T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

KREDİ KARTI KULLANIMINA ETKİ EDEN FAKTÖRLERİN ÇOK AŞAMALI LOJİSTİK REGRESYON ANALİZİ

ALP ABDULLAH ÇERÇİ

YÜKSEK LİSANS TEZİ
İSTATİSTİK ANABİLİM DALI
İSTATİSTİK PROGRAMI

DANIŞMAN
DOÇ. DR. FATMA NOYAN TEKELİ

İSTANBUL, 2018

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

KREDİ KARTI KULLANIMINA ETKİ EDEN FAKTÖRLERİN ÇOK AŞAMALI LOJİSTİK REGRESYON ANALİZİ

Alp Abdullah ÇERÇİ tarafından hazırlanan tez çalışması 25.05.2018 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri İstatistik Anabilim Dalı'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Tez Danışmanı

Doç. Dr. Fatma NOYAN TEKELİ

Yıldız Teknik Üniversitesi

Jüri Üyeleri

Doç. Dr. Fatma NOYAN TEKELİ

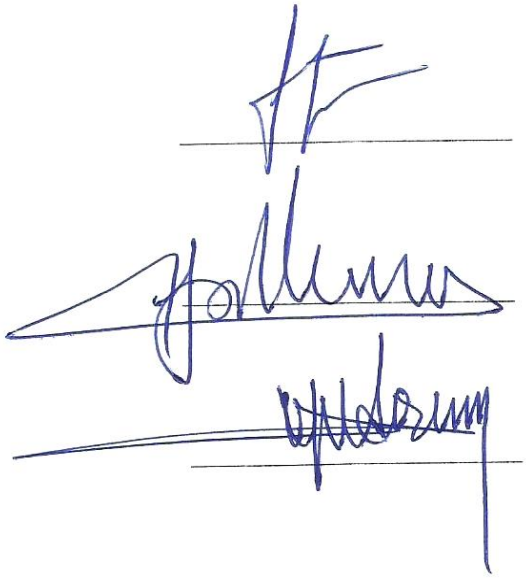
Yıldız Teknik Üniversitesi

Prof. Dr. Gülhayat Gölbaşı ŞİMŞEK

Yıldız Teknik Üniversitesi

Prof. Dr. Hakan YILDIRIM

Marmara Üniversitesi



ÖNSÖZ

Tez çalışmam boyunca, yardımlarını ve görüşlerini benden hiç esirgemeyen, içtenlikle yaptıklarımın arkasında duran ve benimle gurur duyduğuna inandığım, kendisini hayatımın birçok alanında örnek aldığım değerli danışman hocam sayın Doç. Dr. Fatma Noyan Tekeli'ye ve her zaman fikir ve önerilerine değer verdiğim değerli YTÜ İstatistik Bölümü hocalarıma da teşekkürü bir borç biliyorum.

Tez jürimde bana fikirleriyle aydınlık bir yol açan hocalarıma içtenlikle teşekkür ederim.

Hayatım boyunca benden desteklerini maddi ve manevi olarak hiçbir zaman esirgemeyen Annem ve Babam'a teşekkür ederim.

Bu çalışma her an yanımda olan canım aileme armağanımdır.

Mayıs, 2018

Alp Abdullah ÇERÇİ

İÇİNDEKİLER

Sayfa

SİMGE LİSTESİ vi

KISALTMA LİSTESİ ix

ÇİZELGE LİSTESİ x

ÖZET xi

ABSTRACT xii

BÖLÜM 1

GİRİŞ 1

1.1 Literatür Özeti 1
1.2 Tezin Amacı 2
1.3 Hipotez 3

BÖLÜM 2

ÇOK AŞAMALI MODELLER 3

2.1 Çok Aşamalı Regresyon Modelleri 4
2.1.1 Çok Aşamalı Modellerin Sınıflandırılması 7
2.1.1.1 Tesadüfi Etkili Tek Yönlü Anova 7
2.1.1.2 Tesadüfi Katsayılı Regresyon Modeli 9
2.1.1.3 Sabit ve Eğim Parametresi Çıktı Olan Modeller 11
2.1.2 Çoklu Çok Aşamalı Regresyon Modeli 11
2.2 Çok Aşamalı Modeller İçin Tahmin Prensipleri 11
2.2.1 Sabit Etkilerin Tahmini 11
2.2.2 1. Aşama Tesadüfi Katsayılarının Tahmini 15
2.2.3 Varyans- Kovaryans Bileşenlerinin Tahmini 18
2.2.4 Tam En Çok Olabilirlik Yöntemi 18
2.2.5 Sınırlandırılmış En Çok Olabilirlik Yöntemi (Restricted Maximum Likelihood-REML) 20
2.2.6 EM Algoritması Kullanılarak Yapılan ML Tahminleri 22
2.3 Çok Aşamalı İçin Hipotez Testi 24
2.4 Sabit Etkiler İçin Hipotez Testleri 26

2.5 Birinci Aşama Tesadüfi Parametreleri İçin Hipotez Testleri 29
2.6 Varyans Ve Kovaryans Bileşenleri İçin Hipotez Testi.....31

BÖLÜM 3

İKİ AŞAMALI LOJİSTİK REGRESYON ANALİZİ.....33

3.1 1. Aşama Modelleri 33
3.1.1 1.Aşama Örnekleme Modeli 34
3.1.2 1. Aşama Bağlantı Fonksiyonu 34
3.1.3 1. Aşama Yapısal Modeli.....35
3.2 2. Aşama Modelleri 36
3.3 İki Aşamalı Lojistik Regresyon Analizinde Parametre Tahmini 36
3.3.1 Tesadüfi Etkiler yada Varyans-Kovaryans Bileşenlerinin Tahmini 38
3.3.2 Sabit Etkilerin (γ) Tahmini 38
3.4 Hipotez Testleri 38

BÖLÜM 4

UYGULAMA 40

4.1 Veri ve Değişkenlerin Tanımları 40
4.2 Betimsel İstatistikler.....41
4.3 İki Aşamalı Lojistik Regresyon Analizi.....49
4.3.1 1.Model: Koşulsuz Model 49
4.3.2 2. Model: Koşullu Model 50
4.3.3 3.Model (Koşullu Model).....52

BÖLÜM 5

SONUÇ VE ÖNERİLER 56

KAYNAKLAR 58

ÖZGEÇMİŞ.....60

$i = 1,2,.....,n_j$	i indisi 1. aşama birimleri
$j = 1,2,.....,J$	j indisi 2. aşama birimleri
n_j	j . gruptaki birim sayısı
J	Grup sayısı
Y_{ij}	m_{ij} denemedeki başarı sayısı
φ_{ij}	Her denemedeki başarı olasılığı
β_{qj}	1.aşama katsayıları
X_{qij}	1. aşama açıklayıcı değişkenleri
ρ	Gruplar içi korelasyon katsayısı
$\tilde{n}_{grup}$	Gruplar için ortalama örnek büyüklüğü
τ_{00}	Gruplar arası varyans
σ^2	Gruplar içi varyans (1. aşama varyansı)
n_j	j . gruptaki birim sayısını
J	Grup sayısı
Y_{ij}	j .gruptaki i .birime ait gözlenen çıktı (bağımlı değişken)
X_i	1.aşama açıklayıcı değişkeni
β_0	Sabit parametre ya da kesim noktası
β_1	Eğim parametresi
r_i	Hata terimi
r_{ij}	j. gruptaki i. birim için tesadüfi etki
β_{0j}	j. grup için regresyon sabitidir
γ_{00}	Gerçek sabit parametre β_0'ın tahmincisi($E(\beta_0) = \gamma_0$)
γ_{01}	2. aşama değişkeni W_j için eğim parametresi
u_{0j}	j. grubun W_j' ye göre koşullandırılmış, ortalama çıktı üzerindeki etkisi (2. aşama hata payı)
β_{1j}	j. grup için eğim parametresi
γ_{10}	1.aşamadaki eğim katsayılarının ortalama değeri, (gerçek eğim parametresi β_{1j}'nin tahmincisi/ $E(\beta_1) = \gamma_1$)
γ_{11}	2.aşama değişkeni W_j için eğim katsayısı ve

