöntemlerine oldukça benzemekle beraber aralarında önemli farklılıklar vardır. Amaçlarından birisi bağımlı ve bağımsız değişkenler arasındaki ilişkileri araştırmak diğeri ise sınıflandırma olan lojistik regresyon analizinde, bağımlı değişken kategorik veri oluşturmakta ve kesikli değerler almaktadır. Gözlemleri yapılarına göre ait oldukları gruplara en doğru şekilde atamak için kullanılan lojistik regresyon analizi dışında kümeleme analizi ve diskriminant analizinde de benzer atama işlemi yapılmaktadır.

Ayırma (diskriminant) analizi verilerin sınıflandırılmasını ve belirli olasılıklara göre sayısı önceden belirlenmiş iki ve daha fazla sayıda gruplara ayrılmasını sağlayan yöntemdir. Buna karşılık; kümeleme analizinde, gözlemlerin atanacağı grup sayısı bilinmemekte olup, gözlemler benzerliklerine göre kümelenmektedir. Kümeleme analizinde gözlemlerin atanacağı grup sayısı bilinmezken, lojistik regresyon ve diskriminant analizinde grup sayısı bilinmektedir. Buradan hareketle; lojistik regresyon analizi, bağımlı değişkenin 2 ve daha fazla değere sahip olması ve gözlemlerin atanacağı grup sayısının bilinmesi sebebiyle diskriminat analizine benzemekle birlikte bazı noktalarda diskriminant analizinden farklılık gösterir. Çünkü diskriminant analizi çok değişkenli normallik, doğrusallık ve gruplar arası varyans-kovaryans matrisinin eşitliği gibi varsayımlara dayanır. Bu varsayımların sağlanamadığı durumlarda çok değişkenli regresyon analizi uygulanamaz. Lojistik regresyon analizi ise normal dağılım varsayımı, doğrusallık varsayımı gibi ön koşulları gerektirmeyen bir regresyon yöntemidir.

Buradan sonuçla, lojistik regresyonun normallik, doğrusallık, ortak kovaryansa sahip olma gibi varsayımlara uyulmaması durumunda diskriminant analizine, bağımlı değişkenin 0 ve 1 gibi iki değer aldığı veya ikiden çok şık içeren kesikli değişken olduğu durumlarda normal dağılım varsayımının ihlal edilmesi sebebiyle, doğrusal regresyona alternatif oluşturduğu söylenebilir.

3.2.4. Lojistik Regresyon Analizinin Kullanım Sebepleri

Lojistik regresyonun diğer analiz yöntemlerine göre daha çok tercih edilir olmasının nedenleri; (Büyüköztürk ve diğerleri, 2014: 60-61.)

i. Regresyon modelindeki bağımlı değişken kesikli iken bağımsız değişkenin hem sürekli hem de kesikli olduğu durumlarda uygulanabilir olması,
ii. Lojistik regresyon modelindeki parametrelerin kolay yorumlanabilir olması,
iii. Modeldeki bağımsız değişkenlerin olasılık fonksiyonlarının dağılımları üzerinde herhangi bir kısıt olmaması
iv. Lojistik regresyon modeline dayanalı analizleri yapabilmeyi sağlayan çok sayıda bilgisayar paket programı (SPSS, Minitab, vb.) kullanılabilmesi,
v. Lojistik regresyonda bütün olasılık değerlerinin ve 0 ile 1 arasında değişmesi,
vi. Bağımsız değişken ile bağımlı değişken arasındaki ilişkinin doğrusal olma zorunluluğunun olmaması,

lojistik regresyonun popüler bir yöntem olmasının bazı nedenleri olarak gösterilebilir.

3.2.5. Lojistik Regresyon Analizi Türleri

Lojistik regresyon analizinde bağımlı değişkenin kategori sayısına ve ölçeğine göre değişen üç temel yöntem vardır;

a) İkili (iki şıklı - Binary) Lojistik Regresyon Analizi
b) Çoklu (ikiden çok şıklı) Sıralayıcı (Ordinal) Lojistik Regresyon Analizi
c) Çoklu (ikiden çok şıklı) Sınıflayıcı (Nominal) Lojistik Regresyon Analizi

a) İkili (Binary) Lojistik Regresyon Analizi

Bağımlı değişkenin ikili cevap içerdiği yani iki kategorili olduğu durumlarda uygulanan lojistik regresyon analizidir. Bağımlı değişkenin olasılık değeri 0 ile 1

arasında değer alır. Buna karşılık, bağımsız değişkenler kesikli olabileceği gibi, sürekli, sıralanabilir ya da sırasız niteliksel değişken tiplerinde olabilmektedir.

b) Çoklu Sıralayıcı (Multiordinal) Lojistik Regresyon Analizi

Bağımlı değişkeni ikiden fazla değere sahip olup, bu şıkların kendi aralarında örneğin, kısa, orta ve uzun gibi bir sıralamaya tabi tutulduğu lojistik regresyon analizidir. Kategorilendirme yapılırken bir derecelendirme söz konusudur. Eğer söz konusu bağımlı değişkende kısa, orta, uzun gibi üç farklı sıralı kategori olduğu düşünülürse kısa 1 değerini ve orta 2 değerini alırken uzun ise 3 değerini alacaktır. Modelde bağımsız değişken üzerinde herhangi bir kısıtlama olmadığından bağımsız değişkenler faktörler kategorik ya da sürekli değişken olabilir.

c) Çoklu Sınıflayıcı (Multinomial) Lojistik Regresyon Analizi

Çoklu sınıflayıcı lojistik regresyon analizi, bağımlı değişkeni ikiden fazla değere sahip olan ve bu değerler arasında herhangi bir sıralama bulunmayan analiz yöntemidir. Örneğin, bir fakültede görülecek ders türü olarak, üç değerin olduğu istatistik, geometri ve matematik değerleri çoklu sınıflayıcı bir yapıya (multinominal) sahiptir.

3.2.6. Lojistik Regresyon Modeli

Lojistik regresyon modeli doğrusal regresyona alternatif bir modeldir.
Basit doğrusal regresyon modeli;

$$Y = \beta + \beta X_i + \varepsilon$$

eşitlik ile ifade edilir.

Doğrusal regresyon modelinde bağımlı değişken olarak ifade edilen koşullu ortalama $E(Y|x_i)$ sürekli bir rastgele değişkendir ve $(-\infty,+\infty)$ aralığında değer alır.

Basit regresyon modelinde, bağımlı değişkenin, verilen bağımsız değişken ya da değişkenlerin değerlerine göre beklenen değeri;

$$E(\hat{Y}|X) = \beta_0 + \beta_1 X_i$$

Doğrusal regresyon yöntemlerinde x, (−∞,+∞) aralığında değişen değerler alırken E(Y|X)'de mümkün olan her değeri alabilir. Ancak bağımlı değişken ikili değerler (0 ve 1) aldığında beklenen değerin 0 - 1 arasında değerler alması beklenir. Beklenen değeri 0 - 1 arasında sınırlı olasılık değerleri aldığından ve bu değerler, sonsuz değerler alabilen bağımsız değişkenlerle ilişkilendirildiğinden dolayı, bu eşitlik her zaman sağlanamamaktadır. Bu eşitliği sağlamak amacıyla sonuç değeri olarak ifade edilen olasılık değerini (−∞,+∞) aralığında tanımlı olacak şekilde bir dönüşüm yardımıyla 0 - 1 aralığına çekmek mümkündür. Lojistik regresyon modelinde bu dönüşüm lojit adı verilen dönüşüm ile sağlanır.

Bu aşamada kullanılacak basit lojistik modelin spesifik formu;

$$\pi(xi) = E(Y|X_i) = \frac{e^{\beta 0 + \beta 1 \, Xi}}{1+e^{\beta 0 + \beta 1 \, Xi}}$$

şeklinde tanımlanır. Modelin doğrusal olması için bir lojik dönüşüme ihtiyaç vardır. Aşağıdaki gibi lojik dönüşüm yapılırsa;

$$g(x_i) = ln\left(\frac{\pi(xi)}{1-\pi(xi)}\right)$$

$$= ln\,(\pi(xi)) - ln(1-\pi(xi))$$

$$= ln\,(\frac{e^{\beta 0 + \beta 1 \, Xi}}{1+e^{\beta 0 + \beta 1 \, Xi}}) - ln(\frac{1}{1+e^{\beta 0 + \beta 1 \, Xi}})$$

$$= ln\,(e^{\beta 0 + \beta 1 \, Xi})$$

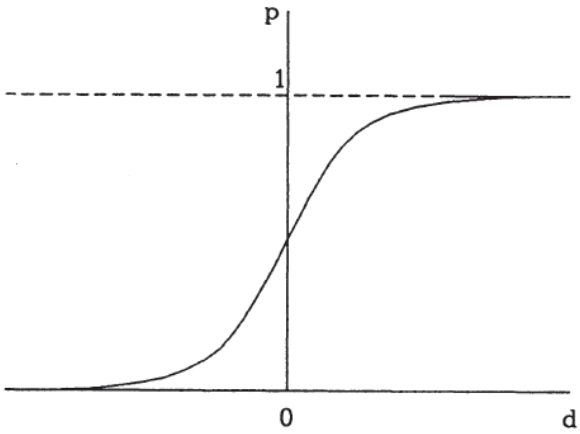
$$g(x_i) = \beta 0 + \beta 1 x_i$$

olur. $\frac{\pi(xi)}{1-\pi(xi)}$ oranına daha önceden de bahsettiğimiz gibi odds değeri denir. Bir olayın olma olasılığının olmama olasılığına oranıdır.

Dönüştürülmüş model g(x), doğrusal regresyon modelinin istenen tüm özelliklerini taşıması açısından önemlidir. Lojit g(x) , parametreleri bakımından doğrusal, sürekli ve x 'in aldığı değerlere bağlı olarak $(-\infty,+\infty)$ aralığında değişebilmektedir.

Olasılık π ile x bağımsız değişkenleri arasındaki ilişkiyi gösteren lojistik fonksiyon, normal dağılan bir olasılık fonksiyonuna benzemektedir. Ancak daha önceden de belirttiğimiz gibi olasılık ile bağımsız değişkenler arasındaki ilişki doğrusal değildir (Alpar, 2013: 645). Lojistik fonksiyonu S biçiminde bir eğriyi andırmaktadır (Jobson, 1992: 284);

Şekil 2: Lojistik Fonksiyonu



3.2.7. Lojistik Regresyon Analizinde Parametre Tahmini

Lojistik regresyon analizinde parametreler farklı tekniklerle tahmin edilebilirler. Bu teknikler arasında yaygın olarak kullanılan teknikler;

a) En Çok Olabilirlik Tekniği,
b) Yeniden Ağırlıklandırılmış İteratif En Küçük Kareler Tekniği,
c) Minimum Logit Ki-kare Tekniği

a) En Çok Olabilirlik Tekniği

Doğrusal regresyon yönteminde parametre tahminleri için en küçük kareler yöntemi kullanılır. EKK yöntemi ile modelden tahminlenen $\hat{Y}$ değerlerinin gözlemlenen Y değerlerinden sapmalarının karesini minimum yapan β_0 ve β_1 parametreleri tahmin edilir. Doğrusal regresyon modelinin varsayımları söz konusu iken, en küçük kareler tahminleri en iyi, doğrusal, sapmasız tahmincilerdir. Ancak daha önce de bahsedildiği gibi bağımlı değişken iki kategorili olduğunda hata terimleri normal dağılım gösteremeyeceğinden en küçük kareler yöntemi kullanılsa da buna dayalı hipotez testleri yapılamaz. Uygulamada birçok yöntem kullanılmakla beraber en çok kullanılanı Maksimum Olabilirlik Yöntemidir.

Maksimum olabilirlik yöntemi, gözlenen bağımlı değişken veri kümesini elde etmenin olasılığını maksimum yapan parametre değerlerini tahmin etmeye yarayan bir tekniktir. Bu metodu uygulamak için önce maksimum olabilirlik fonksiyonu oluşturulur. Oluşturulan fonksiyon, bilinmeyen parametrelerin bir fonksiyonu olarak gözlenen verinin olasılığını verir (Aksaraylı ve Saygın, 2011: 21-37).

Bu fonksiyon aracılığı ile elde edilen parametrelerin tahmini değerleri, fonksiyonu maksimum yapan değerleri bulacak şekilde seçilir. Bu nedenle sonuçta elde edilen tahminciler, gözlenen verilere çok yakın değerlere sahiptir.

İki kategorili bağımlı değişken durumunda; y_i sonuç değişkeni 0 ve 1 ile kodlandıysa, bu durumda $\pi(x)$ ifadesi, verilen herhangi bir x değeri için, y_i'nin 1'e eşit olma koşullu olasılığını vermektedir ve $\pi(x)=P(y = 1/x)$ ile ifade edilir. Buradan hareketle, $[1-\pi(x)]$ ifadesi de, verilen herhangi bir x değeri için y_i'nin 0 olma koşullu olasılığını göstermektedir ve bu olasılık da $[1-\pi(x)]=P(y=0/x)$ şeklinde gösterilir.

