

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

**KATEGORİK VERİ ANALİZİNİN İSTATİSTİKSEL
VERİ ANALİZİ İÇERİSİNDEKİ YERİ VE ÖNEMİ**

Ayşem Ece YALÇINKAYA

Danışman

Yrd.Doç.Dr. Ali Kemal ŞEHİRLİOĞLU

2008

Yemin Metni

Yüksek Lisans Tezi olarak sunduğum “**Kategorik Veri Analizinin İstatistiksel Veri Analizi İçerisindeki Yeri ve Önemi**” adlı çalışmanın, tarafımdan, bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

Tarih

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

Ayşem Ece

YALÇINKAYA

YÜKSEK LİSANS TEZ SINAV TUTANAĞI

Öğrencinin

Adı ve Soyadı :Ayşem Ece YALÇINKAYA
Anabilim Dalı :Ekonometri
Programı :Ekonometri
Tez Konusu :Kategorik Veri Analizinin İstatistiksel Veri Analizi İçerisindeki Yeri ve Önemi
Sınav Tarihi ve Saati :....../....../....

Yukarıda kimlik bilgileri belirtilen öğrenci Sosyal Bilimler Enstitüsü'nün tarih ve sayılı toplantısında oluşturulan jürimiz tarafından Lisansüstü Yönetmeliği'nin 18. maddesi gereğince yüksek lisans tez sınavına alınmıştır.

Adayın kişisel çalışmaya dayanan tezini dakikalık süre içinde savunmasından sonra jüri üyelerince gerek tez konusu gerekse tezin dayanağı olan Anabilim dallarından sorulan sorulara verdiği cevaplar değerlendirilerek tezin,

BAŞARILI OLDUĞUNA	O	OY BİRLİĞİ	O
DÜZELTİLMESİNE	O*	OY ÇOKLUĞU	O
REDDİNE	O**		

İle karar verilmiştir.

Jüri teşkil edilmediği için sınav yapılamamıştır.

O***

Öğrenci sınava gelmemiştir.

O**

* Bu halde adaya 3 ay süre verilir.

** Bu halde adayın kaydı silinir.

*** Bu halde sınav için yeni bir tarih belirlenir.

Evet

Tez burs, ödül veya teşvik programlarına (Tüba, Fulbright vb.) aday olabilir. O

Tez mevcut hali ile basılabilir. O

Tez gözden geçirildikten sonra basılabilir. O

Tezin basımı gerekliliği yoktur. O

JÜRİ ÜYELERİ

İMZA

.....	☐ Başarılı	☐ Düzeltme	☐ Red	
.....	☐ Başarılı	☐ Düzeltme	☐ Red	
.....	☐ Başarılı	☐ Düzeltme	☐ Red	

ÖZET

Yüksek Lisans Tezi

Kategorik Veri Analizinin İstatistiksel Veri Analizi İçerisindeki Yeri ve Önemi

Ayşem Ece YALÇINKAYA

Dokuz Eylül Üniversitesi

Sosyal Bilimler Enstitüsü

Ekonometri Anabilim Dalı

İstatistiksel analiz yöntemleri verilerin ifade ettiği gerçekleri göstermenin bir yolu olarak yıllarca birçok alanda kullanılmıştır. Dünyanın var olan dengesini ispatlamak için fizikçiler, insan davranışlarını ifade edebilmek için davranışçılar, pazarlama dengelerini ve piyasa şartlarını araştırmak için işletmeciler, hastalıkları ve hasta durumlarını incelemek için tıpçılar vb.

Bunlar ve daha birçok alanda 1960'lı yıllara kadar sadece normal dağılıma dayalı analiz yöntemleri üzerinden analizler gerçekleştirilmiştir. Oysaki kullandığımız verilerin büyük bir oranı da kategorik özellik taşımaktadır. Kategorik verilerin dağılımları normal dağılıma uymayacağı gibi, analizlerinde de normal dağılım analizleri kullanılamaz.

Bu alanda en önemli çalışmalar 1990'larda başlamıştır. Bu dönemde birçok yeni analiz yöntemi ve istatistiki prensipler ortaya atılmıştır. Günümüze kadar gelen çalışmalar neticesinde, geç gelişen kategorik veri analiz yöntemleri büyük bir aşama kaydetmiştir.

Türkiye'de sadece kategorik veri analiz yöntemleri ile ilişkin spesifik bir çalışma bulunmamaktadır. Bu alandaki açığın ortadan kaldırılması amacıyla bu çalışma yapılmıştır. Çalışmada kategorik veri analiz yöntemleri ayrıntılandırılmış, uygulamalı gösterimi ise mağazacılık sektöründe bir anket uygulaması ile yapılmıştır. Uygulmanın ikincil hedefi olarak da ordinal veriler üzerinde nominal analiz yöntemlerinin uygulanabileceği ispat edilmiştir.