u_{1j}	j. grubun W_j ' ye göre koşullandırılmış eğim parametresi üzerindeki etkisi (2. aşama hata payı/ tesadüfi etkisi)
W_j	2. aşama açıklayıcı değişkeni
τ_{11}	Eğim parametreleri arasındaki ana kütle varyansı
τ_{01}	Eğim ve sabit terimler arasındaki ana kütle kovaryansı
μ_{Y_J}	j. grup için ortalama çıktı
$\overline{Y}_{\bullet j}$	Örnek ortalaması
V_j	Hata varyansı (β_{0j} 'nin tahmini $\overline{Y}_{\bullet j}$ 'nin varyansı)
Δ_J	$\overline{Y}_{\bullet j}$ 'nin kalıntı varyansıdır yani W_j verildiği durumda $\overline{Y}_{\bullet j}$ 'nin koşullu varyansı
$\tilde{\gamma}_{00}$	γ_{00}'ın tek, minumum varyanslı, yansız tahmin edicisi
Δ_j^{-1}	$\overline{Y}_{\bullet j}$ 'nin kesinliği
$\hat{\gamma}_{00}$	γ_{00}'ın ağırlıklı en küçük kareler tahmincisi (aynı zamanda en çok benzerlik tahmincisi)
$\Sigma\Lambda_j^{-1}$	$\hat{\gamma}_{00}$'ın kesinliği
Var($\hat{\gamma}_{00}$)	$\hat{\gamma}_{00}$' ın varyansı[Var($\hat{\gamma}_{00}$)= $(\sum\Delta_j^{-1})^{-1}$]
$\tilde{\gamma}_{01}$	γ_{01}'in en küçük kareler tahmincisi
$\hat{\gamma}_{01}$	γ_{01}'in ağırlıklı en küçük kareler tahmincisi
Y_j	$n_j \times 1$ boyutlu bağımlı değişken vektörü
β_j	$n_j \times 1$ boyutlu bilinmeyen parametre (katsayılar) vektörü
I	$n_j \times n_j$ boyutlu birim matris olarak tanımlanır
r_j	$n_j \times 1$ boyutlu tesadüfi hatalara ait vektör
$\hat{\beta}_j$	Katsayılar vektörü, β_j ' ye ait sıradan en küçük kareler tahmincisi
T	2. aşama tesadüfi etkilerine ait varyans-kovaryans matrisi
$V_{\hat{\gamma}}$	γ 'nın dispersion (dağılım) matrisi
Z_j,	1. aşama açıklayıcı değişkeni X_j 'nin alt kümesi
λ_j	β_{0j} için yapılan en küçük kareler tahmincisi ($\overline{Y}_{\bullet j}$)'nin güvenirliği
$\beta^*{}_{0j}$	Çekilmiş tahminci (shrinkage estimator)
u^*_{0j}	Ampirik Bayes tahmincisine $\beta^*{}_{0j}$ 'ye karşılık gelen ampirik Bayes kalıntısı
Λ_j	Çok değişkenli güvenirlik matrisi
$V_{qq}{}^*{}_j$	V_j^* 'nin q. köşegen elemanı
N	1. aşama birimlerinin toplam örnek büyüklüğü
l_{ML}	Tam en çok olabilirlik yöntemi ile en çok olabilirlik fonksiyonunu
F	γ sabit etkiler vektöründeki toplam eleman sayısı
$C'\hat{\gamma}$	Contrast vektör
D_i	i. modele ait deviyans istatistiği
L_0	Sıfır hipotezindeki en çok olabilirlik tahminleri için olabilirlik değeri

D_1	Alternatif modeldeki en çok olabilirlik tahminleri için olabilirlik değeri
m	Modeldeki 1. aşama tesadüfi açıklayıcı değişken sayısı
ls_{tot}	j . grup için tahmin edilen kalıntı varyanslarının logaritmasının ağırlıklı ortalaması
s_j^2	j grup için tahmin edilen kalıntı varyansları
df_j	Serbestlik derecesi
d_j	Gruba bağlı kalıntı varyansı s_j^2'nin standardize edilmiş kalıntı dağılımı ölçüsü
H	1. aşama kalıntıları için heterojenlik istatistiği
$\overline{r}_i$	Standardize E.K.K kalıntıları
C_j^{OF}	Verilen veri seti için tahmin edilen sabit etkiler ile j. birim çıkarılarak elde edilen veri seti için tahmin edilen sabit etkiler arasındaki fark (etkili gözlem teşhis değerleri)
$\hat{\Sigma}_F$	Sabit etkilere ait tahmin vektörünün kovaryans matrisi
φ	Tesadüfi kısımdaki tüm parametrelerin vektörü
$\hat{\varphi}_{(-j)}$	Tesadüfi kısımdaki j. birim silindiğinde elde edilen tahminler yani birinci adım tahminleri
C_j^R	Tesadüfi kısımdaki parametre tahminleri üzerinde j. birimin etkisi için teşhis değerleri
C_j	Modelin sabit ve tesadüfi kısmındaki parametreler üzerinde j. grubun birleştirilmiş teşhis etkisi
D_j	j. grup için çok değişkenli kalıntılar

KISALTMA LİSTESİ

AB	Ampirik Bayes
E.K.K.	En Küçük Kareler Yöntemi (OLS)
EÇO	En Çok Olabilirlik Yöntemi
EÇOT	En Çok Olabilirlik Tahmini
EM	Expectation-Maximization Algoritması
G.E.K.K	Genelleştirilmiş En Küçük Kareler Yöntemi
HLM	Hiyerarşik Lineer Modeller (Hierarchical Linear Models)
PQL	Penalized Quasi-Likelihood
SEÇOT	Sınırlandırılmış En Çok Olabilirlik Tahmini
SEÇOY	Sınırlandırılmış En Çok Olabilirlik Yöntemi
TEÇOY	Tam En Çok Olabilirlik Yöntemi

ÇİZELGE LİSTESİ

Sayfa

Çizelge 2.1 Çok aşamalı modeller için test edilebilir hipotez tipleri 25
Çizelge 2.2 Hiyerarşik lineer modeller için genel hipotez testleri..... 26
Çizelge 3.1 İki aşamalı lojistik regresyon modelleri 35
Çizelge 4.1 Kredi kartı kullanımı..... 41
Çizelge 4.2 Banka şubelerinin bulunduğu iller ve müşteri sayıları..... 42
Çizelge 4.3 Banka müşterilerinin cinsiyete göre dağılımı..... 42
Çizelge 4.4 Banka müşterilerinin yaşa göre dağılımı..... 43
Çizelge 4.5 Banka müşterilerinin medeni duruma göre dağılımı..... 43
Çizelge 4.6 Banka müşterilerinin eğitim durumuna göre dağılımı..... 44
Çizelge 4.7 Banka müşterilerinin meslek durumuna göre dağılımı 45
Çizelge 4.8 Banka müşterilerinin mülk sahipliğine durumuna göre dağılımı 46
Çizelge 4.9 Banka müşterilerinin aylık gelir miktarına göre dağılımı 46
Çizelge 4.10 Banka müşterilerinin sosyal güvence sahibi olma durumuna göre dağılımı..... 47
Çizelge 4.11 Kredi kartı kullanımının cinsiyete göre dağılımı 47
Çizelge 4.12 Kredi kartı kullanımının medeni duruma göre dağılımı..... 48
Çizelge 4.13 Kredi kartı kullanımının ailenin aylık gelirine göre dağılımı 48
Çizelge 4.14 Koşulsuz Model- tek yönlü ANOVA modeli sonuçları..... 49
Çizelge 4.15 Koşullu model için sonuçlar 51
Çizelge 4.16 Koşullu model için varyans-kovaryans bileşen tahminleri..... 52
Çizelge 4.17 Koşullu son model için sonuçlar 53
Çizelge 4.18 Koşullu son model için varyans-kovaryans bileşen tahminleri 54