(x_i, y_i) çifti için, y_i =1 olduğunda olabilirlik fonksiyonuna katkısı $\pi(x_i)$ iken, y_i = 0 olduğu zaman ise olabilirlik fonksiyonuna katkısı 1- $\pi(x_i)$ kadardır. (xi, yi) ikilisi için olabilirlik fonksiyonunun dağılımının basit bir gösterimi aşağıdaki gibidir (Hosmer ve Lemeshow,2000:8):

$$\pi(x_i)^{y_i}\,[1-\pi(x_i)]^{1-y_i}$$

Gözlemlerin n tane olup bağımsız oldukları varsayıldığından, olabilirlik fonksiyonu aşağıdaki fonksiyondan elde edilir.

$$l(\beta) = \prod_{i=1}^{n} \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}$$

Maksimum olabilirlik ilkesi bu eşitliği maksimum yapan β değeri tahminini kullanmaktır. Bununla beraber bu eşitliğin logaritması ile çalışmak matematiksel olarak daha kolay olacağından log-olabilirlik fonksiyonu,

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^{n} \{y_i \, ln[\pi(x_i)] + (1 - y_i) \, ln[1 - \pi(x_i)]\}$$

Bu logaritmik fonksiyon ile β parametre tahminleri elde edilebilir. Log-olabilirlik fonksiyonunu maksimum yapan β değerlerini bulmak için log-olabilirlik fonksiyonunun β_0 ve β_1'e göre kısmi türevleri alındıktan sonra, elde edilen ifadeler 0'a eşitlenir. Sonuçta elde edilen eşitlikler aşağıdaki gibidir ve bu eşitlikler olabilirlik eşitlikleri olarak adlandırılır (Hosmer ve Lemeshow, 2000; 9).

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

Elde edilen β'nın değeri en çok olabilirlik tahmini olarak adlandırılır ve $\hat{\beta}$ olarak gösterilir. Örneğin, $\pi(x_i)$ 'nin en çok olabilirlik tahmincisi $\hat{\pi}(x_i)$ ile gösterilir. Bu değer, x 'in x_i gibi bir değere eşit olduğu bilindiği zaman, y_i sonuç değişkeninin 1'e eşit olma koşullu olasılığının tahminini verir.
2. eşitlikten;

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

eşitliği sağlanır. Bu eşitlik bize y 'nin gözlenen değerlerinin toplamının, tahmin edilen değerlerinin toplamına eşit olduğunu göstermektedir.

b) Yeniden Ağırlıklandırılmış İteratif En Küçük Kareler Tekniği

Gruplandırılmış verilerde j tane grubun her birinde n_j denemeden r_j başarı elde edildiğinde başarı oranı;

$$P_j = \frac{r_j}{n_j}$$

olarak tanımlanır. Ve varyansı;

$$\text{Var}\left(\frac{r_j}{n_j}\right) = \frac{Pj(1-Pj)}{n_j}$$

olacaktır. Bağımlı değişken binom dağılıma sahip olduğundan her binom dağılımlı gözlem için varyans değişmektedir. Bu durumda bağımsız değişkenler üzerinde $w_j = \frac{nj}{Pj(1-Pj)}$ ağırlığı ile ağırlıklandırılmış regresyon tahmini yapılmalıdır. Ancak w_j ağırlık değerleri de P_j 'nin bir fonksiyonu olduğu için EKK yöntemi iteratif olarak uygulanabilecektir ve ağırlık değerleri her adımda yeniden elde edilebilecektir. (Tatlıdil, 2002:296).

c) Minimum Logit Ki-kare Tekniği

Berkson tarafından geliştirilmiş olup, ağırlıklı en küçük kareler tahmin yönteminin özel bir biçimidir. 2×J kontenjans tablolarındaki beklenen ve gözlenen lojit değerleri arasındaki farktan yararlanılan minimum lojit ki-kare yöntemi tekrarlı veriler olması durumunda kullanılmaktadır. Yeniden ağırlıklandırılmış iteratif EKK yönteminde verilen P_j olasılığı üzerinden yapılan lojit dönüşümü, bu yöntemde bağımlı değişkenini oluşturmaktadır. Minimum Logit Ki-kare Tekniği logit değeri olarak tanımlanan bağımlı değişkenin ağırlıklandırılmış regresyonundan En Küçük Kareler tahminlerini bulmaya dayanmaktadır. Buradan tek adımda bulunan ağırlıklı EKK yöntemi tahminleri Minimum Logit Ki-kare tahminleri adını almaktadır.

3.2.8. Lojistik Regresyon Analizinde Parametrelerin Anlamlılık Testi

Çoklu ve basit lojistik regresyona ilişkin parametre tahminleri yapıldıktan sonra, modeldeki katsayıların anlamlılığının test edilmesi gerekmektedir. Modelde yer alan bağımsız değişkeni içeren modelin, o değişkeni içermeyen modele göre daha fazla bilgi içerip içermediği söz konusu bağımsız değişkene ait katsayının anlamlılık testi yapıldığında belirlenmiş olur. Bunun için bağımlı değişkene ait gözlenen değerler, her iki modelden elde edilen tahmin değerleriyle karşılaştırılır. Değişken modelde olduğunda daha doğru tahminler elde ediliyorsa, o değişkenin önemli olduğu sonucuna varılır. Modeldeki değişkenlere ilişkin parametrelerin anlamlılığını test etmek için kullanılan yöntemler;

a) Olabilirlik Oranı Testi
b) Wald testi
c) Skor Testi

a) Olabilirlik Oran Testi

Olabilirlik oran testi, modeldeki bağımsız değişkenlerin model için anlamlı olup olmadığına dair hipotezleri test edilir. Log-olabilirlik fonksiyonu ile gözlenen değerlerin ve tahmin edilen değerlerin karşılaştırma işlemi D istatistiği ile yapılmaktadır. D istatistiği, bir anlamda kurulan modelin önemliliğini test eder.

$G = Dx - Dy$

Dx = bağımsız değişkeni içermeyen model,

Dy = bağımsız değişkeni içeren model

G istatistiği:

$$G = -2 \ln \left[\frac{\text{bağımsız değişkeni içermeyen modelin olabilirliği}}{\text{bağımsız değişkeni içeren modelin olabilirliği}} \right]$$

şeklinde de ifade edilebilir.

$\beta 1 = 0$ hipotezi altında G istatistiği 1 serbestlik dereceli ki-kare ($\chi 2$) dağılımı gösterir. Elde edilen G değeri, p tablo değeri ile karşılaştırılır. Tablo değerinden küçük bulunan G değeri bağımsız değişkenin anlamlı olduğu sonucunu doğurur.

b) Wald Testi

Wald istatistiği doğrusal regresyon analizinde parametrelerin anlamlılığını test etmek kullanılan t istatistiği gibi, lojistik regresyon analizi modelindeki her bir β katsayısının ayrı ayrı istatistik olarak önemli olup olmadığını test etmek için kullanılır. Wald istatistiği eğim parametresinin maksimum olabilirlik tahmini β_i 'nin kendi standart hatasına oranıyla elde edilir.

$$W = \frac{\hat{\beta}_j}{S_{\hat{\beta}_j}}$$

Eğim parametresini gösteren $H_0 = \beta_J = 0$ hipotezi için Wald istatistiği standart normal dağılım gösterir. Normal rassal bir değişkenin karesinin alınması 1 serbestlik dereceli $\chi 2$ rassal değişkenine eşit olacağından Wald istatistiği aşağıdaki gibi de ifade edilebilir:

$$W^2 = Z^2 = \left[\frac{\hat{\beta}_j}{S_{\hat{\beta}_j}}\right]^2$$

verilen Wald test istatistiği 1 serbestlik dereceli $\chi 2$ dağılır ve $\chi 2$ dağılımının kritik değerleri ile karşılaştırılarak önemliliği belirlenir. Wald istatistik değeri tablo değerinden büyük ise sıfır hipotezi reddedilir ve parametrenin anlamlı olduğu sonucuna varılır.

Wald istatistiğinin dezavantajı büyük parametre tahminlerinde, tahmin edilen standart hata büyük çıkmakta, bu da H_0 hipotezinin kabul edilmesi olasılığını artırmaktadır. Böyle bir durumda, yanlış olan H_0 hipotezinin kabul edilmesi olasılığı yani II. tip hata olasılığı artar.(Menard, 2002:43).

c) Skor Testi

Lojistik regresyon modelindeki katsayıların anlamlılığını ölçmede kullanılan yöntemlerden biri de skor testidir. Diğer testler gibi yoğun hesaplamalar gerektirmeyen bir test olması skor testinin en büyük avantajı olurken, birçok paket programda bulunmaması bu testin dezavantajıdır. Skor testi, log olabilirliğin türevlerinin dağılım teorisine bağlı olup genel olarak matris hesapları gerektiren çok değişkenli bir testtir.

$$\text{S.T.} = \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y})}{\sqrt{\bar{y}(1-\bar{y})\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

3.2.9.Lojistik Regresyon Analizinde Model İçin Değişken Seçim Yöntemleri

Lojistik regresyon analizi uygulamasında öncelikle modele dahil edilecek bağımsız değişkenleri belirlemek gerekmektedir. Lojistik regresyon modellerinde bulunan bağımsız değişkenlerin tümünün her zaman bağımlı değişkeni açıklamada katkısı olmayabilir. Katkısı olmayan değişkenlerin denklemde tutulmaya devam etmesi denklemi etkilediğinden modelin etkinliğinin ve tahmin gücünün düşmesine sebep olmaktadır. Bu sebeple modele eklenmesi ile bağımlı değişkenin varyansını açıklamada önemli olan değişkenleri bulmak için çeşitli metotlar vardır. Lojistik regresyon modeli için standart (enter) ve adımsal (stepwise) olmak üzere iki temel değişken seçim yöntemi vardır. Adımsal yöntemler de kendi içerisinde ileriye doğru (forward) ve geriye doğru (backward) yöntemler olmak üzere iki gruba ayrılmaktadır.

Enter Yöntemi: Bütün değişkenler bir blok olarak tek aşamada regresyon modeline dahil edilir ve her blok için katsayı tahminleri yapılır.

Adımsal Yöntem: Adımsal yöntemler ileriye doğru (forward) ve geriye doğru (backward) olmak üzere iki gruba ayrılmaktadır. İleriye doğru yöntemlerde modele önce sadece sabit terim dâhil edilerek başlanır. Sonra belirli bir ölçüte göre, değişkenler modele tek tek eklenir. Buradaki ölçüt puan istatistikleridir. En önemli

puan istatistiğine sahip olan değişken ilk önce modele girer. İşlem, anlamlı istatistiği olan değişken kalmayıncaya kadar bu şekilde devam eder. Burada kesme noktası α=0,05'tir. Her bir adımda model dışı bırakılması gereken değişken olup olmadığına bakılır. Bu üç yolla gerçekleşir. Olabilirlik oranı ile ileriye doğru (forward likelihood ratio-forward) yöntemi, durum indeksi ile ileriye doğru (condition index) yöntemi ve Wald istatistiği ile ileriye doğru (forward wald) yöntemidir.

İleriye doğru yöntemlerin tersi ise geriye doğru yöntemlerdir. Öncelikle tüm değişkenler modele alınır, daha sonra koşulları sağlamayan değişkenler tek tek modelden çıkarılır. Değişkenler modelden atılırken koşullu katsayı tahminlerine dayanan olabilirlik oranının olasılığına göre karar verilir.

3.2.10. Lojistik Regresyon Analizinde Açıklama Katsayıları

Doğrusal regresyon analizinde belirlilik katsayısı olan R^2, modeli açıklama gücünü göstermek amacıyla kullanılır. Lojistik regresyon analizinde R^2 değerini, doğrusal regresyondaki gibi en küçük kareler yöntemi kullanarak hesaplamak mümkün olmadığından R^2 lojistik regresyon analizinde bulunmamaktadır. Bu nedenle lojistik regresyon analizi tekniğinde modelin uygunluğunun değerlendirilmesi için farklı R^2 istatistikleri geliştirilmiştir.

Cox ve Snell R2: Çok değişkenli regresyon analizindeki R2 istatistiğine benzemekte olup olabilirlik esasına dayanmaktadır. Cox-Snell R^2, minimum 0 değerini alırken, maksimum değeri 1'den küçük olmaktadır. Maksimum değerinin 1'den küçük olması Cox-Snell R^2 istatistiğinin en önemli dezavantajıdır. Maksimum değerini 1'den küçük olması istatistiğin yorumunu güçleştirmektedir. Cox ve Snell R2 istatistiği aşağıdaki şekilde elde edilmektedir.

$$R^2 = 1 - \left\{\frac{L(M_s)}{L(M_t)}\right\}^{2/N}$$

$L(M_s)$: sadece sabit terim içeren modelin olabilirliği

$L(M_t)$: tüm değişkenleri içeren modelin olabilirliği

Nagelkerke R^2: Cox-Snell R^2 istatistiğinin 0-1 aralığında değer almasını sağlamak amacıyla geliştirilmiş olup, bağımlı değişken ile bağımsız değişken arasındaki ilişkinin miktarını göstermek amacıyla kullanılmaktadır. Nagelkerke R^2 değeri, Cox-Snell R^2 değerinden her zaman daha yüksektir. Cox ve Snell R^2 istatistiğinin 0 – 1 aralığında değerler almasını sağlamak için Cox ve Snell R^2 istatistiği maksimum değeri olan $1 - L(M_s)^{2/N}$'e bölünmelidir. Bu düzeltme ile Cox-Snell R^2'nin en büyük değeri 1 olabilmektedir.
Nagelkarke R^2 istatistiği aşağıdaki gibi elde edilmektedir.

$$R^2 = \frac{1-\left\{\frac{L(M_s)}{L(M_t)}\right\}^{2/N}}{1 - L(Ms)^{2/N}}$$

McFaden R^2: Olabilirlik oranı indeksi olarak da bilinen bu değer doğrusal regresyon analizinde kullanılan R^2 değerine benzemekte olup, tüm değişkenleri içeren modelin, sadece sabit terimi içeren modele göre açıklayıcılık düzeyini göstermektedir. Bu yöntemde, bağımsız değişkenlerin olmadığı sadece sabit değişkenin bulunduğu modelin (M_s) log olabilirliği genel kareler toplamı, bağımsız değişkenlerin olduğu modelin (M_t) log olabilirliği de artık kareler toplamı olarak düşünüldüğünde, R^2 ölçüsü aşağıdaki gibi ifade edilmektedir.

$$R^2 = 1 - \left(\frac{ln\hat{L}(M_t)}{ln\hat{L}(M_s)}\right)$$

Burada $\hat{L}$ tahmin edilen olabilirliği ifade etmektedir.