Anahtar Kelimeler: Kategorik Veri, Loglineer Analiz Yöntemi, Sıralı Lojistik Analiz Yöntemi, Multinomial Regresyon Analiz Yöntemi, Rekabet.

ABSTRACT

Master Thesis

The Place and The Importance of Categorical Data Analysis Methods in Statistical Data Analysis

Ayşem Ece YALÇINKAYA

Dokuz Eylül University

Institute of Social Sciences

Department Econometrics

For years, in many areas the statistical analysis methods is used as a way to show what the statistical datas say actually. Physicists to prove the earth's equilibrium, behaviourists to express the human behaviours, administrators to research the market conditions and marketing rules, doctors to examine the illnesses and patients' situations etc.

In these and much more areas, the analysis had been making by normal distribution and its analysis methods until the 1960's. But, the huge proportion of the datas we used is in categorical type. As the distribution categorical datas do not conform with normal distribution, normal distribution analysis can not be used for them, either.

In this area, the most important studies was began in 1990's. In that period of time many statistical methods and statistical principles were invented. Until now as a result of the studies in statistic, it was made progress in categorical datas analysis methods.

There is not a specific study about categorical data anylysis methods. This study's goal is to cover this hole. In the reasearch, categorical data analysis methods were shown in detail and its practice was made in the shopping sector by a questionnaire. The second goal of the research is to determine if nominal analysis methods can be used on ordinal data sor not.

Key Words: Categorical Data, Loglinear Analysis Methods, Ordinal Logit Analysis Methods, Multinomial Regression Analysis Methods, Competition.

KATEGORİK VERİ ANALİZİNİN İSTATİSTİKSEL VERİ ANALİZİ İÇERİSİNDEKİ YERİ VE ÖNEMİ

YEMİN METNİ.....	ii
TUTANAK.....	iii
ÖZET.....	iv
ABSTRACT.....	v
İÇİNDEKİLER.....	vi
KISALTMALAR.....	xiii
TABLolar LİSTESİ.....	xiv
ŞEKİLLER LİSTESİ.....	xvii
GİRİŞ.....	18

BİRİNCİ BÖLÜM

KATEGORİK VERİLER İÇİN DAĞILIMLAR VE YORUMLAMALAR

1.1. KATEGORİK YANIT VERİLERİ -----	HATA! YER İŞARETİ
TANIMLANMAMIŞ.	
___ 1.1.1. Yanıt – Açıklayıcı Değişken Ayrımı -----	6
___ 1.1.2. Nominal – Ordinal Ölçek Ayrımı -----	7
___ 1.1.3. Sürekli – Kesikli Değişken Ayrımı -----	HATA! YER İŞARETİ
TANIMLANMAMIŞ.	
___ 1.1.4. Kalitatif – Kantitatif Değişken Ayrımı -----	9
1.2. KATEGORİK VERİLER İÇİN DAĞILIMLAR -----	10
___ 1.2.1. Binom Dağılımı -----	10
___ 1.2.2. Çokterimli Dağılım -----	11
___ 1.2.3. Poisson Dağılımı -----	12
___ 1.2.4. Aşırı Yayılım -----	13
___ 1.2.5. Çokterimli Ve Poisson Dağılımları Arasındaki İlişki -----	15
1.3. KATEGORİK VERİLER İÇİN İSTATİSTİKSEL YORUMLAMA -----	16
___ 1.3.1. Olabilirlik Fonksiyonları Ve En Çok Olabilirlik Tahmini -----	16
___ 1.3.2. Binom Parametresi İçin Olabilirlik Fonksiyonu Ve Ml Tahmini -----	18

1.3.3. Wald - Olabilirlik Oran - Skor Testi Üçlüsü -----	19
1.3.4. Güven Aralıklarının Oluşturulması-----	24
1.4. BİNOM PARAMETRELERİ İÇİN İSTATİSTİKSEL YORUMLAMA -----	25
1.4.1. Bir Binom Parametresi İle İlgili Testler -----	25
1.4.2. Binom Parametresi İçin Güven Aralıkları -----	27
1.4.3. Vejetaryenlerin Oranı İle İlgili Örnek-----	28
1.4.4. Kesin Olasılık Küçük Örneklem Yorumlaması-----	29
1.4.5. Orta P-Değerine Dayanan Yorumlama -----	32
1.5. ÇOK TERİMLİ PARAMETRELER İÇİN İSTATİSTİKSEL YORUMLAMA	34
1.5.1. Belirli Bir Çokterimli İçin Pearson İstatistiği-----	35
1.5.2. Örnek: Mendel' İn Teorileri-----	36
1.5.3. Ki-Kare Teorik Açıklaması-----	37
1.5.4. Ki-Kare Olabilirlik Oranı -----	39
1.5.5. Tahminlenen Ve Beklenen Frekanslarla Birlikte Test Etme-----	40