KREDİ KARTI KULLANIMINA ETKİ EDEN FAKTÖRLERİN ÇOK AŞAMALI LOJİSTİK REGRESYON ANALİZİ

Alp Abdullah ÇERÇİ

İstatistik Anabilim Dalı

Yüksek Lisans Tezi

Tez Danışmanı: Doç. Dr. Fatma Noyan TEKELİ

Bu çalışmada, Türkiye'de kredi kartı kullanımı Çok aşamalı lojistik regresyon analizi ile incelenmiştir. Çok aşamalı lojistik regresyon analizi, bağımlı değişkenin iki kategoriye sahip dikotom değişken olduğu kümelenmiş ya da hiyerarşik yapıya sahip verilerin istatistiksel analizinde kullanılır. Hiyerarşik yapıya sahip verilerde, standart istatistiksel testler için gerekli olan gözlemlerin birbirinden bağımsızlığı varsayımı bozulduğundan çok aşamalı lojistik regresyon modellerine başvurulur. Çalışmada özel bir bankanın 59 şubesinde kümelenmiş olarak bulunun 399 müşteriye ait verilerden yararlanılmıştır. Müşterilerin kredi kartı kullanımı, bankaların şubelerine göre farklılık göstermektedir. Çalışmadan elde edilen bulgulara göre, kredi kartı kullanımını müşterinin cinsiyeti, medeni durumu, mesleği ve geliri etkilemektedir. Ortalama kredi kartı kullanımı banka şubelerine göre farklılık gösterirken, cinsiyetin, medeni durumun, mesleğin ve gelirin kredi kartı kullanımı üzerinde etkisi banka şubelerine göre farklılık göstermemektedir.

Anahtar Kelimeler: Çok Aşamalı Lojistik Regresyon Analizi, Kredi Kartı Kullanımı

ABSTRACT

MULTILEVEL LOGISTIC REGRESSION ANALYSIS OF FACTORS INFLUENCING THE USAGE OF CREDIT CARDS

Alp Abdullah ÇERÇİ

Department of Statistics

MSc. Thesis

Adviser: Assoc. Prof. Dr. Fatma Noyan TEKELİ

Credit cards have become one of the most widely accepted, useful and profitable financial products around the world. Many consumers and merchants around the world accept it as a routine payment instrument for the full range of products and services. The study aimed to recognize the factors affecting for usage of credit cards with multilevel logistic regression analysis. Especially, whether gender is effective on the use of credit cards. It is useful to understand hierarchical data structures such as multi-level models, students in class, employees in firms, customers in bank branches. Multilevel (hierarchical) logistic regression analysis is the method of finding the relationship between two variables or more for the multilevel data and the dependent variable of multilevel logistic regression is the dichotomous variable, which is distributed as the Bernoulli distribution such as using credit cards, marrying, adopting a new technology. In this article, we were applied the multilevel analysis approach to the 399 bank customer data and were studied the characteristics of credit card users nested within 59 bank branches. According to the results of the study, gender, marital status, occupation and income was found to be significantly influence on credit card usage. And the Credit card usage averages were varied in bank branches.

Keywords: Multilevel Logistic Regression Analysis, Usage Of Credit Cart

BÖLÜM 1

GİRİŞ

1.1 Literatür Özeti

Lindley ve Smith 1972 yılında çok aşamalı modelleri ilk defa ortaya koymuştur. Dempster, Rubin ve Tsutakawa, hiyerarşik veri yapılarında çok aşamalı modellerin uygulamaları üzerinde çalışmalar yapmışlardır. Mason, Wong ve Etwiste ise 1983'de çok aşamalı modellerin uygulanabilirliği üzerine çalışırken aynı dönemde Atkin ve arkadaşları sosyal bilimlerde çok aşamalı modellerin uygulamalarını geliştirmişlerdir. Bu yıllardan sonra ise özellikle sosyal bilimler, eğitim bilimleri, tıp gibi alanlarda çok aşamalı modeller kullanılmaya başlanmıştır. Teorik altyapısı da 1980'lerden sonra hızla geliştirilmeye devam edilmiştir [1].

Tezde çok aşamalı lojistik regresyon analizinden yararlanarak kredi kartı sahipliğini etkileyen faktöler araştıralacaktır. Litaratürde kredi kartı kullanımı ile müşterilerin demografik özellikleri arasındaki ilişkileri ortaya koymaya yönelik olarak yapılan çalışmalar vardır. Kim ve De Vaney (2001) tarafından yapılan çalışma aylık gelir, varlık miktarı ve kredi kartı sayısı ile kredi kartı borcu arasında ilişki olduğunu göstermiştir. Torlak (2002), eğitim ve meslek gibi demografik değişkenlerin kredi kartı kullanım amacında farklılıklar yarattığını ve gelir düzeyi ile harcama miktarları arasında ilişki olduğunu göstermiştir. Ceylan (2006) yaptığı çalışmada yaş, eğitim düzeyi, mesleğin sahip olunan kredi kartı sayısı ile ilişkili olduğunu göstermiştir. Gökçen ve Kaya (2009) cinsiyet, yaş, eğitim durumu, meslek gibi demografik özelliklerin kredi kartın kullanılma olasılığı üzerindeki etkileri araştırmıştır. Cengiz (2009), kredi kartını elde tutma olasığının yaş ve cinsiyetin etkilediğini ortaya koymuştur. Ucal ve arkadaşları (2011) Türkiye'de üniversite öğrencileri üzerinde yaptıkları çalışma ile cinsiyetin kredi kartı

kullanımı üzerindeki etkisini araştırmıştır ve erkeklerin kredi kartı kullanımının bayanlara göre daha yüksek olduğunu göstermiştir [2]. Aslanoğlu ve Korga (2017), kredi kartı kullanımı üzerinde cinsiyet ve medeni durumun etkisini araştırırken, kredi kartı kullanımının satın alma davranışları üzerindeki etkisini açıklamaya çalışmıştır. Tezde bu literatür göz önünde bulundurularak, banka müşterilerinin kredi kartı kullanımı etkileyeceği düşünülen cinsiyet, medeni durum, eğitim düzeyi, sosyal güvence sahibi olma durumu ve aylık gelir değişkenleri modele dahil edilmiştir.

1.2 Tezin Amacı

Tezin amacı, Türkiye'de kredi kartı kullanımını etkileyen faktörleri ortaya çıkarmak ve bu faktörlerin etkilerinin banka şubelerine göre farklılık gösterip göstermediğini araştırmaktır.

1.3 Hipotez

Bankacılık sektöründe kredi kartı kullanımını etkilyen faktörler çok aşamalı lojistik regresyon analizinden yararlanarak ortaya çıkarılabilir.

BÖLÜM 2

Tezde, ikinci bölümde çok aşamalı modeller iki aşamalı regresyon analizi birinci ve ikinci aşama modelleri, parametre tahminleri ve hipotez testleri çerçevesinde açıklanacaktır. Üçüncü bölümde çok aşamalı lojistik regresyon analizi hakkında bilgi verilecektir. Dördüncü bölümde karşılığı olmadan yapılan hızlı tüketimin bir numaralı aracı olan kredi kartları üzerine bir uygulama yapılacaktır. Beşinci bölümde ise sonuçlar verilecektir.