Düzeltilmiş McFaden R^2: Doğrusal regresyonda kullanılan R^2'de olduğu gibi McFaden R^2 değeri de modele yeni değişken eklendiğinde artmaktadır. Düzeltilmiş McFaden R^2 aşağıdaki gibi elde edilmektedir.

$$R^2 = 1 - \frac{ln\hat{L}(M_t) - p}{ln\hat{L}(M_s)}$$

3.2.11. Lojistik Regresyon Analizinde Model Uyum İyiliğinin Belirlenmesi

Modelde olması gereken bütün değişkenler modele alındıktan sonra, bağımlı değişkeni açıklamadaki etkinliğini incelemeye uyum iyiliği adı verilir.

Aşağıda verilen 4 duruma göre uyum iyiliğinin test edilmesi gerekir (Tatlıdil, 2002:298).

i. Logaritmik dönüşüm yerine bir başka dönüşüm daha uygun olabilir.
ii. Logaritmik dönüşüm uygun olsa bile modeldeki bağımsız değişkenlerin bir kısmı uygun olmayabilir veya bazı etkileşim terimlerinin de modele dahil edilmesi gerekebilir,
iii. Değişkenlerin modelde bulunması uygundur, ancak ölçek hatalı olabilir.
iv. Modelde aykırı değerler olabilir.

Lojistik regresyonda modelin uygunluğunun belirlenmesinde normallik varsayımı sağlamaması ve parametre tahmin yöntemlerinin farklı olmasından dolayı t ve F testleri gibi parametrik testler kullanılamamakta olup doğrusal regresyon analizinden farklı yöntemlere başvurulur.

Lojistik modelin uyum iyiliğini araştırmada kullanılan bazı önemli ölçütler aşağıda verilmiştir (Tatlıdil, 2002:298).

i. Tüm değişkenleri içeren model ile tahmin edilen modele ilişkin olabilirlik oran değerlerinin farkına dayanan ölçütlerin χ^2 dağılacağı düşüncesinden hareketle, kurulan modelin geçerliliği sınanmaktadır. Bu yolla modele girecek bağımsız değişkenlere ve eklenecek karesel terimlere karar verilmektedir.
ii. Hata terimlerinin, x değerlerine veya olasılık değerlerine karşı çizimi ile aykırı değer araştırması yapılmaktadır.
iii. Hata kareler toplamı ve olabilirlik oranına dayalı R^2 türü ölçütler de model uyumunu test etmek amacıyla kullanılmaktadır.
iv. Lojistik model ayrımsama amacıyla kullanıldığında modelin doğru sınıflandırma oranı da bir uyum iyiliği ölçütüdür.

Lojistik regresyonda modelin uygunluğunun belirlenmesinde kullanılan metotlardan bazıları;

a) Pearson Ki-Kare İstatistiği
b) Hosmer-Lemeshow testi
c) Sınıflandırma tablosu

a) Pearson Ki-Kare İstatistiği

Lojistik regresyonda uyum iyiliği, her bir gözlemde başarı ve başarısızlığın gözlenen ve kestirilen olasılıkların karşılaştırılmasını sağlayan Pearson ki-kare istatistiğiyle de incelenebilir. Pearson ki-kare test istatistiği aşağıdaki gibi verilmektedir.

$$X^2 = \sum_{i=1}^{n} \frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i(1-\hat{\pi}_i)}$$

Burada $n_i\hat{\pi}_i$ başarıların beklenen sayısını, $n_i(1\text{-}n_i\hat{\pi}_i)$ ise başarısızlıkların beklenen sayısını belirtmektedir (i =1, 2,..., *n*) . Pearson ki-kare uyum iyiliği test istatistiği (n − p) serbestlik dereceli ki-kare dağılımına sahiptir. Pearson test istatistiğinin küçük değeri modelin uygun olduğunu gösterir.

b) Hosmer ve Lemeshow Testi

"Model Ki-Kare" olarak da adlandırılan bu istatistik lojistik regresyon modelinin bir bütün olarak uyumunu test etmektedir. Teste ilişkin sonucun anlamlı olmaması model veri uyumunu gösterir.

Bu yöntemde gözlemler kestirilen başarı olasılıklarına dayalı olarak g adet grupta sınıflandırılır. Yaklaşık olarak 10 grup kullanılır ve grupları sınır değerleri belirlenerek atamalar yapılır. Hosmer – Lemeshow test istatistiği, gözlenen ve beklenen frekanslardan oluşan, Pearson ki-kare istatistiği olarak hesaplanır.

$$HL = \sum_{i=1}^{n} \frac{(O_j - N_j\overline{\pi}_j)^2}{N_j\overline{\pi}_j(1-\overline{\pi}_j)}$$

Biçiminde ifade edilir. Eşitlikte;

N_j = j. gruptaki toplam gözlem sayısı,

O_j = j. gruptaki gözlenen başarıların sayısı

$\overline{\pi}_j$ = j. grupta kestirilen ortalama başarı olasılığı $\overline{\pi}_j = \sum_{i \in grup}(\pi_i / N_j)$

Hosmer – Lemeshow test istatistiğinde gözlenen başarıların sayısı O_j ve başarısızlıkların sayısı $N_j - O_j$, her bir gruptaki beklenen frekanslar olan $N_j\overline{\pi}_j$ ve $Nj(1-\overline{\pi}_j)$ ile karşılaştırılır. Eğer kestirilen lojistik regresyon modeli doğru ise ve örneklem hacmi büyükse Hosmer – Lemeshow istatistiği g-2 serbestlik dereceli ki-kare dağılımına sahiptir. Hosmer ve Lemeshow test istatistiği ilgili serbestlik derecesi ile Ki-Kare tablo değerinden küçük ise modelin uyumun iyi olduğuna karar verilir. Hosmer – Lemeshow istatistiğinin büyük değerleri verilerin modele yeterli uyumu sağlamadığını gösterir.

c) Sınıflandırma Tablosu

Model uyum iyiliği için kullanılan bir başka yöntem ise sınıflandırma tablolarıdır. Sınıflandırma tabloları bağımlı değişkenin gözlenen gerçek değerleri ile tahmin edilen değerlerinin çapraz sınıflama yapılarak gruplara ayrılmasıdır. Sınıflandırma tablosunda, bağımlı değişkenin gözlenen ve tahmin edilen lojistik olasılıklarından türetilen 0 veya 1 değerleri bulunmaktadır. Türetilen bağımlı değişken değerlerinin bulunması için öncelikle bir kesme değeri tanımlanır. Kesim noktası olarak genellikle 0,50 değeri alınmaktadır. Tahmin edilen değerler 0,5'in üstündeyse 1, değilse 0 değeri alır. Çapraz tablo yardımıyla, gerçekte sonucu pozitif olanların ne kadarının pozitif (duyarlılık), negatif olanların ne kadarının negatif (seçicilik) ve toplamda pozitif ve negatif sonuçların ne kadarının doğru sınıflandığını (doğruluk) hesaplanmaktadır. Model uyumu iyi olduğunda duyarlılık, seçicilik ve doğruluk değerlerinin yüksek olması beklenmektedir.

3.2.12. Lojistik Regresyon Analizinde Parametrelerin Yorumlanması

Model belirlendikten sonra, tahmin edilen katsayıların hesaplanması ve öneminin değerlendirilmesi kadar önemli bir konu da parametrelerin

yorumlanmasıdır. Genel olarak modelde tahmin edilen katsayılar ilgili değişkendeki bir birimlik değişimin, bağımlı değişken üzerinde katsayının işareti yönünde ve katsayı kadar değişim oranını gösterir. Doğrusal regresyon analizinde parametre tahminleri, doğrusallık nedeniyle bağımsız değişkendeki bir birimlik değişmenin bağımlı değişkende meydana getireceği etki kolayca açıklanabilirken lojistik regresyon analizinde bağımlı değişken ile bağımsız değişkenler arasında doğrusal bir ilişki olmadığından yorum yapmak daha zordur.

Lojistik regresyon analizinde, yapılan logaritmik dönüşüm ile bir olayın gerçekleşmesi ve gerçekleşmemesi olasılıkları oranı olan odds oranı, doğrusal regresyon modeli gibi ifade edilip, logaritmik olabilirlik değerinde β1 kadar bir artış veya azalış olacağı şeklinde yorumlanmaktadır. Yani bağımsız değişkendeki değişim bağımlı değişkendeki değişimi değil, bağımlı değişkenin olasılık değerinde meydana gelecek değişimi ifade etmektedir. Diğer taraftan katsayılar yorumlanırken ilgili bağımsız değişkenin kategorik veya sürekli oluşu veya kategorik ise kaç düzeyli olduğu yapılacak yorumu değiştirmektedir.

Modelde İki Şıklı Bağımsız Değişken Olması Durumu

Bağımsız değişkenin iki şıklı olduğu durum en basit durum olup diğer bütün durumlar için de detaylı bir temel teşkil edecektir. x'in 0 ve 1 kodlandığını varsayalım. Bu model altında π(x) ve 1-π(x)'in iki ayrı değeri vardır. Bağımsız değişkenin ikili değer aldığı 2 x 2 tablo aşağıdaki gibi gösterilirse;

Tablo 4: İkili Bağımsız Değişkene Sahip Lojistik Regresyon Modelinin Değerleri

Bağımlı Değişken (Y)	Bağımsız Değişken (x)	
	x= 1	x=0
Y=1	$\pi(1) = \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1+e^{\beta_0}}$
Y=0	$1\text{-}\pi(1) = \frac{1}{1+e^{\beta_0+\beta_1}}$	$1\text{-}\pi(0) = \frac{1}{1+e^{\beta_0}}$
Toplam	1	1

x =1 olduğunda denekler arasında bağımlı değişkenin görülme odds değeri $\pi(1) / [1- \pi(1)]$ olarak tanımlanır.

x=0 olduğunda denekler arasında bağımlı değişkenin görülme odds değeri $\pi(0) / [1- \pi(0)]$ olarak tanımlanır.

Odds değerlerinin logaritması lojit olarak adlandırılır ve odds değerlerine ait lojit fonksiyonları aşağıdaki gibidir.

$$g(1) = \ln \{\pi (1)/ [1-\pi (1)]\}$$
$$g(0) = \ln \{\pi (0)/ [1-\pi (0)]\}$$

x=1 için odds değerinin, x=0 için odds değerine oranı odds oranı olarak tanımlanır. Ve aşağıdaki gibi elde edilir.

$$\theta = \frac{\pi (1)/ [1-\pi (1)]}{\pi (0)/ [1-\pi (0)]}$$

Üstünlük oranının logaritması, log-odds oranı olarak adlandırılır ve aşağıdaki formülle bulunur:

$$ln(\theta) = \ln\left[\frac{\pi (1)/ [1-\pi (1)]}{\pi (0)/ [1-\pi (0)]}\right]$$
$$= \ln \{\pi (1)/ [1-\pi (1)]\} - \ln \{\pi (0)/ [1-\pi (0)]\}$$
$$= g(1) - g(0)$$

bulunur.

Lojistik regresyonda bağımsız değişkenin iki kategorili olması ve 0,1 olarak kodlanması durumunda odds oranını hesaplamak için Tablo 3.7. ile verilen değerler, odds oranı denkleminde yerlerine yazılırsa;

$$\theta = \frac{\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)\Big/\left(\frac{1}{1+e^{\beta_0+\beta_1}}\right)}{\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)\Big/\left(\frac{1}{1+e^{\beta_0}}\right)}$$

$$= \frac{\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)\left(\frac{1}{1+e^{\beta_0}}\right)}{\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)\left(\frac{1}{1+e^{\beta_0+\beta_1}}\right)}$$

$$= \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

elde edilir. Bunun anlamı; iki şıklı bağımsız değişkenin lojistik regresyonu için üstünlük oranı $\theta = e^{\beta_1}$ ve lojistik fark da $\ln(\theta) = \beta_1$ olarak bulunur.

Bu sonuç bize, 1 ve 0 olarak iki şıklı kodlanmış bağımsız değişkenli durumlarda, odds oranını lojistik regresyon katsayısı yardımıyla kolayca elde edilmesini sağlar.

Üstünlük oranının tahmini $\hat{\theta}$, eğik bir dağılıma sahiptir ve örneklem genişliği yeteri kadar büyük olduğu zaman, $\hat{\theta}$ 'nın dağılımı normal olacaktır. Üstünlük oranı için %100(1-α) güven aralığının kestirimi, β_1 katsayısı için güven aralığının alt ve üst noktalarının hesaplanmasından sonra bu değerlerin üssünün alınmasıyla elde edilir.

$$\exp\left[\hat{\beta}_1 \pm z_{1-a/2} \times SE\left(\hat{\beta}_1\right)\right]$$

Burada SE standart hatadır. Odds üstünlük oranının kestirim değerinin $\hat{\theta} = \hat{\beta}_1$ olması sadece bağımsız değişkenin 0 ve 1 olarak iki şıklı kodlandığı zaman doğrudur.

Özetleyecek olursak, ikili bir değişken için önemli olan parametre odds üstünlük oranıdır. Odds üstünlük oranı yaygın olarak kullanılan bir bağlantı ölçümüdür. x=1'in içinde sonucun olma olasılığının, x=0'ın içinde olma olasılığından ne kadar az veya çok olduğunun tahminin yapar. Lojistik regresyon katsayısı ve odds oranı arasındaki ilişki lojistik regresyon sonuçlarının yorumlanması için temel teşkil etmektedir.