İKİNCİ BÖLÜM

OLUMSALLIK (KONTENJANS) TABLOLARI

2.1. OLUMSALLIK TABLOLARI İÇİN OLASILIKLARIN YAPISI -----	45
2.1.1. Duyarlılık Ve Seçicilik-----	49
2.1.2. Kategorik Değişkenlerin Bağımsızlığı -----	50
2.1.3. Poisson, Binom Ve Çokterimli Örnekleme -----	52
2.1.4. İnceleme Türleri-----	56
2.2. İKİ ORANIN KARŞILAŞTIRILMASI -----	59
2.2.1. Oranların Farkı -----	59
2.2.2. Göreli Risk (Relative Risk, Oransal Risk)-----	61
2.2.3. Odds Oranı -----	63
2.2.4. Odds Oranının Özellikleri-----	65
2.2.5. Olgu-Kontrol İncelemeleri Ve Odds Oranı-----	67
2.2.6. Odds Oranı Ve Göreli Risk Arasındaki İlişki -----	68
2.2.7. Odds Oranı İle Logaritmik Odds Oranına Göre Yorumlama -----	70

2.3. TABAKALI 2X2 BOYUTLU OLUMSALLIK TABLOLARINDA KISMİ BİRLİKTELİK -----	72
2.3.1. Kısmi Tablolar-----	73
2.3.2. Ölüm Cezası Örneği -----	74
2.3.3. Koşullu Ve Marjinal Odds Oranları -----	77
2.3.4. Koşullu Bağımsızlığa Marjinal Bağımsızlık -----	78
2.3.5. Homojen Birliktelik-----	81
2.3.6. Cochran-Mantel-Haenszel Metotları -----	83
2.3.7. Akciğer Kanserinin Meta Analizi Örneği-----	87
2.3.8. Ortak Odds Oranının Hesaplanması -----	88
2.3.9. Odds Oranlarının Homojenliğinin Testi -----	89
2.3.10. Koşullu Birliktelikler Konusunda Kesin (Exact) Olasılık Yorumlamaları-----	92
2.4. I X J TABLOLARI İÇİN İLAVELER -----	96
2.4.1. I X J Tablolarındaki Odds Oranları -----	96
2.4.2. Birlikteliğin Özet Ölçümleri-----	98
2.4.3. Ordinal Eğilimler: Uyumlu Ve Uyumsuz İkiliiler -----	99
2.4.4. Birlikteliğin Ordinal Ölçümü: Gamma -----	101
2.5. Kİ-KARE BAĞIMSIZLIK TESTLERİ -----	103
2.5.1. Pearson İstatistiği Ve Ki-Kare Dağılımı -----	103
2.5.2. Olabilirlik-Oran İstatistiği-----	104
2.5.3. Bağımsızlık Testleri-----	106
2.5.4. Artıklar-----	107
2.5.5. Ki-Kare Parçalanımı -----	108
2.5.6. Ki-Kare Testlerinin Yorumlanması-----	110
2.6. ORDİNAL VERİLERDE BAĞIMSIZLIĞIN TEST EDİLMESİ-----	111
2.6.1. Bağımsızlık İçin Doğrusal Eğilim (Trend) Alternatifi -----	111
2.6.2. Alkol Ve Sakat Bebek Örneği-----	113
2.6.3. Ordinal Test İle Ekstra Güç Sağlanması -----	115
2.6.4. Skorların Seçimi -----	116
2.6.5. I X 2 Ve 2 X J Tabloları İçin Eğilim Testleri-----	118
2.6.6. Nominal-Ordinal Tablolar-----	120

2.7. KÜÇÜK ÖRNEKLER İÇİN KESİN YORUMLAMA -----	120
2.7.1. Fisher’ In Kesin (Exact) Olasılık Testi-----	120
2.7.2. Fisher’ın Çay Denemesi-----	122
2.7.3. P-Değerleri Ve Tip-1 Hata Olasılıkları -----	124
2.7.4. Odds Oranı İçin Küçük-Örnek Güven Aralıkları -----	126
2.7.5. Daha Büyük Tablolarda Bağımsızlığın Kesin Testleri-----	127
2.8. ÇOK YÖNLÜ TABLOLAR VE TABLOLAŞTIRILAMAYAN YANITLAR İÇİN İLAVELER -----	127

ÜÇÜNCÜ BÖLÜM

GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER

3.1. GENELLEŞTİRİLMİŞ LİNEER BİR MODELİN BİLEŞENLERİ -----	131
3.1.1. Rassal Bileşen -----	131
3.1.2. Sistematiik Bileşen -----	132
3.1.3. Bağıntı -----	132
3.1.4. Normal Glm-----	134
3.2. İKİLİ VERİLER İÇİN GENELLEŞTİRİLMİŞ LİNEER MODELLER -----	135
3.2.1. Doğrusal Olasılık Modeli -----	135
3.2.2. Lojistik Regresyon Modeli-----	HATA! YER İŞARETİ TANIMLANMAMIŞ.
3.2.3. Probit Modeller -----	141
3.3. SAYILABİLİR VERİLER İÇİN GENELLEŞTİRİLMİŞ LİNEER MODELLER: POISSON REGRESYON -----	142
3.3.1. Poisson Regresyon -----	144
3.3.2. Oran Verileri İçin Poisson Regresyonu-----	144
3.4. MODEL YORUMU VE MODEL KONTROLÜ -----	145
3.4.1. Wald, Olabilirlik Oranı (Likelihood-Ratio), Skor Testi -----	146
3.4.2. Poisson Model Kontrolü -----	148
3.4.3. Model Artıkları -----	148
3.4.4. Poisson Regresyonunda Aşırı Yayılım-----	150
3.5. GENELLEŞTİRİLMİŞ LİNEER MODELLERİN UYUMU -----	151
3.5.1. Newton-Raphson Algoritması-----	151