2. ÇOK AŞAMALI MODELLER

Çok aşamalı veri setlerinde gruplar, gruplar içinde birimler şeklinde devam eden hiyerarşik bir yapı vardır. İlk olarak yüksek seviyeli birimlerden (gruplar) daha sonra alt birimlerden örnekler çekilerek çok aşamalı örnekler oluşturulduğu varsayılırsa, bu tip örneklerde aynı grup içinde bulunan birimler, basit tesadüfi örnekleme yolu ile elde edilen birimlere göre daha fazla benzerlik gösterme eğiliminde olacaklardır [3]. Aynı grupta bulunan birimlerin çevre, deneyim, demografik özellikler gibi benzer karakteristiklere sahip olmaları birçok analiz için gerekli olan gözlemlerin birbirinden bağımsızlığı varsayımının çiğnenmesine neden olur [4]. Bu yüzdende, birden fazla aşamadan meydana gelen hiyerarşik veri setinde, gözlemlerin birbirine bağımlı olmasının yaratacağı olumsuz etkilerden kurtulmak için çok aşamalı modellemeden yararlanılabilir [5].

Okullardaki öğrenciler üzerinde yapılan eğitim çalışmaları, ailelerin fertleri üzerinde yapılan sosyolojik çalışmalar, şirketlerin farklı birimlerinde çalışan personel üzerinde yapılan performans çalışmaları ve hastanelerin farklı kliniklerindeki hastalar üzerinde yapılan sağlık çalışmaları gibi okulların, ailelerin, şirketlerin, kliniklerin grupları

oluşturduğu kümelenmiş veri seti üzerinde yapılan çalışmalar çok aşamalı modellere örnek oluşturabilir.

Literatürde, " Tesadüfi katsayılı modeller", " varyans bileşenleri modelleri", "hiyerarşik lineer modeller", "çok aşamalı regresyon modelleri" ve "karma lineer modeller" gibi farklı isimler kullanılmaktadır.

Çok aşamalı veri setinde, ilk aşamada birimler diğer aşamalarda ise birimlerin içinde bulunduğu gruplar, grupların içinde bulunduğu toplumlar şeklinde devam eden bir hiyerarşik bir yapı vardır. Bu hiyerarşik yapı içerisinde aşamalar farklı şekillerde adlandırılabilir. Çok aşamalı modeller için grupların sınırlarının temiz yani belirgin olması gerekir. Eğer düzensiz değişkenler varsa bu değişkenler kendileri için uygun seviyelere yerleştirilmelidir [3]. Çok aşamalı modellerde, birey veya birim aşamasından başlayarak, birimler, birimlerin meydana getirdiği gruplar, grupların meydana getirdiği topluluklar vb. şeklinde devam eden hiyerarşik yapının her aşaması için ayrı ayrı modeller kurulur. Ve daha sonra bu modeller birleştirilerek birleşik modele ulaşılır. Kavramsal olarak, hiyerarşik lineer model, regresyon modelinin hiyerarşik bir sistemi olarak ifade edilebilir [6]. Çok aşamalı modellerde, klasik regresyon analizinden farklı olarak sadece birimler bazında değil gruplar bazında da değişkenlerden yararlanılmaktadır. Ve çok aşamalı modellerde kullanılan tüm bu değişkenler bulundukları aşamalara göre farklı şekillerde adlandırılabilirler. Çok aşamalı modeller için, HLM, MlwİN, M-plus, Lisrel, SPSS, SAS gibi paket programlarından yararlanılabilir.

2.1 Çok Aşamalı Regresyon Modelleri

Tüm çok aşamalı regresyon modelleri, hiyerarşik veri setleri ile en düşük seviyede ölçülmüş bağımlı değişken ve varolan tüm seviyelerde ölçülmüş açıklayıcı değişkelerin varlığı varsayımına dayanmaktadır. Çok aşamalı regresyon modellerinde ilk aşamayı birimler bazında kurulan modeller oluştururken, diğer aşamaları ise gruplar bazında kurulan modeller meydana getirir. Kavramsal olarak, model regresyon eşitliğinin hiyerarşik bir sistemi olarak görünmektedir [7]. Tezde, iki aşamalı lojistik regresyon modelleri üzerinde durulacaktır. İlk olarak, iki aşamalı klasik regresyon modeli hakkında kısaca bilgi verilecektir.

İki aşamalı modelde, 1. aşama birimleri "birimler", 2. aşama birimleri ise "gruplar" olarak tanımlanabilir. $i = 1, 2, ..., n_j$ olmak üzere i indisi 1. aşama birimlerini; $j = 1, 2, ..., J$ olmak üzere j indisi ise 2. aşama birimlerini yani grupları göstermektedir. n_j, j . gruptaki birim sayısını, J ise grup sayısını gösterir. Gruplardaki birim sayıları farklı olabilir($j \neq k$ için $n_j = n_k$ olmak zorunda değildir.) [8].

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}) + r_{ij} \qquad (2.1)$$

Y_{ij}, j .gruptaki i .birime ait gözlenen çıktı (bağımlı değişken), X_i ise 1.aşama açıklayıcı değişkeni olmak üzere 1.aşama modeli için regresyon eşitliği (1.aşama modeli) (2.1) de verilmiştir. β_{0j}, sabit parametre olarak adlandırılır ve regresyon doğrusunun y eksenini kestiği noktayı gösterir. Yani 1.aşama X_i açıklayıcı değişken değerleri 0 olduğunda, bağımlı değişkenin aldığı değerleri gösterir. β_{1j} ise eğim parametresidir ve X_i değişkeninde meydana gelen bir birim artışa yada azalışa karşılık Y_{ij} değişkeninde beklenen değişikliği göstermektedir. β_{1j}, 1.aşama açıklayıcı değişkenlerinin, Y_{ij}'nin tahminine (kestirimine) yaptıkları göreceli katkıyı gösterdiğinden kısmi regresyon katsayısı olarak da adlandırılabilirler.

Hata terimi r_{ij}, i. 1.aşama birimi için tesadüfi etkiyi göstermektedir. Çok aşamalı regresyon modellerinde r_{ij}'nin 0 ortalama ve σ^2 varyans ile normal ve birbirinden bağımsız dağıldığını varsayılır [$r_i \sim N(0, \sigma^2)$].

(2.1)'de görüldüğü üzere, klasik regresyon analizinden farklı olarak çok aşamalı regresyon modellerinde 2.aşama birimleri (gruplar) değiştikçe onlara ait olan parametre tahminleri de değişecektir. Modelde eğim ve sabit parametreler hangi gruba ait olduklarını gösteren j indisine sahiptir. Çok aşamalı regresyon modellerinde, sabit ve eğim katsayılarının gruplara göre değiştiği varsayıldığından bunlar tesadüfi katsayılar olarak adlandırılır [3]. Hiyerarşik regresyon modelinin ikinci adımını, (2. aşama) grup seviyesinde tanımlanan açıklayıcı değişken ile regresyon katsayılarındaki (β_j) gruplara göre değişimlerin tahmin edilmesi oluşturur. 2. aşama modeli (2.2) de verilmiştir.

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_J + u_{0j} \quad (2.2a)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_J + u_{1j} \quad (2.2b)$$

β_{0j}, j. grup için regresyon sabitidir ve (2.2a) da bu sabitin gruplara karşı tesadüfi olarak değiştiği ve bu değişimin bir kısmının gruplara ait değişkenlerle açıklanabildiği model kurulmuştur. Bu modelde; γ_{00}: 1. aşama çıktılarının ortalama değeri, γ_{01}: 2. aşama (okul seviyesi) değişkeni W_j için eğim parametresi, u_{0j}: j. grubun W_j' ye göre koşullandırılmış ortalama çıktı üzerindeki etkisini göstermektedir [9].