Modelde İkiden Fazla Düzeyli Bağımsız Değişken Olması Durumu

Bağımsız değişkenin ikiden fazla düzeyli olduğu durumlarda (k > 2), nominal ölçekli bağımsız değişkenleri sürekli değişkenlermiş gibi modele dahil

etmek doğru olmayacaktır. Bu yüzden ikiden çok kategorisi olan bağımsız değişkenin, kukla değişkenler kullanılarak modele dahil edilmesi gerekmektedir.

Referans hücre metodunda, bir referans grup belirlenir ve referans grup için bütün kukla değişkenlerini sıfıra eşitlenir. Diğer grupların her biri için ise sadece bir tane kukla değişkenini 1'e eşitlerken, diğerlerini 0'a eşitlemek gerekmektedir. Örneğin ırklar için bir tablo oluşturulmak istenirse; (Hosmer ve Lemeshow, 2000; 57)

Tablo 5: İkiden Fazla Düzeyi Olan Bir Değişken İçin İlk Düzeyi Referans Grup Olarak Kullanarak Kukla Değişkenlerinin Oluşturulması

	Kukla Değişkenleri		
IRK	D_1	D_2	D_3
Beyaz(1)	0	0	0
Siyah(2)	1	0	0
İspanyol(3)	0	1	0
Diğerleri(4)	0	0	1

Siyahların beyazlara göre karşılaştırılması aşağıdaki gibidir;

$$\ln[\hat{\theta}(siyah,beyaz)] = \hat{g}(\text{siyah}) - \hat{g}(\text{beyaz})$$
$$= \left[\hat{\beta}_0 + \hat{\beta}_{11} \times (D_1 = 1) + \hat{\beta}_{12} \times (D_2 = 0) + \hat{\beta}_{13} \times (D_3 = 0)\right] -$$
$$= \left[\hat{\beta}_0 + \hat{\beta}_{11} \times (D_1 = 0) + \hat{\beta}_{12} \times (D_2 = 0) + \hat{\beta}_{13} \times (D_3 = 0)\right]$$
$$= \hat{\beta}_{11}$$

Üstünlük oranı için güven aralıkları ikili bağımsız değişkenlerde kullanılan yaklaşıma benzer bir yaklaşımla bulunur. İlk olarak log odds (lojistik regresyon katsayısı) için güven aralıklarının alt ve üst sınırları elde edilir ve daha sonra bu sınırların üssü alınarak üstünlük oranı için güven sınırları bulunur. Genellikle lojistik regresyon katsayısı için %100 (1-α) güven sınırları;

$$\hat{\beta}_{ij} \pm z_{1-a/2} \times S\hat{E}(\hat{\beta}_{ij})$$

şeklindedir.

Odds oranı için %100(1−α) güven aralığı bu sınırları üssü alınarak;

$$\exp\left[\hat{\beta}_{ij} \pm z_{1-a/2} \times S\hat{E}\left(\hat{\beta}_{ij}\right)\right]$$

elde edilir.

Referans hücre metodu literatürde kukla değişkenlerin kodlanmasındaki en yaygın kullanılan metottur. Bunun sebebi diğer yöntemlere göre daha kolay bir yöntem olması ve diğer grupların riskini, kontrol grubunun riskine göre tahmin etmesidir.

Kukla değişkenlerin kodlanmasında kullanılan bir diğer yöntem ise "ortalamadan sapma" yöntemidir. Bu yöntem "grup ortalamasının", "genel ortalamadan" sapmasının etkisini verir. Lojistik regresyon analizinde "grup ortalaması" grubun logiti ve "genel ortalama" ise ortalama logittir. Seçilen kategorilerin birisi için bütün kukla değişkenleri (-1)'e eşitlenir ve sonra kalan kategoriler için 0, 1 kodlaması yapılarak kukla değişkenleri elde edilir. Yine ırk örneğini dikkate alınacak olursa; (Hosmer ve Lemeshow, 2000; 59)

Tablo 6: İkiden Fazla Düzeyi Olan Bir Değişken İçin Ortalama Lojitten Sapma (Marjinal) Metodu Kullanarak Kukla Değişkenlerinin Oluşturulması

	Kukla Değişkenleri		
IRK	D_1	D_2	D_3
Beyaz(1)	-1	-1	-1
Siyah(2)	1	0	0
İspanyol(3)	0	1	0
Diğerleri(4)	0	0	1

Siyahların beyazlara göre log odds değerini tahmin işlemi aşağıdaki gibidir;

$\ln [\hat{\theta}(siyah,beyaz)] = \hat{g}$ (siyah)$- \hat{g}$ (beyaz)

$$= \left[\hat{\beta}_0 + \hat{\beta}_{11} \times (D_1 = 1) + \hat{\beta}_{12} \times (D_2 = 0) + \hat{\beta}_{13} \times (D_3 = 0)\right] -$$

$$= \left[\hat{\beta}_0 + \hat{\beta}_{11} \times (D_1 = -1) + \hat{\beta}_{12} \times (D_2 = -1) + \hat{\beta}_{13} \times (D_3 = -1)\right]$$

$$= 2\hat{\beta}_{11} + \hat{\beta}_{12} + \hat{\beta}_{13}$$

şeklindedir.

Odds oranı için güven aralığı bulunan denklemdeki katsayıların toplamı için güven limitlerinin alt ve üst noktalarının üssü alınarak elde edilir. Bu yöntemle kestirilen log odds ve standart hata değeri referans hücre metodu ile tahminlenen değerlerle aynı olacaktır.

Kategorik değişken en azından sıralı ölçeğe sahip olduğu zaman kullanılabilecek diğer bir kukla değişkeni yaratma yöntemi de ortogonal polinomlardır. Ortogonal polinomlar genellikle, aralıklı ölçeğe sahip bir bağımsız değişkenin artan seviyelerinin, cevap değişkenindeki eğilimini değerlendirmek için kullanılırlar. Tahminlenen lojistik regresyon katsayıları, bağımsız değişken ile logit arasındaki ilişkinin önemli sayılabilecek doğrusal, ikinci dereceden ve kübik elemanlara sahip olup olmadığını test etmeyi sağlar.

Bu yöntemlerle odds oranlarının tahmini, referans hücre yöntemine göre daha karmaşık hesaplamalar gerektirmektedir. Bu yüzden de referans hücre yöntemi daha yaygın olarak kullanılmaktadır.

Modelde Sürekli Bağımsız Değişken Olması Durumu

Lojistik regresyon modeli sürekli bir bağımsız değişken içerdiği zaman, tahmin edilen katsayıların yorumlanması değişkenin modele nasıl girdiğine bağlıdır. Sürekli bir değişken için katsayının yorumlanması amacıyla geliştirilecek yöntemlerde lojitin sürekli değişkenle doğrusal olduğu varsayılacaktır.

Logitin sürekli değişkenle (x) doğrusal olduğu varsayımı altında, logit için eşitlik;

$$g(x) = \beta_0 + \beta_1 x$$

şeklindedir. Eğim katsayısı (β_1), x 'deki 1 birimlik değişimin log odds değerinde meydana getireceği değişimi verir. x 'in herhangi bir değeri için,

β_1= g(x+1) –g(x)

dir. Sürekli ölçekli bir birlikte değişene ait geçerli yorumu sağlayabilmek için birlikte değişkendeki "c" birimlik bir değişim için nokta ve aralık tahmin yöntemlerinin geliştirilmesi zorunluluğu vardır.

x 'deki "c" birimlik bir değişim, log odds değeri lojit farktan elde edilir.

$c\beta_1$= g(x+c) –g(x)

olur. Ve üstünlük oranı da bu logit farkının üssü alınarak elde edilir.

θ(c) = θ (x+c, x)= exp ($c\beta_1$)

θ(c)′nin tahmini için %100 (1−α) güven aralığın alt ve üst sınır noktaları,

$\exp\left[c\hat{\beta}_1 \pm z_{1-a/2} \times cS\hat{E}\ (\hat{\beta}_1)\right]$

olarak bulunur.

3.2.13. Çok Değişkenli Lojistik Regresyon Analizi

Birçok regresyon tekniğinde olduğu gibi lojistik regresyon analizinde de birden fazla bağımsız değişkenin bağımlı değişken üzerinde olan etkisi incelenebilmektedir. İki veya daha fazla bağımsız değişken olması durumunda çok değişkenli lojistik regresyon analiz yönteminden yararlanılmaktadır. Çok değişkenli lojistik regresyon analizi yöntemi, tek değişkenli regresyon analizinin genişletilmiş hali olarak değerlendirilebilir.

Çok değişkenli lojistik regresyon modeli için en önemli şey modeldeki katsayıların tahmini ve katsayıların anlamlılıklarının test edilmesi olacaktır. X' = (x_1, x_2,……,x_p) vektörü ile gösterilen p tane bağımsız değişken toplandığı ve bu

değişkenlerin sürekli olduğunu varsayılsın.Ayrıca bağımlı değişkenin (Y_i) 0 ve 1 olarak kodlanarak iki sonuçlu olduğu varsayılsın. (Yi =1) durumundaki koşullu olasılık P(Y=1/x) =π (x)'e eşittir. Çok değişkenli lojistik regresyon modelinin logiti ise aşağıdaki gibi elde edilmektedir.

$$g(x)= \beta_0 +\beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_P X_P$$

Bu durumda lojistik regresyon modeli;

$$\pi(x) = \frac{e^{g(x)}}{1+e^{g(x)}}$$

şeklindedir. Odds oranı ise;

$$\theta = \frac{\pi(x_i)}{1-\pi(x_i)}$$

şeklindedir. Aşağıdaki gibi lojik transformasyon yapılırsa;

$$\begin{aligned} g(x) &= \ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) \\ &= \ln\left(\frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots\ldots\ldots+\beta_p X_P}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots\ldots\ldots+\beta_p X_P}}\right) - \ln\left(\frac{1}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots\ldots\ldots+\beta_p X_P}}\right) \\ &= \ln(e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots\ldots\ldots+\beta_p X_P}) \\ &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots\ldots\ldots + \beta_p X_P \end{aligned}$$

olarak bulunur. Her bir parametrenin exp(β_i) i=1,2,...,n değeri odds değeri olarak değerlendirilebilir. Hesaplanan odds değeri Y değişkeninin X_p değişkeninin etkisi ile kaç kat daha fazla ya da yüzde kaç oranla fazla gözlenme olasılığına sahip olduğunu gösterir (İyit, 2003).

Bağımsız değişkenlerden bazıları kesikli nominal ölçekli (cinsiyet, ırk, vb.) ise bunları sürekli değişken gibi kabul ederek modele dahil etmek doğru olmayacaktır. Burada çeşitli seviyeleri göstermek için kullanılan sayılar yalnızca

tanımlayıcıdır ve herhangi bir sayısal değerleri bulunmamaktadır. Bağımsız değişkenler sayısal olarak sınıflandırıldığında farklı kukla değişkenleri kesikli olan bu değişkenleri temsil etmek amacıyla kullanılmaktadır.

Örneğin, ırk değişkenini bağımsız değişkenlerden biri olduğunu düşünülürse, değişkene ait 3 kategori "siyah", "beyaz" veya "diğerleri" olarak kodlanmış olsun. Irk değişkeninin şık sayısı 3 olduğu için D1 ve D2 olmak üzere, 2 tane kukla değişken kullanılması gerekecektir. Genel olarak, eğer nominal bir değişken k kategoriye sahipse, o zaman k -1 tane kukla değişkenine ihtiyaç vardır. J'inci x_j bağımsız değişkeninin k_j tane kategoriye sahip olduğunu düşünelim. k_j-1 kukla değişken, D_{ju} şeklinde gösterilir ve parametreleri de β_{ju} (u=1,2,…., k_j-1) şeklinde gösterilmek üzere, P değişkenli ve j'inci değişkeni kesikli olan bir model için lojit aşağıdaki gibidir.

$$g(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \sum_{u=1}^{k_j-1} \beta_{ju} D_{ju} + \beta_p X_P$$

Çoklu lojistik modellerde ikili lojistik modellerin kullanılabilmesi de mümkündür. Örneğin, bağımlı değişken y=0,1,2 gibi üç seviyeli olsun. Burada iki tane ikili lojistik regresyon modeli bulunmaktadır. Birincisi y=1'e karşı y=0 için ikincisi ise y=2'ye karşı y=0 içindir. Yani y=0 temel grup iken y=2'ye karşı y=1'i karşılaştıran lojistik fonksiyon iki karşılaştırmaya ilişkin lojistik fonksiyonların farklarından elde edilmektedir.

$$g_1(x) = \log\left(\frac{P(y\,1/x)}{P(y\,0/x)}\right) = \beta_{10} + \beta_{11}X_1 + \beta_{12}X_2 + \cdots\ldots\ldots + \beta_{1p}X_P$$

$$g_2(x) = \log\left(\frac{P(y\,2/x)}{P(y\,0/x)}\right) = \beta_{20} + \beta_{21}X_1 + \beta_{22}X_2 + \cdots\ldots\ldots + \beta_{2p}X_P$$

şeklindedir.

Sonuç değerleri için koşullu olasılıklar üç grup durumunda j=0,1,2 için aşağıdaki gibidir.

$$\pi_j(x) = \pi(y = j / x) = \frac{\exp(g_j(x))}{\sum_{k=0}^{2} \exp(g_k(x))} \qquad j = 0,1,2,$$

Birbirinden bağımsız n tane (x_i , y_i) i =1,2,..., n olmak üzere gözlem çiftinin olduğu varsayılsın. Modelin kurulması için tek değişkenli modelde olduğu gibi kestirim vektörünün $\beta'=(\beta_0, \beta_1, ..., \beta_p)$ elde edilmesi gerekmektedir. Çok değişkenli lojistik regresyon modelinde de tek değişkenli lojistik regresyon modelinde olduğu gibi, parametre tahmininde en çok olabilirlik yöntemi kullanılacaktır. Çok değişkenli lojistik regresyon modeli katsayılarının en çok olabilirlik yöntemine göre tahmin edilmesi için öncelikle olabilirlik fonksiyonunun oluşturulması gerekmektedir. Çok değişkenli lojistik regresyon modeli için log olabilirlik fonksiyonunun p+1 katsayıya göre türevi alınarak, değişken sayısından bir fazla olmak üzere (p +1) tane olabilirlik denklemi elde edilir.