3.5.2. Kullanılan Olabilirlik Fonksiyonunun Yorumlanması -----	153
3.5.3. Sapma -----	154
3.5.4. Glm' Lerin Avantajları -----	156

DÖRDÜNCÜ BÖLÜM

LOJİSTİK REGRESYON

4.1. LOJİSTİK REGRESYON MODELİNİN YORUMLANMASI -----	157
4.1.1. Lineer Yaklaşım Yorumları-----	158
4.1.2. Odds Oranının Yorumlanması-----	159
4.1.3. Olgu-Kontrol İncelemeleri İçin Lojistik Regresyon -----	160
4.2. LOJİSTİK REGRESYON İÇİN YORUMLAMA -----	161
4.2.1. Etkiler İçin Güven Aralıkları -----	161
4.2.2. Anlamlılığın Test Edilmesi -----	162
4.2.3. Olasılık Tahminlerinin Dağılımları-----	163
4.3. MODEL KONTROLÜ -----	163
4.3.1. Uyum İyiği Ve Olabilirlik Oran Model Karşılaştırma Testleri -----	165
4.3.2. Lojit Modeller İçin Artıklar-----	166
4.3.3. Etkinin Teşhis Edilmesindeki Ölçüler-----	167
4.4. KALİTATİF TAHMİN EDİCİLER İÇİN LOJİT MODELLER -----	168
4.4.1. Lojit Modellerde Kukla Değişkenlerin Kullanımı-----	169
4.4.2. Azt Ve Aıds Örneği -----	170
4.4.3. Faktörlerin Anova Tipi Gösterimi -----	171
4.5. ÇOKLU LOJİSTİK REGRESYON -----	174
4.5.1. Birden Fazla Tahmin Ediciye Sahip Model Seçimi-----	175
4.5.2. Tahmin Edicilerin Geriye Doğru Elenmesi -----	176
4.5.3. Tahmin Edicilik Gücünün Korelasyon Özeti -----	177
4.6. LOJİSTİK REGRESYON İÇİN KESİN YORUMLAMA -----	178
4.6.1. Koşullu Maksimum Olabilirlik Yorumlaması -----	178
4.6.2. Olumsuzluk Tabloları İçin Kesin Yorumlama-----	179

BEŞİNCİ BÖLÜM

OLUMSALLIK TABLOLARI İÇİN LOG-LİNEER MODELLER

5.1. İKİ YÖNLÜ TABLOLAR İÇİN LOG LİNEER MODELLER -----	182
5.1.1. Bağımsız Model -----	182
5.1.2. Parametrelerin Yorumlanması -----	184
5.1.3. Parametre Kısıtları -----	186
5.1.4. Doygun Model -----	187
5.2. ÜÇ YÖNLÜ TABLOLAR İÇİN LOG LİNEER MODELLER -----	190
5.2.1. Log Linear Modeller Ve Bağımsızlık Yapısı -----	190
5.2.2. Model Parametrelerinin Yorumlanması -----	192
5.2.3. Alkol, Sigara Ve Marihuana Kullanımına İlişkin Örnek -----	193
5.3. LOG LİNEER MODELLER İÇİN YORUMLAMA -----	197
5.3.1. Log Linear Modellerin Uygunluğu -----	197
5.3.2. Ki-Kare İyi Uyum Testleri -----	198
5.3.3. Log Linear Artıkları -----	199
5.3.4. Kısmi Birliktelik Testleri -----	202
5.3.5. Odds Oranları İçin Güven Aralıkları -----	203
5.4. DAHA YÜKSEK BOYUTLAR İÇİN LOG LİNEER MODELLER -----	205
5.5. LOG LİNEER – LOJİT İLİŞKİSİ -----	206
5.5.1. Log Linear Modelleri Yorumlamak İçin Lojit Modellerin Kullanımı -----	206
5.5.2. Log Linear Modeller Ve Lojit Modeller Arasındaki Benzerlik -----	208
5.5.3. Model Seçimindeki Stratejiler -----	209