β_{1j}: j. grup için eğim parametresidir ve bu parametrede gruplara karşı meydana gelen değişimlerin bir kısmı gruplara ait özellikler bir kısmı ise tesadüfi etkiler ile açıklanmaktadır. (2.2b) de ; γ_{10}:1.aşamadaki eğim katsayılarının ortalama değeri, γ_{11}: 2.aşama değişkeni W_j için eğim katsayısı ve u_{1j}: j. grubun W_j' ye göre koşullandırılmış eğim parametresi üzerindeki etkisidir.

u_{0j} ve u_{1j} 0 ortalamalı, sırasıyla τ_{00}ve τ_{11} varyanslarına sahip ve kovaryansları τ_{01} olan tesadüfi değişkenler olduğu varsayılır. (2.2) deki varyans-kovaryans bileşenleri şu anda, tesadüfi ya da koşullu bileşenlerdir. Yani W_j (2.aşama açıklayıcı değişkeni) için yapılan kontrollerden sonra β_{0j}, β_{1j} de arta kalan değişkenliği gösterirler.

(2.2a) ve (2.2b) de regresyon parametreleri olan γ'ların gruptan gruba değişmediği varsayılır.(γ'lar hangi gruba ait olduklarını gösteren j alt indisine sahip değillerdir yani tüm gruplar için aynı değere sahiptirler.) Bu yüzden γ'lar sabit katsayılar (fixed cefficients) olarak adlandırılırlar. Hata payı terimleri u_j lerin (hangi gruba ait olduklarını gösteren j indisine sahip) varlığında 2.aşama değişkeni W_j ile yapılan tahmin işlemelerinden sonra, β katsayıları gruplar arası değişimlerden arındırılmış olur[9].

(2.2a) ve (2.2b) deki eşitlikler (2.1) de yerine konulduğunda birleştirilmiş modele ulaşılır. 2 aşama için birleştirilmiş model (2.3) deki eşitlikte verilmiştir:

Sabit kısım Tesadüfi kısım

$$Y_{ij} = \overbrace{\gamma_{00} + \gamma_{01} W_j + \gamma_{10}(X_{ij} - \bar{X}_{\bullet j})}^{} + \underbrace{\gamma_{11} W_j (X_{ij} - \bar{X}_{\bullet j})}_{\text{Etkileşim terimi}} + \overbrace{u_{0j} + u_{1j}(X_{ij} - \bar{X}_{\bullet j}) + r_{ij}}^{} \quad (2.3)$$

Çok aşamalı modellerde, eşitlik (2.3) de görüldüğü üzere birleştirilmiş model çok kompleks bir yapıya sahiptir ve yorumlanması oldukça zordur. Çok aşamalı analizde hesaplamalarda ve yorumlamalarda kolaylık sağlaması amacıyla alt modellerden yararlanılabilir [9].

Kullandığımız terminolojiyi iki aşamalı hiyerarşik modellere uygun hale getirirsek, (2.1) deki eşitlik 1. aşama modeli, (2.2) deki eşitlik ise 2. aşama modelini ve (2.3) deki eşitlik ise birleştirilmiş modeli gösterir. 1. aşamadaki birimler, 2. aşamada ise gruplar vardır. r_{ij}, 1. aşama tesadüfi etkileri, u_{0j} ve u_{1j} ise 2. aşama tesadüfi etkilerdir. Bundan başka, Var(r_{iJ}) 1. aşama varyansını, Var(u_{0j}), Var(u_{1j}) ve Cov(u_{0j},u_{1j}) ise 2. aşama varyans kovaryans bileşenlerini gösterir. 1. aşamadaki β parametreleri, 1. aşama katsayılarını ve γs ise 2. aşama katsayılarını gösterir.

2.1.1 Çok Aşamalı Modellerin Sınıflandırılması

2.1.1.1 Tesadüfi Etkili Tek Yönlü Anova

Uygulamada hiyerarşik lineer modellerin özel hallerinden meydana gelen çeşitli alt modellerden yararlanarak, hiyerarşik yapıya sahip veri seti için en uygun model kurulmaya çalışılır. İlk olarak tesadüfi etkili tek yönlü ANOVA modeli kurularak çok aşamalı modellere başlanılır amaç bağımlı değişkendeki değişkenliği hiyerarşinin farklı aşamalarına göre ayırmaktır. Bu model sadece tesadüfi gruplar ve bu gruplar içerisindeki değişkenliği içerir [9].

Y_{ij}, j. gruptaki i. birime ait gözlenen çıktı olmak üzere 1.aşama modelinde bağımlı değişken, grup ortalamaları artı tesadüfi hatanın fonksiyonu olarak modellenir. 1. aşama modeli (birim aşaması modeli);

$$Y_{ij} = \beta_{0j} + r_{ij} \quad (2.4)$$

şeklindedir.

Her 1. aşama hata terimi r_{ij}'nin, 0 ortalama ve σ^2 1. aşama sabit varyansı ile normal ve birbirinden bağımsız dağıldığını dağıldığını varsayılır. β_{0j}, j. grup için ortalama çıktıyı gösterir, yani $\beta_{0j} = \mu_{Y_j}$.

2. aşama modeli bölümlerin sabit parametreleri arasındaki değişkenliği ifade eder. β_{0j}, yani bölümlere ait ortalama çıktı, büyük ortalama (grand mean) etrafında tesadüfi olarak değişmektedir.

Tesadüfi etkili tek yönlü ANOVA için 2. aşama modeli aşağıdaki gibidir: (γ_{01}=0 olan)

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{2.5}$$

γ_{00}, (grand- mean) ana kütle içindeki büyük ortalama (J tane grubun ortalamalarının ortalaması)

u_{0j} ise, j. gruba ait 0 ortalamalı ve τ_{00} varyanslı tesadüfi etkiyi gösterir[9].

(2.5) deki eşitliği (2.4) de yerine koyduğumuzda ise;

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \tag{2.6}$$

şeklindeki γ_{00}büyük ortalamalı, u_{0j}(2. aşama)grup etkili ve r_{iJ} (1. aşama) birey etkili tek yönlü ANOVA modeli elde edilir. Bu model tesadüfi etkili modeldir(random-effects) çünkü grup etkileri tesadüfi olarak yorumlanır. Çıktıya ait varyans ise;

$$Var(Y_{ij}) = Var(u_{0j} + r_{ij}) = \tau_{00} + \sigma^2 \tag{2.7}$$

şeklindedir.

(2.5) ve (2.6) daki hiyerarşik model eşitlikleri, 1. yada 2. aşamaya ait hiçbir tahminci içermedikleri için "tamamen koşulsuz model" (fully uncondational) ya da "boş model-empty model" olarak da adlandırılırlar[9].

Tek yönlü ANOVA modelinin tahmini, hiyerarşik veri analizinin hazırlayıcı adımı olarak yararlıdır. Büyük ortalama γ_{00} (grand mean) için nokta ve güven aralığı tahmini yapar. Her iki aşamada çıktıdaki değişkenlikle ilgili bilgi verir. σ^2 parametresi gruplar içi

değişkenliği (gruplar içi varyans), τ_{00} ise gruplar arası değişkenliği (gruplar arası varyans) gösterir. Gruplar içi varyans, grupların kendi ortalamaları etrafındaki dağılım ölçüsü ve gruplar arası varyans ise grup ortalamaları arasındaki varyanstır [9].

Tek yönlü ANOVA modelinden elde edilen yararlı bir parametre ise, gruplar içi korelasyon katsayısıdır.(intracluster correlation coefficient). Ve aşağıdaki gibi hesaplanır [10].

$$\rho = \rho(Y_{ij}, Y_{i'j}) = \frac{\tau_{00}}{(\tau_{00} + \sigma^2)} \tag{2.8}$$

yani;

ρ = gruplar arasındaki ana kütle varyansı/ toplam varyanstır. (2.aşama değişkenliğinin toplam değişkenliğe oranı) Bu parametre "gruplar içi korelasyon katsayısı" (intraclass correlation) olarak adlandırılır. Çünkü, tesadüfi olarak seçilen gruptan tesadüfi olarak seçilen birimler arasındaki korelasyona eşittir.