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

$$\sum_{i=1}^{n} x_{ij}\left[y_i - \pi(x_i)\right] = 0 \qquad j = 1,2, \dots , p$$

Olabilirlik denklemlerinin çözümü için tek değişkenli lojistik regresyon modelinde olduğu gibi bilgisayar programlarından yararlanılabilir. $\hat{\beta}$ olabilirlik denklemlerinin çözümü sonucunda elde edilen tahmin vektörüdür. $\hat{\pi}_j(x)$ ise lojistik regresyon modeli için tahminlenen değerleri ifade eder. Tahmin edilen katsayıların varyans ve kovaryanslarının tahminleri ise elde edilen log-olabilirlik fonksiyonlarının ikinci derecen kısmi türevleri alınarak elde edilen matris yardımıyla bulunur. Söz konusu türevlerin genel şekli aşağıdaki gibidir.

$$\frac{\partial^2 L(\beta)}{\partial \beta^2_j} = -\sum_{i=1}^{n} x^2{}_{ij}\, \pi_i(1-\pi_i)$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_u} = -\sum_{i=1}^{n} x_{ij} x_{iu}\, \pi_i(1-\pi_i)$$

Bu eşitliklerle verilen ifadelerin negatiflerini içeren I(β) şeklinde gösterilen (p +1) x (p +1) boyutundaki matris "bilgi" (information) matrisi olarak adlandırılır. Tahminlenen katsayıların varyans ve kovaryansları bu matrisin tersinden elde edilir ve $I^{-1}(\beta) = \sum \mathrm{I}(\beta)$ 'dir.

Bilgi matrisi, $I(\hat{\beta})$ = X'VX formundadır. X matrisi [n × (p +1)] boyutunda bir matristir ve her bir denek için verileri içermektedir.

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \ldots & x_{2p} \\ \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & \ldots & x_{np} \end{bmatrix}$$

V matrisi [n × n] boyutunda genel elemanı $\hat{\pi}(1-\hat{\pi}_i)$, i =1,2, ... ,n olan diagonal (köşegen) bir matristir.

$$V = \begin{bmatrix} \hat{\pi}_1(1-\hat{\pi}_1) & 0 & \ldots & 0 \\ 0 & \hat{\pi}_2(1-\hat{\pi}_2) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & \ldots & \hat{\pi}_n(1-\hat{\pi}_n) \end{bmatrix}$$

Bilgi matrisi model kurulurken ve modelin uyum iyiliği belirlenirken kullanılır (İyit, 2003).

DÖRDÜNCÜ BÖLÜM
UYGULAMA

4.1. ARAŞTIRMA VE YÖNTEM

4.1.1. Araştırmanın Amacı

Bu çalışmada bireylerin sigara içmesine neden olabilecek bağımsız değişkenler yardımıyla Türkiye'deki kişilerin sigara içmesinde etkili olan faktörlerin belirlenmesi için bir analiz yapılmıştır. Çalışmamızda bağımlı değişken iki düzeyli kategorik değişken olduğundan, bu tür verilerin analizinde uygulanan lojistik regresyon analizi kullanılarak, sigara içmede etkili olan en önemli değişkenlerin belirlenmesi amaçlanmıştır.

4.1.2. Araştırmanın Evreni ve Örneklemi

Araştırmanın evrenini, Türkiye'ye bağlı 79 ilde yaşayan 18 yaş ve üzeri 6228 erkek ve 7969 kadın olmak üzere 14197 kişi oluşturmaktadır.

4.1.3. Araştırmanın Yöntemi

Çalışma kapsamında yapılan görüşmeler, 79 ilde önceden belirlenmiş adreslerde gerçekleştirilmiştir. Katılımcı seçimi anketörlere verilecek olan linkler üzerindeki bir program sayesinde yapılmıştır. Anketörler adres veri tabanının bulunduğu rastgele hane halkı bireyi seçimi sağlayan programa girdiklerinde, görüşmenin yapılacağı adresi veri tabanından seçmiş ve hanede yaşayan birey sayısı ile hanede yaşayan 18 yaş (18 yaşını doldurup 19 yaşından gün almış) ve üstü birey sayısını programa kaydetmiştir.

Kişi sayısı girildikten sonra hanede bulunan kişilerin adı, soyadı ve doğum yılı girilmiş. Hanedeki tüm bireylerin bilgisi kaydedildikten sonra program rassal olarak bir katılımcı belirlemiştir. Bu belirlenen katılımcı ile görüşme yapılması için izin istenmiş ve onaylaması halinde görüşme gerçekleşmiştir. Her haneden bir kişi

çevrimiçi bir yazılım üzerinden rastgele seçilmiş ve seçilen kişinin görüşmeyi kabul etmesi durumunda anket gerçekleşmiştir.

Veriler bilgisayar destekli yüz yüze görüşme tekniği CAPI (Computer Assisted Personal Interviewing) ile tablet bilgisayarlar aracılığı ile toplanmıştır.

Adreslerde katılımcı seçiminin gerçekleştirileceği hane bireylerini seçmek için daimi ikametgâh adresi ve hane halkı kavramları tanımlanmıştır. Buna göre; bir kişinin daimi ikametgâhı sürekli yaşadığı ya da yıl içerisinde en uzun süre kaldığı yerdir, adrestir. Çalışma kapsamında, bir kişi yıl içinde birden fazla adreste ikamet ediyorsa, en uzun süre kaldığı adres daimi ikametgâh adresi olarak kabul edilmiştir. Ayrıca bir kişinin en az 6 ay kalmak niyetiyle bulunduğu yer de daimi ikamet adresi olarak kabul edilmiştir.

Hane halkı, aralarında akrabalık bağı bulunsun veya bulunmasın aynı konutta ikamet eden bir veya birkaç kişinin oluşturduğu topluluktur. Bu ifadeye dayalı olarak 18 yaş ve üstü; askerliğini yapan veya ailesinin yanından ayrılıp şehir dışında okuyan öğrenciler hane halkına dâhil edilmemiştir. Hane halkı tanımına uymayan kişiler, katılımcı seçme linkinde o adreste ikamet eden kişiler olarak yazılmamıştır.

Anket uygulaması sonucunda elde edilen veriler SPSS 25.0 programında değerlendirilmiştir.

Araştırmada öncelikle kişilerin sosyo-demografik verilerine ilişkin bulgular frekans tabloları ile verilmiş olup, kişilerin yaşları, eğitim durumları, medeni durumları vb. bilgiler tanımsal istatistikler ile belirtilmiştir. Kişilerin sigara kullanımına ilişkin verileri kadınlara ve erkeklere göre ayrı ayrı frekans tablolarında verilmiştir.

Verilerin normal dağılıma uygunluğu Kolmogorov – Smirnov Z testi ile test edilmiş normal dağılıma uymadığı tespit edilmiştir.

Analizde bağımlı değişken sigara kullanılması ve sigara kullanılmamasıdır. Bağımlı değişkenin iki şıklı değişken olmasından dolayı ikili (binary) lojistik regresyon analizinden yararlanılmıştır. SPSS programında lojistik regresyon analizi enter (tam) ve stepwise (adımsal) olarak 2 şekilde yapılabilmektedir. Bu çalışmadaki analiz sonuçları enter yöntemi ile incelenmiştir.

Öncelikle tüm değişkenler için basit lojistik regresyon analizi uygulanmış anlamlı olmayan değişkeler denklemden çıkarılmıştır. Tek değişkenli basit lojistik

regresyon analizi sonunda önemli bulunan değişkenlerle çok değişkenli lojistik regresyon analizi uygulanmıştır.

Çoklu lojistik regresyon analizi uygulandıktan sonra modelin uyum iyiliği Hosmer – Lemeshow testi ile test edilmiştir.

4.2.ARAŞTIRMANIN TEMEL ANALİZLERİ

4.2.1. Sosyo-demografik ve Kategorik Verilere İlişkin Frekans Tabloları

Bu bölümde araştırmaya katılan kişilerin sosyo-demografik bilgilerine ve özelliklerine ilişkin çeşitli frekans tablolarına yer verilmiştir.

Tablo 7: Sosyo-demografik verilere ilişkin tanımlayıcı istatistikler

		Frekans (n)	Yüzde (%)
Cinsiyet	Kadın	7969	56,1
	Erkek	6228	43,9
Yaş	18-29	3302	23,3
	30-39	3175	22,4
	40-49	2552	18,0
	50-59	2222	15,7
	60-69	1806	12,7
	70 ve üstü	1140	8,0
Medeni Durum	Bekar	4612	32,5
	Evli	9585	67,5
Eğitim Durumu	Okuma yazma bilmiyor/okuma yazma biliyor fakat bir okul bitirememiş	1604	11,3
	İlkokul mezunu, ilköğretim, ortaokul veya mesleki ortaokul mezunu	6784	47,8
	Lise veya mesleki lise mezunu	3776	26,6
	Yüksekokul, fakülte mezunu	1934	13,6
	Yüksek lisans, doktora mezunu	99	0,7
Yerleşim Birimi	Kent	11661	82,1
	Kır	2536	17,9
Çalışma Durumu	Çalışıyor	4126	29,1
	Çalışmıyor	10071	70,9
	Toplam	14197	100,0

Tablo 7 incelendiğinde araştırmaya 14197 kişinin katıldığı ve katılan kişilerin %56,1'inin kadın % 43,9'unun ise erkek olduğu görülmektedir.

Araştırmaya katılan kişilerin yaş gruplarına bakacak olursak araştırmaya katılan kişilerin çoğunluğunu %23,3 ile 18-29 yaş arası kişiler ve %22,4 ile 30-39 yaş kişiler oluşturmaktadır.

Tablo 7 incelendiğinde çalışmaya katılan kişilerin %67,5'i evli kişilerden oluşmaktadır.

Araştırmaya katılan 14197 kişiden %11,3'ü okuma yazma bilmiyor, %47,8'i İlkokul, ilköğretim, ortaokul veya mesleki ortaokul mezunu, %26,6'sı lise veya mesleki lise mezunu, %13,6'sı yüksekokul, fakülte mezunu, %0,7'si ise yüksek lisans doktora mezunudur.

Çalışmaya katılan kişilerin %82,1'i kent bölgesinde yaşıyor olup %19,9'u ise kır bölgesinde yaşamaktadır.

Tablo 7 incelendiğinde çalışmaya katılan kişilerin %29,1'inin gelir getirici bir işte çalıştığı kalan 70,9'nun ise çalışmadığı görülmektedir.

Tablo 8: Diğer değişkenlere göre tanımlayıcı istatistikler

		Frekans (n)	**Yüzde (%)**
Solunum Yolları Hastalıkları (Astım-KOAH)	Var	506	3,6
	Yok	13691	96,4
Depresyon	Var	171	1,2
	Yok	14026	98,8
Kamu Spotu Etkisi	Düşünüyor	1246	8,8
	Düşünmüyor	12951	91,2
Uyku Düzeni	Çok Düzensiz	494	3,5
	Düzensiz	2064	14,5
	Bazen düzenli, bazen düzensiz	4481	31,6
	Düzenli	6672	47,0
	Çok düzenli	486	3,4
Sigara Dışı Tütün Kullanımı	Kullanıyor	446	3,1
	Kullanmıyor	13751	96,9
	Toplam	14197	100,0

Tablo 8 incelendiğinde çalışmaya katılan kişilerin %3,6'sında tanısı konulmuş solunum yolu hastalığı mevcuttur ve çalışmaya katılan 14197 kişinin %1,2'sinde tanısı konulmuş depresyon hastalığı bulunmaktadır.

14197 kişiden %8,8'i Sağlık Bakanlığının radyo ve televizyonlarda yer alan kamu spotu ve reklam kampanyalarının, sigara kullanımı üzerinde bir davranış

değişikliği oluşturduğunu düşünürken, %91,2'si kamu spotu ve reklam kampanyalarının davranış değişikliği üzerinde etkisi olmadığını düşünmektedir.

Çalışmaya katılan kişilerin %3,5'inin uyku düzeninin çok düzensiz, %14,5'inin düzensiz, %31,6'sının bazen düzenli bazen düzensiz, %47'sinin düzenli, %3,4'ünün çok düzenli olduğu görülmektedir. Çalışmaya katılan kişilerin %0,7'si ise uyku düzeni hakkında fikri olmadığını belirtmiştir.

Tablo 8 incelendiğinde çalışmaya katılan kişilerin %3,1'i nargile, sarma sigara, elektrik sigara, puro vb. sigara dışı tütün ürünleri kullanırken %96,9'u bu tütün ürünlerini kullanmamaktadır.

4.2.2. Sigara Kullanım Oranına İlişkin Frekans Tabloları

Çalışmaya konu olan sigara kullanımına ilişkin frekans tabloları aşağıdaki gibi verilmiştir.

Tablo 9: Sigara kullanım oranı frekans tablosu

	Frekans	**Yüzde %**
İçmiyor (0)	10493	73,9
İçiyor (1)	3704	26,1
Total	14197	100,0

Çalışmaya katılan 14197 kişinin %26,1'i sigara kullanıyorken geriye kalan %73,9'u sigara kullanmamaktadır.