ALTINCI BÖLÜM

KATEGORİK VERİ ANALİZİ KONUSUNDA YİRMİNCİ YÜZYILDAKİ GELİŞMELER

6.1. PEARSON-YULE BİRLİKTELİK TARTIŞMASI -----	212
6.2. R.A. FISHER' IN KATKILARI -----	214
6.3. LOJİSTİK REGRESYON -----	216
6.4. ÇOK YÖNLÜ OLUMSALLIK TABLOLARI VE LOG LİNEER MODELLER -----	219
6.5. SON (VE GELECEKTEKİ) GELİŞMELER -----	222

YEDİNCİ BÖLÜM

MAĞAZACILIK SEKTÖRÜNDE BİR KATEGORİK VERİ ANALİZİ UYGULAMASI

7.1. ANKETLERİN GEÇERLİLİĞİ VE GÜVENİLİRLİĞİ	227
7.1.1. Soruların Zorluk Derecesi Testi	227
7.1.2. Anket Güvenirlilik Analizi :	227
7.1.3. Soru-Bütün Korelasyon Testi.....	228
7.2. ARAŞTIRMANIN EVRENİ VE ÖRNEKLEM.....	228
7.3. PROBLEM CÜMLESİ	229
7.4. VERİLERİ TOPLAMA VE ANALİZ YÖNTEMİ.....	229
7.5. ARAŞTIRMA SINIRLILIKLARI	230
7.6. ARAŞTIRMA VARSAYIMLARI	230
7.7. VERİLERİN ANALİZİ VE BULGULAR	230
7.7.1.Kontenjans (Olumsuzluk) Tabloları ve Ki-Kare Analizleri.....	242
7.7.2. Sıralı Lojit Analiz Yöntemi Uygulaması	257
7.7.3. Multinomial Logistic Regresyon Analiz Yöntemi Uygulaması.....	260
SONUÇ VE YORUMLAR.....	265
KAYNAKÇA.....	267
EK.....	270

KISALTMALAR

ASE	: Asymtotic Standart Error
BD	: Breslow-Day
CDA	: Categorical Data Analysis
CDF	: Birikimli Dağılım Fonksiyonu
CMH	: Cochran-Mantel- Haenszel İstatistiği
DF	: Serbestlik Derecesi
EİANTE	: En İyi Asimptotik Normal Tahmin Edici
GLM	: Genelleştirilmiş Lineer Model
MH	: Mantel-Haenszel
MI	: Kalp Kası Bozukluğu
ML	: Maximum Likelihood
OR	: Odds Oranı
SE	: Standart Error

TABLolar LİSTESİ

Tablo 1.1 İlk Ve İkinci Enfeksiyona Yakalanan Buzağlar	41
Tablo 1.2 Hipoteze Göre Olasılık Yapısı	42
Tablo 2.1 Cinsiyete Göre Ölümden Sonraki Hayata Olan İnançların	
Çapraz Sınıflaması	45
Tablo 2.2 Tablo-2.1 İçin Notasyonlar	48
Tablo 2.3 Göğüs Kanseri Tanılarına Göre Tahminlenen Koşullu Dağılımlar	49
Tablo 2.4 Marjinal, Koşullu Ve Ortak Olasılıkların Notasyonları	51
Tablo 2.5 Emniyet Kemeri Kullanımı Ve Otomobil Çarpışmalarının Sonuçları	55
Tablo-2.6 Akciğer Kanseri İle Sigara İçimin Çapraz Sınıflaması	56
Tablo 2.7 Kalp Kası Bozukluğu (Myocardial İnfraction) Ve Aspirin	
Kullanımının Çapraz Sınıflandırılması	60
Tablo 2.8 Katilin Irkına Ve Kurbanların Irkına Göre Ölüm Cezası Kararı	75
Tablo 2.9 Marjinal Bağımsızlığı İçermeyen Koşullu Bağımsızlık İçin	
Beklenen Değerlerin Gösterilmesi	80
Tablo 2.10 Cochran-Mantel-Haenszel Testi İle İlgili Olarak Çinli Sigara	
Kullanıcıları Ve Akciğer Kanseri Yakalananların İncelenmesi	84
Tablo 2.11 Terfilerle Çalışanların Irklarına İlişkin Olumsuzluk Tablosu	93
Tablo 2.12 Gelir Aracılığı İle İş Memnuniyetinin Çapraz Sınıflaması	100
Tablo 2.13 Bebek Sakatlığı Ve Annenin Alkol Tüketimi	114
Tablo 2.14 Fisher'ın Çay-Tatma Deneyi İçin Olumsuzluk Tablosu	123
Tablo 2.15 Tablo 2.14'in Marjinlerine Sahip Tablolar	
İçin Hipergeometrik Dağılım	123
Tablo 2.16 Veri Çizelgesine İlişkin Örnek Tablo	130
Tablo-3.1 Horlama Ve Kalp Rahatsızlığı Arasındaki İlişki	137
Tablo-3.2 İstatistiksel Analizde Model Türleri	156
Tablo 4.1 Azt Kullanımına Ve Irka Göre Aids Belirtilerinin Gelişimi	170
Tablo 4.2 Tablo 4.1' E Göre Uydurulan Lojit Model İçin Parametre Tahminleri	173