2.1.1.2 Tesadüfi Katsayılı Regresyon Modeli

Tesadüfi katsayılı regresyon modelinde ise, 1. aşama sabit ve eğim parametrelerinin 2. Aşama da tesadüfi olarak değiştiği modeller kurulur. 1. aşama modeli (2.1) deki modelin aynısıdır. 2. aşama modeli ise (2.2) deki modelin basitleştirilmiş halidir. Tesadüfi katsayılı regresyon modelinde, 1. aşama parametreleri 2. aşama birimleri yani gruplara karşı değişir fakat bu değişim 2. aşama birimlerine ait özelliklere bağlı değildir tesadüfidir. 2. aşama model aşağıdaki verilmiştir:

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{2.9a}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{2.9b}$$

2. aşama tesadüfi etkilerden etkilerine ait varyans- kovaryans matrisi T ile ifade edilir ve aşağıdaki gibi gösterilir [9].

$$Var\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{bmatrix} = T \tag{2.10}$$

Var(u_{0j})= τ_{00} = 1. aşama sabit parametresi için koşulsuz varyans,

Var(u_{1j})= τ_{11}= 1. aşama eğim parametresi için koşulsuz varyans,

Cov(u_{0j} , u_{1j})= τ_{01}= 1. aşama eğim ve sabit parametreleri arasındaki koşulsuz kovaryanstır [9].

Bu varyanslar koşulsuz varyans olarak adlandırılırlar, çünkü 2. aşama tahmincileri ne (2.9a) ve (2.9b) eşitliklerinde yoktur. (2.9a) ve (2.9b)' yi (2.2) deki modelde yerine koyduğumuzda birleştirilmiş model:

$$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \overline{X}_{\bullet j}) + u_{0j} + u_{1j}(X_{ij} - \overline{X}_{\bullet j}) + r_{ij} \qquad (2.11)$$

şeklindedir. Bu modele göre bağımlı değişken, u_{0j} (j. birimin ortalama üzerindeki tesadüfi etkisi), u_{1j} 'nin β_{1j} eğim parametresi üzerindeki tesadüfi etkisi olduğu $u_{1j}(X_{ij} - \overline{X}_{\bullet j})$ ve 1. aşama hatası r_{iJ} şeklindeki üç bileşenden meydana gelen tesadüfi hata bileşenleri ile ortalama regresyon eşitliğinin ($\gamma_{00} + \gamma_{10}(X_{ij} - \overline{X}_{\bullet j})$ toplamının bir fonksiyonudur [9].

2.1.1.3 Sabit ve Eğim Parametresi Çıktı Olan Modeller

Tesadüfi etkili regresyon modeli, 2. aşama birimlerine karşılık gelen regresyon katsayılarındaki (eğim ve sabit) değişkenliği tahmin etmemize izin verir. Bundan sonraki adım ise; değişkenliğin modelinin kurulmasıdır. Bu model ise 2. aşama açıklayıcı değişkenleri ile kurulur. Yani, modelde sabit ve eğim parametresindeki değişkenliği açıklamak amacıyla 2.aşama açıklayıcı değişkenlerinden yararlanılır. Bu duruma dikkat çekmek amacıylada model, sabit ve eğim parametrelerinin "çıktı" olduğu modeller olarak adlandırılır. Bu modellere tam hiyerarşik lineer modeller de denilmektedir ve (2.2) ve (2.4) deki modeller geçerlidir [9].

Uygulamada bu modeller için önerilen kesin bir yöntem yoktur. Ama modellemeye çok aşamalı bir modele ihtiyaç duyulup duyulmadığını araştırmak için tesadüfi etkili ANOVA modeli ile başlanması önerilir. Daha sonraki adımda ileriye dönük yada geriye dönük adım adım yöntemlerle veriye en uygun model kurulabilir.

2.1.2 Çoklu Çok Aşamalı Regresyon Modeli

1. aşama regresyon modeli [9];

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_2 X_{2ij} + ... + \beta_{Qj} X_{Qij} + r_{ij}$$

$$= \beta_{0j} + \sum_{q=1}^{Q} \beta_{qj} X_{qij} + r_{ij} \qquad r_{ij} \sim N(0,\sigma^2) \tag{2.12}$$

şeklindedir. 2. aşama modelinde her bir β_{qj} aşağıdaki gibi modellenir (2. aşama açıklayıcı değişkenler W_{sj}=1,2,...,S_q olmak üzere)

$$\beta_{qj} = \gamma_{q0} + \gamma_{q1} W_{1j} + \gamma_{q2} W_{2j} + ... + \gamma_{qS_q} W_{S_q j} + u_{qj}$$

$$= \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qS} W_{Sj} + u_{qj} \tag{2.13}$$

Birleştirilmiş model ise (2.14)' da verilmiştir[9].

$$Y_{ij} = \gamma_{00} + \sum_{q=1}^{Q} \gamma_{q0} X_{qij} + \sum_{s=1}^{S_q} \gamma_{0s} W_{sj} + \sum_{s=1}^{S_q} \sum_{q=1}^{Q} \gamma_{qs} W_{sj} X_{qij} + u_{0j} + \sum_{q=1}^{Q} u_{qj} X_{qij} + r_{ij} \tag{2.14}$$

2.2 Çok Aşamalı Modeller İçin Tahmin Prensipleri

2 aşamalı çok aşamalı modellerde 3 tip parametre için tahmin yapılır. Bunlar; 2. aşama sabit etkileri için genelleştirilmiş en küçük kareler tahminleri, 1. aşama tesadüfi katsayıları için ampirik bayes tahminleri ve varyans-kovaryans bileşenleri için en çok olabilirlik tahminleridir[9].

2.2.1 Sabit Etkilerin Tahmini

i. Nokta tahmini

Q tane açıklayıcı değişkene sahip 1. aşama modeli matris notasyonu ile aşağıdaki gibi açıklanabilir.

$$Y_j = X_j \beta_j + r_j, \qquad r_j \sim N(0, \sigma^2 I) \tag{2.15}$$

$$Y_j = \begin{bmatrix} Y_{ij} \\ Y_{2j} \\ . \\ . \\ Y_{n_j j} \end{bmatrix},\ X_j = \begin{bmatrix} 1 & X_{1j} & \ldots & X_{1j,p-1} \\ 1 & X_{2j} & \ldots & X_{2j,p-1} \\ . & . & \ldots & . \\ . & . & \ldots & . \\ 1 & X_{n_j j} & .... & X_{n_j j,p-1} \end{bmatrix},\ \beta_j = \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \\ . \\ . \\ \beta_{n_{j-1} j} \end{bmatrix},\ r_j = \begin{bmatrix} r_{1j} \\ r_{2j} \\ . \\ . \\ r_{n_j j} \end{bmatrix}$$

Burada;

$Y_j : n_j \times 1$ boyutlu çıktı (bağımlı değişken) vektörü

$X_j :\ n_j \times (Q+1)$ boyutlu açıklayıcı değişken matrisi (bu matrisin ilk sütunu 1'lerden diğer sütunları x'lerden oluşur.)

$\beta_j :\ n_j \times 1$ boyutlu bilinmeyen parametre (katsayılar) vektörü

I : $n_j \times n_j$ boyutlu birim matris olarak tanımlanır.

r_j, $n_j \times 1$ boyutlu, 0 ortalama ile köşegen elemanları σ^2, diğer elemanları 0 olan varyans-kovaryans matrisine sahip ve normal dağıldığı varsayılan tesadüfi hatalara ait vektördür.

β_j' ye ait sıradan en küçük kareler tahmincisi $\hat{\beta}_j$ (katsayılar vektörü)'dir ve bu katsayılar

$$\hat{\beta}_j = (X_j^T X_j)^{-1} X_j^T Y_j \qquad (2.16)$$

ile bulunur [9].