Tablo 10: Demografik verilere göre kadınların sigara içme oranı

		İçiyor	İçmiyor	Toplam
Yaş	18-29	298 (%17,2)	1436 (%82,8)	1734
	30-39	411 (%21,6)	1495 (%78,4)	1906
	40-49	366 (%23,8)	1172 (%76,2)	1538
	50-59	200 (%15,5)	1093 (%84,5)	1293
	60-69	77 (%8,2)	857 (%91,8)	934
	70 ve üstü	35 (%6,2)	529 (%93,8)	564
Medeni Durum	Bekar	421 (%18,4)	1871 (%81,6)	2292
	Evli	966 (%17,0)	4711 (%83,0)	5677
Eğitim Durumu	Okuma yazma bilmiyor/okuma yazma biliyor fakat bir okul bitirememiş	119 (%9,3)	1164 (%90,7)	1283
	İlkokul mezunu, ilköğretim, ortaokul veya mesleki ortaokul mezunu	630 (%15,9)	3336 (%84,1)	3966
	Lise veya mesleki lise mezunu	437 (%24,8)	1322 (%75,2)	1759
	Yüksekokul, fakülte mezunu	192 (%20,9)	725 (%79,1)	917
	Yüksek lisans, doktora mezunu	9 (%20,5)	35 (%79,5)	44
Yerleşim Birimi	Kent	1251 (%18,2)	5610 (%81,8)	6861
	Kır	136 (%12,3)	972 (%87,7)	1108
Çalışma Durumu	Çalışıyor	325 (%31,0)	723 (%69,0)	1048
	Çalışmıyor	1062 (%15,3)	5859 (%84,7)	6921
	Toplam	1387 (%17,4)	6582 (%82,6)	7969

Tablo 10 incelendiğinde 18-29 yaş aralığındaki kadınların %17,2'sinin sigara kullanırken %82,8'inin sigara kullanmadığı görülmektedir. Kadınlarda sigara içen kişi sayısının en çok 30-39 yaş aralığındaki kişilerden oluştuğu görülmektedir.

Çalışmaya katılan kadınların büyük çoğunluğunu evli kadınlar oluşturmakla beraber evli kadınların %17'sinin sigara kullandığı görülmektedir.

Eğitim durumuna bakacak olursak çalışmaya katılan kadınların büyük çoğunluğu İlkokul mezunu, ilköğretim, ortaokul veya mesleki ortaokul mezunu olduğu görülmektedir ve İlkokul mezunu, ilköğretim, ortaokul veya mesleki ortaokul mezunu olan kadıların %15,9'u sigara kullanırken % 84,1'i sigara kullanmamaktadır.

Yine tablo 10 incelendiğinde çalışan kadınların 3 te 1'inin sigara kullandığını söyleyebiliriz.

Tablo 11: Diğer değişkenlere göre kadınların sigara içme oranı

		İçiyor	İçmiyor	Toplam
Solunum Yolları Hastalıkları (Astım-KOAH)	Var	65 (%19,7)	265 (%80,3)	330
	Yok	1322 (%17,3)	6317 (%82,7)	7639
Depresyon	Var	27 (%20,5)	105 (%79,5)	132
	Yok	1360 (%17,4)	6477 (%82,6)	7837
Kamu Spotu Etkisi	Düşünüyor	147 (%21,9)	525 (%78,1)	672
	Düşünmüyor	1240 (%17,0)	6057 (%83,0)	7297
Uyku Düzeni	Çok Düzensiz	63 (%23,2)	208 (%76,8)	271
	Düzensiz	276 (%22,7)	941 (%77,3)	1217
	Bazen düzenli, bazen düzensiz	449 (%17,6)	2106 (%82,4)	2555
	Düzenli	567 (%15,3)	3131 (%84,7)	3698
	Çok düzenli	32 (%14,0)	196 (%86,0)	228
Sigara Dışı Tütün Kullanımı	Kullanıyor	63 (%100)	0 (%0,0)	63
	Kullanmıyor	1324 (%16,7)	6582 (%83,3)	7906
	Toplam	1387 (%17,4)	6582 (%82,6)	7969

Tablo 11 incelendiğinde tanısı konulmuş solunum yolları hastalığına sahip kadınların %19,7'si sigara kullanmaktadır. Yine tanısı konmuş depresyon hastalığına sahip olan kadınların % 20,5'i sigara kullanırken %79,5'inin sigara kullanımından uzak durdukları görülmektedir. Kadınların büyük bir çoğunluğu Sağlık Bakanlığının radyo ve televizyonlarda yer alan kamu spotu ve reklam kampanyalarının sigara kullanımı üzerinde etkisi olmadığını düşünmektedir. Kamu spotu reklamlarının sigara kullanımına etkisi olduğunu düşünen kadınların %78,1'inin sigara

kullanmadığı görülmektedir. Sigara dışı tütün ürünleri kullanan kadıların hepsinin sigara kullandığı görülmektedir.

Tablo 12: Demografik verilere göre erkeklerin sigara içme oranı

		İçiyor	İçmiyor	Toplam
Yaş	18-29	672 (%42,9)	896 (%57,1)	1568
	30-39	559 (%44,1)	710 (%55,9)	1269
	40-49	438 (%43,2)	576 (%56,8)	1014
	50-59	343 (%36,9)	586 (%63,1)	929
	60-69	207 (%23,7)	665 (%76,3)	872
	70 ve üstü	98 (%17,0)	478 (%83,0)	576
Medeni Durum	Bekar	884 (%38,1)	1436 (%61,9)	2320
	Evli	1433 (%36,7)	2475 (%63,3)	3908
Eğitim Durumu	Okuma yazma bilmiyor/okuma yazma biliyor fakat bir okul bitirememiş	94 (%29,3)	227 (%70,7)	321
	İlkokul mezunu, ilköğretim, ortaokul veya mesleki ortaokul mezunu	1043 (%37,0)	1775 (%63,0)	2818
	Lise veya mesleki lise mezunu	820 (%40,7)	1197 (%59,3)	2017
	Yüksekokul, fakülte mezunu	346 (%34,0)	671 (%66,0)	1017
	Yüksek lisans, doktora mezunu	14 (%25,5)	41 (%74,5)	55
Yerleşim Birimi	Kent	1817 (%37,9)	2983 (%62,1)	4800
	Kır	500 (%35,0)	928 (%65,0)	1428
Çalışma Durumu	Çalışıyor	1361 (%44,2)	1717 (%55,8)	3078
	Çalışmıyor	956 (%30,3)	2194 (%69,7)	3150
	Toplam	2317 (%37,2)	3911 (%62,8)	6228

Tablo 12 ile tablo 10' karşılaştırdığımızda çalışmaya katılan erkeklerin sigara içme oranlarının kadınlara göre daha yüksek olduğu görülmektedir.

Çalışmaya katılan erkeklerin büyük çoğunluğunu 18-29 yaş aralığındaki erkekler oluşturmaktadır. 18-29 yaş aralığındaki erkeklerin %42,9'u sigara kullanırken %57,1'i sigara kullanmamaktadır. Erkeklerde sigara içme oranının en az olduğu yaşların 70 ve üzeri yaşlar olduğu görülmektedir.

Çalışmaya katılan erkeklerin eğitim durumları incelendiğinde sigara içme oranının en az %25,5 ile yüksek lisans doktora mezunlarında olduğu görülmektedir.

Kent ve kırlarda yaşayan erkeklerin 3 te 1'inden fazlasının sigara kullandığı görülmektedir. Yine aynı şekilde çalışan erkeklerin %44,2'sinin ve çalışmayan erkeklerin ise %30,3'ünün sigara kullandığını söyleyebiliriz.

Tablo 13: Diğer değişkenlere göre erkeklerin sigara içme oranı

		İçiyor	İçmiyor	Toplam
Solunum Yolları Hastalıkları (Astım-KOAH)	Var	59 (%33,5)	117 (%66,5)	176
	Yok	2258 (%37,3)	3794 (%62,7)	6052
Depresyon	Var	12 (%30,8)	27 (%69,2)	39
	Yok	2305 (%37,2)	3884 (%62,8)	6189
Kamu Spotu Etkisi	Düşünüyor	215 (%37,5)	359 (%62,5)	574
	Düşünmüyor	2102 (%37,2)	3552 (%62,8)	5654
Uyku Düzeni	Çok Düzensiz	106 (%47,5)	117 (%52,5)	223
	Düzensiz	366 (%43,2)	481 (%56,8)	847
	Bazen düzenli, bazen düzensiz	723 (%37,5)	1203 (%62,5)	1926
	Düzenli	1046 (%35,2)	1928 (%64,8)	2974
	Çok düzenli	76 (%29,5)	182 (%70,5)	258
Sigara Dışı Tütün Kullanımı	Kullanıyor	382 (%99,7)	1 (%0,3)	383
	Kullanmıyor	1935 (%33,1)	3910 (%66,9)	5845
	Toplam	2317 (%37,2)	3911 (%62,8)	6228

Tablo 13 incelendiğinde tanısı konulmuş solunum yolu hastalıklarına sahip olan erkeklerin %33,5'i sigara kullanırken %66,5'i sigara kullanmamaktadır.

Erkeklerin büyük bir çoğunluğu Sağlık Bakanlığının radyo ve televizyonlarda yer alan kamu spotu ve reklam kampanyalarının sigara kullanımı üzerinde etkisi olmadığını düşünmektedir. Kamu spotu reklamlarının sigara kullanımına etkisi olduğunu düşünen kadınların %62,5'inin sigara kullanmadığı görülmektedir. Uyku düzeni çok düzensiz olan erkeklerin %47,5'i sigara kullanırken %52,5'i sigara

kullanmamaktadır. Sigara dışı tütün ürünü kullanan erkeklerin %99,7'si sigara kullanmaktadır.

Tablo 14: Demografik verilere göre toplam sigara içme oranı

		İçiyor	**İçmiyor**	**Toplam**
Cinsiyet	Kadın	1387 (%17,4)	6582 (%82,6)	7969
	Erkek	2317 (%37,2)	3911 (%62,8)	6228
Yaş	18-29	970 (%29,4)	2332 (%70,6)	3302
	30-39	970 (%30,6)	2205 (%69,4)	3175
	40-49	804 (%31,5)	1748 (%68,5)	2552
	50-59	543 (%24,4)	1679 (%75,6)	2222
	60-69	284 (%15,7)	1522 (%84,3)	1806
	70 ve üstü	133 (%11,7)	1007 (%88,3)	1140
Medeni Durum	Bekar	1305 (%28,3)	3307 (%71,7)	4612
	Evli	2399 (%25,0)	7186 (%75,0)	9585
Eğitim Durumu	Okuma yazma bilmiyor/okuma yazma biliyor fakat bir okul bitirememiş	213 (%13,3)	1391 (%86,7)	1604
	İlkokul mezunu, ilköğretim, ortaokul veya mesleki ortaokul mezunu	1673 (%24,7)	5111 (%75,3)	6784
	Lise veya mesleki lise mezunu	1257 (%33,3)	2519 (%66,7)	3776
	Yüksekokul, fakülte mezunu	538 (%27,8)	1396 (%72,2)	1934
	Yüksek lisans, doktora mezunu	23 (%23,2)	76 (%76,8)	99
Yerleşim Birimi	Kent	3068 (%26,3)	8593 (%73,7)	11661
	Kır	636 (%25,1)	1900 (%74,9)	2536
Çalışma Durumu	Çalışıyor	1686 (%40,9)	2440 (%59,1)	4126
	Çalışmıyor	2018 (%20,0)	8053 (%80,0)	10071
	Toplam	3704 (%26,1)	10493 (%73,9)	14197

Tablo 14 incelendiğinde çalışmaya katılan kadınları %17,4'ü sigara kullanırken erkeklerin ise %37,2'si sigara kullanmaktadır.

Yaş gruplarına göre sigara kullanım oranlarına bakacak olursak en çok sigara kullanım oranlarının 18-29, 30-39 ve 40-49 yaşları arasında olduğu görülmektedir.

Tablo 15: Diğer değişkenlere göre toplam sigara içme oranı

		İçiyor	**İçmiyor**	**Toplam**
Solunum Yolları Hastalıkları (Astım-KOAH)	Var	124 (%24,5)	382 (%75,5)	506
	Yok	3580 (%26,1)	10111 (%73,9)	13691
Depresyon	Var	39 (%22,8)	132 (%77,2)	171
	Yok	3665 (%26,1)	10361 (%73,9)	14026
Kamu Spotu Etkisi	Düşünüyor	362 (%29,1)	884 (%70,9)	1246
	Düşünmüyor	3342 (%25,8)	9609 (%74,2)	12951
Uyku Düzeni	Çok Düzensiz	169 (%34,2)	325 (%65,8)	494
	Düzensiz	642 (%31,1)	1422 (%68,9)	2064
	Bazen düzenli, bazen düzensiz	1172 (%26,2)	3309 (%73,8)	4481
	Düzenli	1613 (%24,2)	5059 (%75,8)	6672
	Çok düzenli	108 (%22,2)	378 (%77,8)	486
Sigara Dışı Tütün Kullanımı	Kullanıyor	445 (%98,8)	1 (%0,02)	446
	Kullanmıyor	3259 (%23,7)	10492 (%76,3)	13751
	Toplam	3704 (%26,1)	10493 (%73,9)	14197

Tablo 15 incelendiğinde doktor tarafından tanısı konulmuş solum yolları hastalıklarına sahip olan kişilerin %24,5'i sigara kullanırken %75,5'i sigara kullanmamaktadır. Tanısı konulmuş depresyon hastalığı olan kişilerin ise %22,8'i sigara kullanmaktadır.

Sağlık Bakanlığının radyo ve televizyonlarda yer alan kamu spotu ve reklam kampanyalarının sigara kullanımı üzerinde etkisi olduğunu düşünen kişilerin %29,1'i sigara kullanırken %70,9'u sigara kullanmamaktadır.

4.2.3. Verilerin Normal Dağılıma Uygunluğunun Test Edilmesi

Kişilerin sigara kullanımına ilişkin tutumlarının belirlenmesi için uygulanan anket sonucu verilerinin öncelikle Normal dağılıma uyup uymadığı test edilmiştir. Normallik sınaması için Kolmogorov-Smirnov testi uygulanmıştır.