Tablo-5.1 Ölümünden Sonraki Hayata İnanmanın (Sütunlar) Cinsiyet (Satırlar) İle Olan İlişkisinin Yer Aldığı Tablo 2.1 İçin Uydurulan Bağımsız Modele Göre Parametre Tahminleri-----	184
Tablo-5.2 Doygun Log Lineer Model İçin Eşdeğer Birliktelik Parametreleri -----	189
Tablo-5.3 Lise Son Sınıf Öğrencileri İçin Alkol(A), Sigara(C) Ve Marihuana (M) Kullanımı (Kaynak: Agresti, 1996) -----	193
Tablo-5.4 Tablo-5.3' Teki Verilere Uygulanan Log Lineer Modeller İçin Beklenen Değerler -----	194
Tablo-5.5 Tablo-5.4' Te Yer Alan Log Lineer Modeller İçin Tahmin Edilmiş Odds Oranları -----	196
Tablo 5.6 Alkol (A), Sigara (C) Ve Marihuana (M) Kullanımına İlişkin Oluşturulan Log Lineer Modeller İçin Uyum İyiliği Testleri -----	199
Tablo 5.7 İki Log Lineer Model İçin Ayarlanmış Artıklar -----	201
Tablo-5.8 İki Sonuçlu Y Yanıt Değişkeni İle Üç Yönlü Bir Tablo İçin Eşdeğer Log Lineer Ve Lojit Modeller -----	209
Tablo 7.1 Cinsiyetlerine Göre Frekans ve Yüzde Değerleri	231
Tablo 7.2 Çocuk Durumlarına Göre Frekans ve Yüzde Değerleri	232
Tablo 7.3 KMO and Bartlett's Test Çıktısı	233
Tablo 7.4 Açıklanan Toplam Varyans Çıktısı	233
Tablo 7.5 Çevrilmiş Faktör Matrisi (Rotated Component Matrix)	234
Tablo 7.6 Faktör Grupları ve Ortalama Değerleri Tablosu	235
Tablo 7.7 Çeşitlilik-Hizmet ile Müşteri Odaklılık İçin Olumsuzluk Tablosu	243
Tablo 7.8 Çeşitlilik-Hizmet Faktörü ile Müşteri Odaklılık İçin Ki-Kare Analiz Tablosu	244
Tablo 7.9 Çeşitlilik-Hizmet ile Müşteri Odaklılık Arasındaki Simetrik Ölçüler Tablosu	244
Tablo 7.10. Kalite-Müşteri Odaklılık İçin Olumsuzluk Tablosu	245
Tablo 7.11 Kalite ve Müşteri Odaklılık Ki-Kare Analiz Tablosu	246

<u>Tablo 7.12 Kalite-Müşteri Odaklılık Simetrik Ölçüler Tablosu</u>	247
<u>Tablo 7.13 Beden-Müşteri Odaklılık Olumsuzluk Tablosu.....</u>	248
<u>Tablo 7.14 Beden-Müşteri Odaklılık Ki-Kare Analiz Tablosu.....</u>	248
<u>Tablo 7.15 Beden-Müşteri Odaklılık Simetrik Ölçüler Tablosu.....</u>	249
<u>Tablo 7.16 Konum-Müşteri Odaklılık Olumsuzluk Tablosu</u>	250
<u>Tablo 7.17 Konum-Müşteri Odaklılık Ki-Kare Analiz Tablosu.....</u>	250
<u>Tablo 7.18 Konum-Müşteri Odaklılık Simetrik Ölçüler Tablosu</u>	251
<u>Tablo 7.19 Cinsiyet-Müşteri Odaklılık Olumsuzluk Tablosu.....</u>	252
<u>Tablo 7.20 Cinsiyet-Müşteri Odaklılık Ki-Kare Analiz Tablosu</u>	252
<u>Tablo 7.21 Cinsiyet-Müşteri Odaklılık Simetrik Ölçüler Tablosu</u>	253
<u>Tablo 7.22 Çocuk Durumu -Müşteri Odaklılık Olumsuzluk Tablosu</u>	254
<u>Tablo 7.23 Çocuk Durumu-Müşteri Odaklılık Ki-Kare Analiz Tablosu</u>	255
<u>Tablo 7.24 Çocuk Durumu-Müşteri Odaklılık Simetrik Ölçüler Tablosu</u>	256
<u>Tablo 7.25 Sıralı Lojit Model Tablosu (1.Deneme).....</u>	258
<u>Tablo 7.26 Sıralı Lojit Model Tablosu (2.Deneme).....</u>	259
<u>Tablo 7.27 Multinomial Lojistik Regresyon Model Tablosu (1.Deneme).....</u>	261
<u>Tablo 7.28 Multinomial Lojistik Regresyon Model Tablosu (2.Deneme).....</u>	262