Ve $\hat{\beta}_j$'ya ait dağılım matrisi; (parametre tahminlerinin parametre etrafındaki dağılım matrisi) yani $\hat{\beta}_j$'ya ilişkin varyans-kovaryans matrisi $Var(\hat{\beta}_j)$ yardımı ile hesaplanır.

$$Var(\hat{\beta}_j) = V_j = \sigma^2 (X_j^T X_j)^{-1} \qquad (2.17)$$

(2.16) da (2.17) yi yerine koyduğumuzda 1. aşama modeli aşağıdaki gibi olur.

$\hat{\beta}_j = \beta_j + e_j$, ej~N(0, V_j) (2.18)

V_j, hata varyans matrisidir ve β_j' yi tahmin eden $\hat{\beta}_j$'nin dağılım matrisinden meydana gelir.

β_j için 2. aşama genel modeli:

$$\beta_j = W_j\gamma + u_j, \qquad u_j \sim N(0, T) \tag{2.19}$$

olarak verilebilir.

$$w_j' = (1, W_{j1}, ..., W_{j,Q-1}),\ \gamma_p = (\gamma_{0p}, \gamma_{1p}, \ldots, \gamma_{Q-1,p})$$

$$W_j = \begin{bmatrix} w_j' & 0 & ... & 0 \\ 0 & w_j' & ... & 0 \\ . & . & ... & . \\ . & . & ... & . \\ 0 & 0 & ... & w_j' \end{bmatrix}_{P\times PQ}, \quad \gamma = \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ . \\ . \\ \gamma_{P-1} \end{bmatrix}_{PQ\times 1},$$

$$\mathrm{var}(u_j) = T = \begin{bmatrix} \tau_{00} & \tau_{01} & \cdots & \tau_{0,p-1} \\ \tau_{10} & \tau_{11} & \cdots & \tau_{1,p-1} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ \tau_{p-1,0} & \tau_{p-1,1} & \cdots & \tau_{p-1,p-1} \end{bmatrix}$$

(2.19)' u (2.18)' de yerine koyduğumuzda birleştirilmiş model;

$$\hat{\beta}_j = W_j\gamma + u_j + e_j \tag{2.20}$$

olarak yazılır [11].

W_j verildiği durumda $\hat{\beta}_j$'nin dağılımı yani $\hat{\beta}_j$'ye ilişkin koşullu varyans-kovaryans matrisi

$$\mathrm{Var}(\hat{\beta}_j) = \mathrm{Var}(u_j + e_j) = T + V_j = \Delta_J \tag{2.21}$$

= parametre varyansı (dağılımı)+ hata varyansı (dağılımı- dispersion)

şeklinde yazılır.

- Eğer her grup eşit sayıda gözleme sahipse ve β_j' nin her bir bileşenini için 2. aşama açıklayıcı değişkenleri aynı değerleri alıyorsa, her $\hat{\beta}_j$ de aynı dağılıma (Δ) sahip olur.

$$\Delta = T + V = T + \sigma^2 (X^T X)^{-1} \tag{2.22}$$

Bu durumda γ için tek, minumum varyanslı, yansız sıradan en küçük kareler regresyon tahmincisi (2.23) deki gibidir.

$$\hat{\gamma} = (\sum W_j^T W_j)^{-1} \sum W_j^T \hat{\beta}_j \quad (2.23)$$

- Eğer her gruptaki örnek büyüklükleri farklı ise, Δ_J gruptan gruba farklılık gösterir ve her Δ_J'nin bilindiği varsayımı altında γ'nın tek, minumum varyanslı, yansız genelleştirilmiş en küçük kareler kareler tahmincisi;

$$\hat{\gamma} = (\sum W_j^T \Delta_j^{-1} W_j)^{-1} \sum W_j^T \Delta_j^{-1} \hat{\beta}_j \quad (2.24)$$

ile hesaplanır.

Genelleştirilmiş en küçük kareler tahmincisi, ağırlıkları her gruptaki verinin kesinliğine (Δ_j^{-1}) (precision) yani varyans-kovaryans matrisinin tersi olan matrise göre belirler. (2.16) ve (2.19)' daki eşitlikler için normallik varsayımı geçerli olduğu durumda (2.24) deki γ' nın en çok olabilirlik tahminidir [9].

ii. Aralık tahmini

γ için güven aralığı γ'nın dağılım matrisi $V_{\hat{\gamma}}$ ye bağlıdır.

$$V_{\hat{\gamma}} = Var(\hat{\gamma}) = (\sum W_j^T \Delta_j^{-1} W_j)^{-1} \quad (2.25)$$

Örneğin, belirli bir γ_h elemanı için %95 güven aralığı;

$$Var(\gamma_h) \pm 1{,}96 (V_{hh})^{1/2} \quad (2.26)$$

V_{hh}, $V_{\hat{\gamma}}$ matrisinin h. köşegen elemanıdır.

Tüm 1. aşama katsayılarının tesadüfi olması ve her 2. aşama biriminin en küçük kareler tahmini ($\hat{\beta}_j$,2.17) için yeterli örneği içermesi durumunda $\hat{\gamma}$'nın formülü (2.24) ve $V_{\hat{\gamma}}$'nin formülü (2.25) geçerlidir. Fakat bazı 1. aşama değişkenleri sabit etkilere sahipken, tesadüfi etkilere sahip olmayabilir. (2.27) deki modelde bazı 1. aşama değişkenleri sabit etkilere sahip fakat tesadüfi etkilere sahip değildir.

$$Y_j = X_j W_j \gamma + Z_j u_j + r_j \quad (2.27)$$

Z_j, X_j'nin alt kümesidir ve Z_j'nin sütunları tesadüfi etkilere sahip 1. aşama değişkenlerinden meydana gelirken, X_j'nin sütunlarında ise sadece sabit etkiler sahip 1. aşama değişkenleri vardır.

Y_j' nin marjinal varyansı;

$$Var(Y_j) = V_{yj} = Z_j T Z_j^T + \sigma^2 I_j \tag{2.28}$$

şeklindedir.

I , (n_{j*} n_j) boyutunda birim matristir. Sabit etkiler için genelleştirilmiş en küçük kareler tahmini ise aşağıdaki hale dönüşür.

$$\hat{\gamma} = (\sum_{j=1}^{J} W_j^T X_j^T V_{yj}^{-1} X_j W_j)^{-1} \sum_{j=1}^{J} W_j^T X_j^T V_{yj} - Y_j \tag{2.29}$$

Varyans- kovaryans matrisi ise;

$$Var(\hat{\gamma}) = (\sum_{j=1}^{J} W_j^T X_j^T V_{yj}^{-1} X_j W_j)^{-1} \tag{2.30}$$

ile hesaplanabilir [9].

Eğer Z_j=X_j ise

$$X_j^T V_{yj}^{-1} X_j = \Delta_j^{-1} \tag{2.31}$$

eşitliği geçerli olacak ve (2.30) daki eşitlik (2.31) deki tüm 1. aşama değişkenlerinin tesadüfi etkilere sahip olduğu hale dönüşecektir [9].