Tablo 16: Normallik Testi

Kolmogorov-Smirnov Z	0,078
p	0,000

Kolmogorov-Smirnov testi uygulamasına göre (sig.) p=0,000 < 0,05 olarak elde edilmiş, verilerin Normal dağılıma uygun olmadığı sonucuna ulaşılmıştır.

4.2.4. Basit Lojistik Regresyon Uygulama

Araştırma kapsamında ankete katılan kişilerin sigara kullanımına ilişkin tutumları etkileyen değişkenlerin tespit edilmesi amacıyla Lojistik regresyon analizi uygulanmıştır.

Bireylerin sigara kullanmaları ve kullanmamaları bağımlı değişken olarak, sosyo-demografik bilgiler ile diğer kategorik değişkenler ise bağımsız değişkenler olarak belirlenmiştir.

Tablo 17'de sigara kullanım modelinde kullanılan değişkenler, aldıkları değerler ve özelliklerini gösterilmektedir.

Tablo 17: Lojistik Regresyon Analizinde Kullanılan Değişkenler ve Özellikleri

Değişken	Tip	Aldığı Değerler
Durum	Kategorik	0 İçmiyor 1 İçiyor
Cinsiyet	Kategorik	1 Kadın 2 Erkek
Yaş	Kategorik	1 18-29 2 30-39 3 40-49 4 50-59 5 60-69 6 70 ve üstü
Medeni Durum	Kategorik	1 Bekar 2 Evli
Eğitim Durumu	Kategorik	1 Okuma yazma bilmiyor/okuma yazma biliyor fakat bir okul bitirememiş 2 İlkokul, ilköğretim, ortaokul 3 Lise veya mesleki lise 4 Yüksekokul, fakülte 5 Yüksek lisans, doktora
Yerleşim Birimi	Kategorik	1 Kent 2 Kır
Çalışma Durumu	Kategorik	0 Çalışmıyor 1 Çalışıyor
Solunum Yolları Hastalıkları	Kategorik	0 Yok 1 Var
Depresyon	Kategorik	0 Yok 1 Var
Kamu Spotu	Kategorik	0 Düşünmüyor 1 Düşünüyor
Uyku Düzeni	Kategorik	1 Çok düzensiz 2 Düzensiz 3 Bazen düzenli bazen düzensiz 4 Düzenli 5 Çok Düzenli
Sigara dışı tütün kullanımı	Kategorik	0 Kullanmıyor 1 Kullanıyor

Çok değişkenli lojistik regresyon modelini oluşturmak üzere, modele girecek değişkenleri belirlemek amacıyla, her bir değişkene ait tek değişkenli basit lojistik regresyon analizi sonuçları Tablo 18' de verilmiştir.

Tablo 18: Sigara Kullanımına İlişkin Basit Lojistik Regresyon Analizi

Değişken	β□	Std. Hata	Wald	sd	p	Exp (β□)
Sabit	-1,040	0,020	2774,900	1	0,000	0,353
Cinsiyet	1,034	0,039	684,842	1	0,000*	2,811
Sabit	-0,502	0,039	161,775	1	0,000	0,605
Yaş	-0,189	0,013	225,835	1	0,000*	0,827
Sabit	-1,013	0,020	2528,762	1	0,000	0,363
Medeni Durum	-0,167	0,040	17,216	1	0,000*	0,846
Sabit	-1,653	0,058	826,167	1	0,000	0,191
Eğitim Durumu	0,245	0,021	131,895	1	0,000*	1,278
Sabit	-1,062	0,025	1776,016	1	0,000	0,346
Yerleşim Birimi	-0,064	0,050	1,637	1	0,201	0,938
Sabit	-0,877	0,020	1895,041	1	0,000	0,416
Çalışma Durumu	1,014	0,040	634,016	1	0,000*	2,757
Sabit	-1,082	0,053	423,153	1	0,000	0,339
Solunum Yolları Hastalıkları	-0,087	0,105	0,682	1	0,409	0,917
Sabit	-1,129	0,092	151,868	1	0,000	0,323
Depresyon	-0,180	0,183	0,965	1	0,326	0,835
Sabit	-0,974	0,033	883,973	1	0,000	0,377
Kamu Spotu	0,163	0,066	6,207	1	0,013*	1,177
Sabit	2,464	0,501	24,230	1	0,000	11,757
Sigara Dışı Tütün Kullanımı	7,267	1,001	52,674	1	0,000*	1432,630
Sabit	-0,517	0,072	51,956	1	0,000	0,596
Uyku Düzeni	-0,159	0,021	56,310	1	0,000*	0,853

Tablo 18 incelendiğinde yerleşim birimi, solunum yolları hastalıkları ve depresyon değişkenlerinin anlamlı olmadığı ($p > 0,05$) olduğu görülmektedir.

4.2.5. Çoklu Lojistik Regresyon Uygulama

Tek değişkenli basit lojistik regresyon analizi sonunda önemli bulunan değişkenlerle oluşturulan Çok değişkenli lojistik regresyon modellerine ait sigara kullanımına ilişkin sonuçlar Tablo 19' da verilmiştir.

Tablo 19: Sigara Kullanımına İlişkin Çoklu Lojistik Regresyon Analizi

Değişken	β□	Std. Hata	Wald	sd	p	Exp (β□)
Sabit	2,159	0,506	18,234	1	0,000	8,659
Cinsiyet (1)	0,757	0,048	248,115	1	0,000	2,131
Yaş Grupları			154,868	5	0,000	
Yaş Grupları(1)	0,215	0,064	11,128	1	0,001	1,240
Yaş Grupları(2)	0,315	0,069	21,068	1	0,000	1,370
Yaş Grupları(3)	0,027	0,074	0,132	1	0,716	1,027
Yaş Grupları(4)	-0,486	0,088	30,412	1	0,000	0,615
Yaş Grupları(5)	-0,830	0,115	52,362	1	0,000	0,436
Medeni Durum(1)	-0,050	0,050	1,002	1	0,317	0,951
Eğitim Durumu			46,383	4	0,000	
Eğitim Durumu(1)	0,362	0,087	17,316	1	0,000	1,436
Eğitim Durumu(2)	0,439	0,094	21,720	1	0,000	1,551
Eğitim Durumu(3)	0,113	0,104	1,195	1	0,274	1,120
Eğitim Durumu(4)	-0,325	0,283	1,323	1	0,250	0,722
Çalışma Durumu(1)	0,495	0,053	88,616	1	0,000	1,641
Kamu Spotu(1)	0,053	0,072	0,553	1	0,457	1,055
Sigara Dışı Tütün kullanımı(1)	7,015	1,002	49,011	1	0,000	1113,314
Uyku Düzeni			55,727	4	0,000	
Uyku Düzeni(1)	-0,066	0,118	0,309	1	0,578	0,936
Uyku Düzeni(2)	-0,326	0,112	8,414	1	0,004	0,722
Uyku Düzeni(3)	-0,434	0,111	15,373	1	0,000	0,648
Uyku Düzeni(4)	-0,717	0,162	19,675	1	0,000	0,488

Tablo 19’ da, modelde yer alan bağımsız değişkenlerin katsayıları, bu katsayılara ait standart hatalar (SE), Wald istatistikleri, Wald istatistiklerinin p değerleri, serbestlik dereceleri, katsayıların anlamlılık düzeyleri, Odds Oranları

(Exp($\hat{\beta}$)) verilmiştir. Bağımlı değişken iki sonuçlu olduğu için ikili lojistik regresyon analizi uygulanmıştır.

Bu çalışmada kategorik bağımsız değişkenler için referans kategori olarak her değişkenin ilk kategorisi bağımlı değişken için ise içiyor (1) ele alınmıştır.

Tablo 19 incelendiğinde sabit katsayı, cinsiyet, yaş grupları, eğitim durumu, çalışma durumu, sigara dışın tütün kullanımı, uyku düzeni değişkenlerinin anlamlı olduğu ($p < 0,05$) medeni durum ve kamu spotu değişkelerinin ise anlamlı olmadığı ($p > 0,05$) görülmektedir. Bir değişkenin her kategorisi anlamlı olmayabilir, sadece tek bir kategorisi anlamlı olduğu için modele katkı sağlamış olarak kabul edilmektedir.

Erkeklerin referans kategori olan kadınlara göre sigara içme oranı 2,131 kat daha fazladır.

30-39, 40-49 ve 50-59 yaş grubundaki kişilerin referans grubu olan 18-29 yaş grubundaki kişiler göre sigara içme oranları sırasıyla 1,240, 1,370 ve 1,027 kat fazla iken, 60-69 ve 70 yaş üstü yaş grubundaki kişilerin referans grubundaki kişilere göre sigara içme oranları 0,615 ve 0,436 daha kat azdır.

Eğitim durumu ilkokul/ilköğretim/ortaokul olan kişilerin referans grubu olan okuma yazma bilmiyor/okul bitirmemiş kişilere göre sigara içme oranı 1,436 kat fazladır. Yine eğitim durumu lise/mesleki lise ve yüksekokul/fakülte olan kişilerin referans grubu olan okuma yazma bilmiyor/okul bitirmemiş kişilere göre sigara içme oranı sırasıyla 1,551 ve 1,120 kat fazladır. Eğitim durumu yüksek lisans/doktora olan kişilerin ise okuma yazma bilmiyor/okul bitirmemiş kişilere göre sigara içme oranı 0,722 kat daha azdır.

Çalışan kişilerin referans grubu olan çalışmayan kişilere göre sigara içme oranı 1,641 kat daha fazladır.

Sigara dışı tütün ürünü kullanan kişilerin referans grubu olan sigara dışı tütün kullanmayan kişilere göre sigara kullanımı 1113,314 kat daha fazladır.

Uyku düzeni düzensiz, bazen düzenli bazen düzensiz, düzenli ve çok düzenli olan kişilerin referans grubu olan uyku düzeni çok düzensiz olan kişilere göre sigara içme oranları sırasıyla 0,936, 0,722, 0,648 ve 0,488 kat daha azdır.

4.2.6.Modelin Uyum İyiliği

Ki-kare anlamlılık düzeyi model, blok ve adım için hesaplanır. Enter yönteminde bağımsız değişkenler hep birden modele dahil olduğundan model katsayıları testinin her aşamasında eşittir.

Tablo 20: Model katsayılarının Omnibus testi

		Ki - Kare	Serbestlik Derecesi	P
Adım 1	Adım	2292,349	18	0,000
	Blok	2292,349	18	0,000
	Model	2292,349	18	0,000

Tablo 21 incelendiğinde ki-kare anlamlılık düzeyi 0,00 < 0,05 olduğu için anlamlıdır. Elde edilen sonuçlara göre başlangıçtaki model ile bağımsız değişkenler eklendikten sonraki iki modelin -2LogL istatistikleri arasındaki fark olan G istatistiği, 18 serbestlik derecesiyle 2292,34'tür. Bu istatistik değeri $\chi 2$ (0,05; 18) = 28,869 değerinden büyük olduğundan dolayı H_0 hipotezi ($H_0=\beta_0=\beta_1=\ldots.=\beta_n=0$) reddedilmiştir. Lojistik regresyon katsayılarının hepsi aynı anda sıfıra eşit olmayıp, en azından birisi sıfırdan farklıdır.

Tablo 21: Model Özeti

Adım	-2 Log likelihood	Cox & Snell R^2	Nagelkerke R^2
1	14005,705[a]	0,149	0,218

Lojistik regresyon model tablosu incelendiğinde bağımsız değişkenlerin, bağımlı değişkeni açıklama oranının Nagelkerke R^2 değerine göre % 21 olduğu görülmektedir.

Tablo 22: Hosmer- Lemeshow Testi

Adım	Ki-kare	Serbestlik derecesi	P
1	14,484	8	0,070

Tablo 22 incelendiğinde, Hosmer-Lemeshow testinin sonucu anlamlı değildir (0,07>0,05). Bu değerin anlamlı çıkmaması, modelin kabul edilebilir bir uyuma sahip olduğunu, model ve veri uyumunun yeterli düzeyde olduğunu belirtir. Diğer bir deyişle gözlenen değerler ve model tarafından tahmin edilen değerler arasında anlamlı bir fark yoktur.

Tablo 23: Lojistik Regresyon Analizi Sonucu Elde Edilen Sınıflandırma Tablosu

	Gözlenen		Tahmin		
			Sigara		Doğru Yüzdesi
			İçmiyor	İçiyor	
Adım 1	Sigara	İçmiyor	10372	121	98,8
		İçiyor	3138	566	15,3
	Genel Yüzde				77,0

Lojistik regresyon analizi sonucunda elde edilen sınıflandırmaya göre, bağımsız değişkenlerin modele eklenmesiyle, sigara kullanmayan grupta yer alan 10493 kişinin 10372'si doğru iken 121'i yanlış sınıflandırılmıştır ve doğru sınıflandırılma oranı %98,8'dir. Sigara kullanan grubunda bulunan 3704 kişiden 566'sı doğru iken 3138'i yanlış sınıflandırılmıştır ve doğru sınıflandırılma oranı %15,3'tür.Elde edilmek istenen yani amaçlanan modele ilişkin toplam doğru sınıflandırılma oranı, doğru atanma oranını göstermektedir ve %50 değerinden büyük olması gerekir. Elde ettiğimiz modelin sınıflandırma tablosunun değeri %77,0 olup, modelde doğru bir atanma yapıldığını göstermektedir.

SONUÇ VE ÖNERİLER

Bu çalışmada kategorik ile lojistik regresyon analizlerden ve yöntemlerinden bahsedilmesinin ardından basit ve çok değişkenli lojistik regresyon analizleri ile ilgili uygulama yapılmıştır.