ŞEKİLLER LİSTESİ

<u>Şekil-1.1</u> $H_0 : \beta = 0$ 'In Sınanmasında Üç Test Çeşidi İçin Kullanılan Log- Olabilirlik Fonksiyonu Ve Bilgisi.	23
<u>Şekil-1.2</u> $N=25$ Olduğunda π Binom Parametresi İçin Nominal %95 Güven Aralıklarına Göre Kapsama Olasılıklarının Grafiği	32
<u>Şekil-2.1</u> Ölüm Cezası Alanların Yüzdesi	75
<u>Şekil 3.1</u> Kalp Rahatsızlığı Ve Horlama Verileri İçin Modellerin Uyumu	138
<u>Şekil 3.2</u> Lojistik Regresyon Fonksiyonu	139
<u>Şekil 3.3</u> Wald, Olabilirlik Oranı Ve Skor Testlerinde Kullanılan Bilgi	154
<u>Şekil 7.1</u> Cinsiyetlerine Göre Dağılım Grafiği	231
<u>Şekil 7.2</u> Çocuk Durumlarına Göre Dağılım Grafiği	232
<u>Şekil 7.3</u> Sıralı Lojit Modeli	263
<u>Şekil 7.4</u> Sıralı Lojit Model Dörtlü Uygunluk Test Grafiği.....	263

GİRİŞ

Günümüzde; sosyal bilimler alanında insan düşünce ve davranışlarına etki eden faktörlerin ölçülmesi ve değerlendirilmesinden tıp araştırmalarında özellikle tedavi yöntemlerinin tespit edilmesine kadar çok çeşitli alanlarda sıklıkla “kategorik veri analiz yöntemleri” kullanılmaktadır. Evrendeki doğal dengenin yapı taşı konumunda “normal dağılım” bulunmasına rağmen; belirli kıstaslar altında elde edilebilen “kategorik verileri” analiz edebilmek için farklı analiz yöntemlerine ihtiyaç duyulmaktadır. Uygulamada karşılaşılan veri problemlerinin çoğu kategorik veri içermektedir. Yapısı gereği kategorik özellik gösteren başlıca verilere anket yöntemi ile toplanan veriler örnek olarak gösterilebilir. Pazar araştırmasında kişilere şu şekilde sorular sorulabilir: “Bir ürünü kullanıp kullanmadıkları (Evet, Hayır)”, “Ne sıklıkla kullandıkları (Asla, haftada bir kereden az, ayda 4-10 kez, ayda 10 kereden fazla)”, “Bu üründen memnuniyetlerinin nasıl olduğunu (Kesinlikle kötü, bazen kötü, yansız, bazen memnun edici, kesinlikle memnun edici)”, “Rakip ürünlerden hangisini kullandıkları (A marka, B marka vs.)”. Bu örneklerin sonuncusu ve ilki nominal kategorik değişken, ikinci ve üçüncüsü ordinal kategorik değişkendir.

Kategorik veri analiz yöntemlerinin tarihsel gelişim sürecine bakıldığında, çok uzak bir geçmişi olmadığı görülmektedir. Sürekli verilere uygulanan metotlar yirminci yüzyılın başlarında karmaşık bir seviyeye ulaşırken, kategorik verilere uygulanacak istatistiksel metotların gelişiminde geç kalmıştır. Bu alanda, İngiliz istatistikçi Karl Pearson tarafından 1900’lü yıllarda etkili çalışmalar yapılmasına rağmen, 1960’lara kadar kategorik verilere uygulanacak istatistik metotlarının gelişimi yavaş kalmıştır. Kategorik veri analizi için istatistiksel metotların geliştirilmesinde; istatistik, biyoistatistik, ekonomi, psikoloji ve sosyoloji gibi farklı alanlardaki bilim adamlarının katkılarından büyük ölçüde yararlanılmıştır. Bu disiplinler arası köken benzer metodolojiler için çoklu yorumlar oluşturmuştur. Sonuç olarak, kategorik veri analizi bilgi olarak zengindir ve çalışma alanı geniştir. Bu disiplinler arası yapı aynı zamanda mevcut tekniklerin sentezini ve birleşimini yapmaktadır. Fakat disiplinler arası terminolojilerin ve uygulamaların farklılığı bu durumu zorlaştırmaktadır.

Kategorik veri analizi ile ilgili istatistiksel metot ve teknikler son 25-30 yılda hızlı bir gelişim göstermiş ve hala göstermektedir. Uygulamalı araştırmalarda bu metot ve tekniklerin

kullanılması son yıllarda olağan hale gelmekle beraber özellikle büyük bir bölümü uygun ticari yazılım ve masrafsız bilgisayar hesaplamaları ile yapılmaktadır.

Analizlerde, kategorik veriler için genellikle sayılabilir özet tablolar kullanılmaktadır. Bu tip verilerde, standart normal dağılım verilerinde kullanılan istatistiksel analiz yöntemlerinin kullanılması kesinlikle uygun değildir. Kategorik verilerin analizi için çeşitli istatistik metotların geliştirilmesi gerekmektedir (Simonoff, 2003:3). Bu çalışma ise; kategorik verilerin analiz yöntemlerine ilişkin, yeni gelişen ve disiplinler arası bölünmüş olan belirli metotları ve teorileri özet bir şekilde göstermek ve kategorik verilerin analizi için özel araçları kullanan sosyal bilimcilerin uygulamalarına yardım etmek amacını taşımaktadır.

Çalışmanın literatür kısmı altı bölümden oluşmaktadır. Birinci bölümde; kategorik verilerin önemli türleri ile kullanılan dağılımlar ve istatistiksel yorumlama yöntemleri tanıtılmıştır. Bu bölümde amaç hala tartışma konusu olan kategorik veri türlerinin genel bir tanımlanmasını sağlamak, yorumlayıcı veri analizinin gerektirdiği dağılım varsayımlarını kategorik veriler için ortaya koymak ve parametreler konusunda yorumlamalar yapmak için örneklem verilerinin kullanımına ilişkin metotları incelenmiştir. Bu bölümde, kategorik yanıtlar için “*Binom, Çokterimli ve Poisson dağılımı*” şeklinde üç ana dağılım gözden geçirilmiş ve istatistiksel yorumlama açısından en çok olabilirlik metodu ile bu metoda ilişkin testler ele alınmıştır. Son olarak binom ve çok terimli parametrelerin istatistiksel yorumlama yöntemleri ayrıntıları ile ele alınmıştır.

İkinci bölümde ise, kategorik değişkenler arasındaki ilişkileri gösteren olumsuzluk (kontenjans) tabloları ve birliktelikleri özetleyen parametreler tanımlanmıştır. Ayrıca bu bölümde ki-kare test yöntemlerine de yer verilmiştir. Öncelikle iki kategorik değişken arasında ki birliktelikler ve birlikteliği ifade eden parametreler tanıtılarak bu parametreler için yorumlama metotları gösterilmiştir. Sonra üçüncü bir değişkenin kontrol edilmesi ile alan genişletilmiştir. Çünkü kontrol altında bir birliktelik ciddi olarak değişebilmektedir. Daha sonra nominal ve ordinal çok kategorili değişkenler için parametreler ele alınmıştır.

Üçüncü bölümde, kategorik yanıt değişkenlerin modellenmesi üzerine odaklanılmış ve bir önceki bölümde açıklayıcı değişkenlerin kategorik yanıt değişkenler üzerindeki etkilerini araştırırken kullandığımız analizlerin temelini oluşturan modeller incelenmiştir. Modeller ile birçok açıklayıcı değişkenin etkilerinin eş zamanlı olarak analiz edilmesi gibi daha karmaşık durumlar ele alınmıştır. Modellerdeki açıklayıcı değişkenler sürekli ya da kategorik ya da her iki türde de olabilmektedir. Burada ele alınan modellerin neredeyse tümü genelleştirilmiş doğrusal modellerin özel biçimleridir. Bu modellerin genel bir durumudur ve kategorik yanıt değişkenler için oluşturulan modelleri kapsadığı gibi sürekli yanıt değişkenler için kullanılan sıradan regresyon ve ANOVA modellerini de kapsamaktadır. Ayrıca uyumluluk testlerine de yer verilmiştir.

Kategorik analiz yöntemleri arasında en çok kullanım yeri bulan lojistik regresyon analiz yöntemi dördüncü bölümde, bir diğer önemli analiz yöntemi olan Loglineer analiz yöntemi ise beşinci bölümde ayrıntılandırılmıştır. Altıncı bölümde, kategorik veri analizinin kısa bir tarihsel gelişimi sunulmuştur. Yirminci yüzyılın başlarından günümüze kadar yapılan çalışma ve katkılara değinilerek araştırmacılar için kronolojik bir kaynak listesi verilmiştir. Ayrıca bundan sonraki dönemlerde kategorik veri analizi konusunda günümüzdeki çalışmalara dayanarak hangi yönde gelişim sağlanabileceğine ilişkin tahminleme yapılmıştır.

Çalışmanın son bölümü ise kategorik veriler üzerinden yapılan bir uygulamadan meydana gelmektedir. Uygulama alanı olarak “departman mağazacılığında rekabet olanaklarının geliştirilmesi” maksadıyla bu alanda hizmet veren bir mağaza ve müşterileri seçilmiştir. Bu anlamda mağazaya ait temel süreçlere ilişkin kriterler, belirli bir anket üzerinden müşteriler ile yapılan karşılıklı görüşmeler sonucu belirlenmiştir. Anket sonuçlarının değerlendirilmesi ile iyileştirme sürecinde yapılan analizlerde kategorik veri analizi yöntemlerinden yararlanmıştır. Çalışma sonunda eldeki anket verilerine uygulanan sıralı lojit analiz yöntemi ile multinominal regresyon analiz yönteminin kullanımının karşılaştırılması hedeflenmiştir.