Bireylerin sigara içmelerini etkileyen faktörleri belirlemek ve bu faktörlerin her birinin sigara içme üzerindeki etkisini ortaya koyma amacıyla yapılan bu çalışmada Türkiye'nin 79 ilinden mahalleler seçilmiş ve bu mahallelerde oturan kişilere anket uygulanmıştır. Elde edilen veriler SPSS 25'te analiz edilmiştir. Sigara kullanımı üzerinde etkisi olabileceği düşünülen 6'sı demografik bilgiler içeren 11 bağımsız değişken belirlenmiştir. Belirlenen kategorik değişkenler ile araştırmaya katılan 14197 kişiye ait sosyo-demografik bilgilerine ve özelliklerine ilişkin çeşitli frekans tabloları Tablo 7 ve Tablo 8'de verilmiştir. Daha sonra 14197 kişiye ait sigara içme oranı hesaplanmış kadınlara, erkeklere ve toplam örnekleme göre sigara içme oranlarına ait çapraz tablolar Tablo 10, Tablo 11, Tablo 12, Tablo 13, Tablo 14 ve Tablo 15'de verilmiştir. Erkeklerin kadılara göre sigara içme oranlarının daha yüksek olduğu görülmüştür.

Bağımlı değişkenimiz kişilerin sigara kullanıp kullanmaması olup, iki şıklı kategorik bir değişkendir. Bu nedenle, çalışmada kişilerin sigara kullanımını etkileyen faktörleri belirlemek amacıyla, normal dağılım gerektirmeyen ve diğerlerine kıyasla daha esnek birçok değişkenli analiz tekniği olan iki şıklı (binary) lojistik regresyon analizinden yararlanılmıştır. Analizde lojistik regresyon yöntemlerinden enter yöntemi uygulanmıştır. Öncelikle belirlenen 11 değişken için her bir değişkene ait tek değişkenli ikili basit lojistik regresyon uygulaması yapılmıştır. Uygulama sonucundan 11 değişkenden yerleşim birimi, solunum yolları hastalıkları ve depresyon değişkenlerinin anlamlı olmadığı ($p > 0,05$) görülmüştür. Cinsiyet, yaş, medeni durum, eğitim durumu, çalışma durumu, kamu spotu, sigara dışı tütün kullanımı ve uyku düzeni değişkenlerinin anlamlı olduğu ($p < 0,05$) Tablo 19'da gösterilmiştir.

Sonrasında tek değişkenli basit lojistik regresyon analizi sonunda önemli bulunan 8 değişkenlerle çok değişkenli lojistik uygulaması yapılarak bireylerin sigara kullanımını etkileyen faktörler belirlenmiştir. Elde edilen lojistik regresyon modeli

sonuçlarına göre cinsiyet, yaş, eğitim durumu, çalışma durumu, sigara dışı ürün kullanma durumu ve uyku düzeni kişilerin sigara kullanımı üzerinde etkili olduğu belirlenmiştir. Odds oranlarına göre, kişilerin sigara içmelerindeki en önemli faktörün sigara dışı tütün ürünü kullanımı olduğu görülmüştür. Bu katsayı sigara dışı tütün ürünü kullanan kişilerin sigara dışı tütün kullanmayan kişilere göre sigara kullanımı 1113,314 kat daha yüksek olduğunu gösterir. Medeni durum ve kamu spotu değişkenlerinin ise sigara kullanımı üzerinde etkili olmadığı tespit edilmiştir.

Yapılan analizler sonunda lojistik regresyona göre doğru sınıflandırma oranı %77 olarak bulunmuştur. Bu oran %50 'den büyük olup modelde doğru bir atanma yapıldığını göstermektedir. Ayrıca Hosmer-Lemeshow testi sonucuna göre modelin kabul edilebilir bir uyuma sahip olduğu, model ve veri uyumunun yeterli düzeyde olduğu görülmüştür.

Bu çalışma sırasında karşılaşılan yeni problemlere ve bakış açılarına dayalı olarak aşağıda belirttiğim başlıklar bundan sonraki araştırmacılara öneri olarak listelenmiştir;

- Kişilerin sigara kullanımına etkili olan faktörleri ve bu faktörlerin etkilerini belirlemek amacıyla belirlenen değişken sayısı arttırılarak daha geniş kapsamlı bir araştırma yapılabilir.
- Bu çalışmada kişilerin sigara kullanımını etkileyen faktörleri belirlemek amacıyla standart (enter) yöntemi kullanılmıştır. Çalışma adımsal (stepwise) yöntemiyle de yapılıp sonuçlar karşılaştırılabilir.

KAYNAKÇA

Afifi, A.A., Clark,V. (1997). *Computer - Aided Multivariate Analysis.* New York: Chapman & Hall.

Agresti, A. (1996). *Categorical Data Analysis*. Canada: John Wiley & Sons Publication.

Akçagöz H, Özkan B, Kızılay H, (2005). Tarımsal Üretimde Çiftçi Davranışları: Çiftçiliği Uygulama Ölçeği (FIS). *Bahçe Dergisi.*34(2): 63-71.

Aksaraylı, M., Saygın, Ö. (2011). Algılanan Hizmet Kalitesi ve Lojistik Regresyon Analizi ile Hizmet Tercihine Etkisinin Belirlenmesi. *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*. 13(1): 21-37.

Aktaş, R. (1997). *Mali Başarısızlık (İşletme Riski) Tahmin Modelleri.* Ankara: Türkiye İş Bankası Kültür Yayınlar.

Albayrak, A.S., Eroğlu, A., Kalaycı, Ş., Küçüksille,E., AK, B., Karaatlı, M., Keskin, H.Ü., Çiçek, E.U., Kayış,A., Antalyalı, Ö.L., Uçar, N., Demirgil, H., İşler,D. B., Süngür, O., (2014). *SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri.* Ankara: Asil Yayın Dağıtım Ltd. Şti.

Alberg, A.J., Samet, J.M. (2003). Epidemiology of Lung Cancer. *Chest Journal* 123(1): 1.

Alpar, R. (2013). *Uygulamalı Çok Değişkenli İstatistiksel Yöntemler.* Ankara: Detay Yayıncılık.

Altaş, D., Yıldırım E.Z. (2003). Lisansüstü Eğitime GirişSınavı (LES) Sonuçlarının Üç Yönlü Çapraz Sınıflandırma Tablosu ile İncelenmesi. *Marmara Üni. Sos. Bil. Enst. Hakemli Dergisi.* 5(20): 213-223.

Aytemur, Z.A., Akc□ay, Ş., Elbek, O. (2010). *Tu□tu□n ve Kontrolu□*. I□stanbul: Tu□rk Toraks Derneg□i Yayınları.

Baloğlu, S. ve Mccleary, K.W. (1999) U.S. International Pleasure Travelers' Images of Four Mediterranean Destinations: A Comparison of Visitors and Nonvisitors. *Journal of Travel Research*. 38(2): 144-152.

Baydemir, M. B., (2014). *Lojistik Regresyon Analizi Üzerine Bir İnceleme.* (Yayınlanmamış Yüksek Lisans Tezi). Malatya: İnönü Üniversitesi Sosyal Bilimler Enstitüsü.

Bergh, H., Baigi, A., Mansson, J., Mattsson, B., Marklund, B. (2007). Predictive factors for long-term sick leave and disability pension among frequent and normal attenders in primary health care over 5 years. *Public Health.* 121(1): 25–33.

Büyüköztürk, Ş., Çokluk, Ö., Şekercioğlu, G. (2014). *Sosyal Bilimler İçin Çok Değişkenli İstatistik SPSS ve LISREL Uygulamaları.* Ankara: Pegem Akademi.

Cankurt, M., Miran , B. (2010). Aydın Yöresinde Çiftçilerin Traktör Satın Alma Eğilimleri Üzerine Bir Araştırma. *Ege Üniversitesi Ziraat Fakültesi Dergisi.* 47(1): 43-51.

Chen, T. (1999). Critical Success Factors for Various Strategies in the Banking Industry. *International Journal of Bank Marketing.* 17(2): 83-91.

Chatfield C. Collins, A.J. (1980). *Introduction to Multivariate Analysis*. London: Cahpman and Hall

Choi, T.Y. ve Chu, R. (2000). Levels of Satisfaction among Asian and Western Travellers. *International Journal of Quality & Reliability Management.* 17(2): 116-131.

Dubey, S., Charles, A. (2008). Update in Lung Cancer 2007. *American Journal of Respiratory and Critical Care Medicine.* 177(1): 941-46.

Freeman, D. H. (1987). *Applied Categorical Data Analysis.* New York: Marcel Dekker Inc.

Gatty, R. (1966). Multivariate Analysis for Marketing Research. *An Evaluation, Applied Statistics.* 15(3): 157-172.

Gauch, H. G., JR. (1982). *Multivariate Analysis in Community Ecology*. New York: Cambridge University Press.

Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C. (1998). *Multivaritate Data Analysis*. New Jersey: Prentice-Hall International Inc.

Hamarat, B. (1998). *Türkiye' de Sağlık Açısından Homojen İl Gruplarının Belirlenmesine İlişkin İstatistiksel Bir Yaklaşım.* (Yayınlanmamış Yüksek Lisans Tezi). Eskişehir: Anadolu Üniversitesi Fen Bilimleri Enstitüsü.

Harris, R.J. (2001). *A Primer of Multivariate Statistics*. London: Lawrence Erlbaum Associates.

Hosmer, D.W., Lemeshow S. (2000). *Applied Logistic Regression*. New York: John Wiley & Sons Inc

Huang, X., Soutar, G.N. ve Brown, A. (2001). Resource Adequacy in New Product Development: A Discriminant Analysis. *European Journal of Innovation Management.* 4(1): 53-59

İyit, N. (2003). *Lineer Olmayan Lojistik Regresyon Analizinde Model Kurma Stratejileri ve Bir Uygulaması.* (Yayınlanmamış Yüksek Lisans Tezi). Konya: Selçuk Üniversitesi

Jobson, J.D. (1992). A*pplied Multivariate Data Analysis*. New York: Springer Science+ Business Media.

Johnson, D. E. (1988). *Applied Multivariate Methods for Data Analysts.* California: Duxbury Press.

Johnson, R. A.,Wichern D.W. (2002). *Applied Multivariate Statistical Analysis*, New Jersey: Prentice-Hall, Inc.

Kachigan, S.K. (1991). *Multivariate Statistical Analysis, A Conceptual Introduction.* New York: Radius Press.

Karlıkaya, C., Öztuna, F., Solak, Z., Özkan, M., Örsel, O. (2006).Tu□tu□n Kontrolu□. *Toraks Dergisi.* 7(1): 51-64.

Legendre, L., Legendre,P. (1983). *Numerical Ecology*. Amsterdam: Elsevier

Menard, S. (2002). *Applied Logistics Regression Analysis.* London: Sage Publications.

Ohlson, J.A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research.* 18(1): 109-131.

Orloci, L. (1978). *Multivariate Analysis in Vegetation Research.* Boston W. Junk.

Özcebe, H. (2008). Genc□ler ve Sigara. *Sag□lık Bakanlıg□ı Yayını.* 731(1): 9-10.

Özdamar, K. (1999). *Paket Programlar Ile Istatistiksel Veri Analizi-2 (Çok Değişkenli Analiz)*. Eskişehir: Kaan Kitabevi.

Özkul, I□., Sarı, Y. (2008). Tu□rkiye' de Tu□tu□n Sekto□ru□nu□n Durumu, Sorunları ve Ço□zu□m Önerileri. *2. Ulusal I□ktisat Kongresi.* Düzenleyen Dokuz Eylu□l Üniversitesi I□ktisadi ve I□dari Bilimler Faku□ltesi. I□zmir. 20-22 Şubat 2008.

Özlu□, T., Metintas□, M., Karadag□, M., Kaya, A. (2010). *Solunum Sistemi ve Hastalıkları Temel Bas□vuru Kitabı.* İstanbul: I□stanbul Tıp Kitabevi.

Özyardımcı, N. (2002), *Sigara ve Sag□lık*. Bursa: Nadir Kitap

Pielou, E.C. (1984). *The Interpretation of Ecological Data*. New York: Wiley-Interscience.

Sağlam, M. (2013). Çok Değişkenli İstatistiksel Yöntemler ile Toprak Özelliklerinin Gruplandırılması. *Toprak Su Dergisi.* 2(1): 7-14.

Shaw, P. J. A. (2003). *Multivariate Statistics for the Environmental Sciences.* New York: Hodder Arnold.

Sheth, J.N. (1971). The Multivariate Revolution in Marketing Research. *Journal of Marketing,* 35(1): 13-19.

Tatlıdil, H.(2002). *Uygulamalı Çok Değişkenli İstatistiksel Analiz.* Ankara: Akademi Matbaası.

Temple, J. (2004). Explaining the private health insurance coverage for older Australians. *People and Place.* 12 (2): 13-23

Timm , N.H. (2002). *Applied Multivariate Analysis.* New York. Springer -Verlag,

Türkiye İstatistik Kurumu (TÜİK). Küresel yetişkin tütün araştırması, 2008-2012

Walsh, G., Thurau, T.H., Mitchell, V.W., Wiedmann, K.P. (2001) Consumers' Decision Making Style as a Basis for Market Segmentation. *Journal of Targeting, Measurement and Analysis for Marketing.* 10(2): 117-131.

Whittaker, R.H. (1978). *Classification of Plant Communities.* The Hague: Junk

Whittaker, R.H. (1978). *Ordination of Plant Communities*. The Hague: Junk

WHO. (2008). World Health Organization. WHO Report On The Global Tobacco Epidemic, The MPOWER Package. Geneva, 2008;48.

WHO. (2015). WHO: Tobacco.

Williams, W.T. (1976). *Pattern Analysis in Agricultural Science*, New York: Elsevier.

Withey, J.J. ve Panitz E. (1995). Face-to-Face Selling: Making It More Effective. *Industrial Marketing Management.* 24(1): 239-246.