2.2.2 1. Aşama Tesadüfi Katsayılarının Tahmini

1. aşama modellerinde, çıktılar 2. aşama birimlerine karşı değişen katsayılara bağlıdır. 2. aşama birimleri olan gruplara göre değişen 1. aşama katsayılarının tahmini bu bölümde açıklanacaktır.

i. Nokta tahminleri

β_j için en iyi tahmincileri bulmak amacıyla çekilme (shrinkage) prensiplerini (2.16) ve (2.19) daki eşitlikler yani tam hiyerarşik lineer modeller için genelleştirebiliriz. Yine β_j

tahmini için iki alternatif vardır. Birinci tahminci, j. grup verisine esas alan basit sıradan en küçük kareler tahmincisidir.($\hat{\beta}_j$) Tam hiyerarşik lineer modeller için (2.32) deki eşitlik yardımı ile hesaplanabilir [9].

$$\hat{\beta}_j = (X_j^T X_j)^{-1} X_j^T Y_j \tag{2.32}$$

İkinci tahminci, W_j deki grup karakterlerinin verildiği durumda, β_j' ye ait açıklayıcı değerdir.

$$\hat{\hat{\beta}}_j = W_j \hat{\gamma} \tag{2.33}$$

$\hat{\gamma}$, genelleştirilmiş en küçük kareler yöntemi ile tahmin edilmiştir. Bu iki tahmincinin optimal kombinasyonu ise

$$\beta_j^* = \Lambda_j \hat{\beta}_j + (I - \Lambda_j) W_j \hat{\gamma} \tag{2.34}$$

yardımı ile hesaplanır ve formülde yer alan Λ_j ise;

$$\Lambda_j = T(T + V_j)^{-1} \tag{2.35}$$

ile elde edilir. Λ_j, β_j için parametre dağılım matrsinin, toplam dağılım matrisine oranıdır. Λ_j, çok değişkenli güvenirlik matrisi olarak adlandırılır [9].

ii. Birinci aşama sıradan en küçük kareler katsayılarının güvenirliği

τ_{qq} ve v_{qqj} olarak gösterilen T ve V_j 'nin köşegen elemanları, sırasıyla Q+1 tane sıradan en küçük kareler 1. aşama katsayılarının güvenirliğinin hesaplanması için kullanılır.

$$güvenirlik(\hat{B}_{qj}) = \tau_{qq} / (\tau_{qq} + v_{qqj}),\ q=0,1,2,...,Q \tag{2.36}$$

$\hat{\beta}_{qj}$'nin örnek varyansı v_{qqj}, J birim arasında farklılık gösterebilir, her bir 2. aşama biriminin güvenirliği farklı indise sahiptir. J tane 2. aşama birimlerine karşılık toplam güvenirlik (2.37) daki gibi hesaplanabilir [9].

$$güvenirlik(\hat{B}_q) = \frac{1}{J} \sum_{j=1}^{J} \tau_{qq} / (\tau_{qq} + v_{qqj}),\quad q=0,1,2...,Q \tag{2.37}$$

iii. Kalıntı tahmini

2. aşama kalıntılarının sıradan en küçük kareler tahmini, 1. aşama tesadüfi parametreleri β_j'lerin sıradan en küçük kareler tahmininin bir fonksiyonudur.

$$\hat{u}_j = \hat{\beta}_j - W_j\hat{\gamma} \tag{2.38}$$

Ampirik Bayes kalıntıları her j. birim için;

$$u_j^* = \beta_j^* - W_j\hat{\gamma} \tag{2.39}$$

şeklindedir.

2. aşama kalıntılarına benzer olarak, 1. aşama kalıntıları da hem en küçük kareler hem de ampirik Bayes yöntemleri ile tahmin edilebilir. Bu tahminler 1. aşama verilerinin dağılım varsayımlarının kontrolü için gereklidir[9].

(2.40) daki genel modele göre en küçük kareler kalıntılarının tahmini;

$$\hat{r}_{ij} = Y_{ij} - X_{ij}^T W_j\hat{\gamma} - Z_{ij}^T\hat{u}_j \tag{2.40}$$

şeklindedir. X_{ij}^T ve Z_{ij}^T sırasıyla X_j ve Z_j nin i. satır elemanlarını içerir. Her birim için ampirik Bayes kalıntı tahmini ise aynı forma sahiptir;

$$r^*_{ij} = Y_{ij} - X_{ij}^T W_j\hat{\gamma} - Z_{ij}^T u^*_j \tag{2.41}$$

iv. Aralık tahmini

β_j için güven aralığı, koşullu varyans-kovaryans matrisi V_j^* ye göre oluşturulur.

$$V_j^* = (V_j^{-1} + T^{-1})^{-1} + (1-\Lambda_j)[Var(W_j\hat{\gamma})](I-\Lambda_j)^T \tag{2.42}$$

Herhangi bir β_{qj} için, 95% güven aralığı;

$$\beta^*_{qj} \pm 1{,}96(V^*_{qq_j})^{1/2} \tag{2.43}$$

şeklinde verilir. $V^*_{qq_j}$, V_j^*'nin q. köşegen elemanıdır. (3.30) daki eşitliğin geçerli olması için; β_j'nin tüm elemanlarının tesadüfi olması ve her j grubun, sıradan en küçük kareler ortalamaları ile $\hat{\beta}_j$ hesaplamalarına yetecek veriye sahip olması gereklidir [9].

Tüm β_j elemanlarının tesadüfi olduğunu fakat 2. aşama j. biriminin , β_j 'nin sıradan en küçük kareler tahmini için yeterli veriye sahip olmadığını farz edelim. Bu durumda da, Ampirik Bayes tahmincisi (2.44) de verildiği gibi hesaplanır [9]:

$$\beta_j^* = (X_j^T X_j + \sigma^2 T^{-1})^{-1}(X_j^T Y_j + \sigma^2 T^{-1} W_j \hat{\gamma}) \qquad (2.44)$$

2.2.3 Varyans- Kovaryans Bileşenlerinin Tahmini

Yij'lerin normal dağıldığı iki aşamalı modellerde, bilinmeyen üç tip parametre vardır. Bunlar; 1. aşama tesadüfi parametreleri β lar, 2. aşama sabit parametreleri γ lar ve 3 varyans kovaryans parametreleridir. Geçmişte sadece gruplardaki birim sayısı eşit olduğu durumlarda varyans kovaryans bileşenleri için tahmin yapılabiliyorken, günümüzde, gruplardaki örnek büyüklükleri farklı olduğu durumlarda iteratif sayısal prosedürlere verimli tahminler yapılabilmektedir.(en çok olabilirlik yöntemi gibi) Bu durum da, hiyerarşik lineer modellerin kullanımı arttırmıştır [9].

Varyans kovaryans parametreleri de, 1. aşama parametrelerinin varyans kovaryans matrisi T ve 1. aşama kalıntı varyansı σ^2 den meydana gelir. Bu parametrelerin tahminlerine ulaşabilmek için iki tip genel stratejiden yaralanılabilir. Bu stratejilerden birincisi, γ, T ve σ^2'nin tam en çok olabilirlik tahmini (TEÇOT yani EÇOT) ve β'nın ampirik Bayes (AB) tahminine dayanmaktadır. İkincisi ise, T ve σ^2'nin sınırlandırılmış en çok olabilirlik tahmini (SEÇOT) ve γ, β'nın ise ampirik Bayes tahminine dayanmaktadır [12].

2.2.4 Tam En Çok Olabilirlik Yöntemi

En çok olabilirlik yönteminin temel fikri, ana kütle Y ve ana kütleden çekilen örnek arasındaki benzerlik ilişkisinden yararlanarak bu örneğin elde edilme ihtimalinin maksimum yapan γ, T ve σ^2 parametrelerine ait tahmini seçmektir. γ, T ve σ^2 için EÇO tahminlerinin verildiği durumda, 1.aşama katsayıları β'lar için ampirik Bayes tahminlerini hesaplayabiliriz.

Tahminin dayandığı yaklaşım bazı özelliklere sahiptir. Bunlar:

i. Genel koşullar altında, γ, T ve σ^2'nin EÇO tahminleri tutarlıdır: 2. aşama örnek büyüklüğü J arttığında, EÇO tahminleri gerçek parametre değerlerine yakınsar.

Bu da, EO tahminlerinin asimptotik olarak yansız olması yani, sapmaların J arttığında ortadan kaybolmasını beraberinde getirmektedir. (Yeterli veriye sahip olunduğunda, bu tahminler yüksek olasılıkla gerçek parametrelere yaklaşırlar.)

ii. Bu tahminler asimptotik etkinliğe sahiptirler.(Veriye ait büyük örnekler verildiğinde, en çok olabilirlik tahminleri minimum varyans ile yansızdırlar.)

iii. Büyük J için, bu tahminlerin örnek dağılımı yaklaşık olarak normaldir bu da T'nin EÇO tahminlerinin pozitif tanımlı olmasını sağlar.

iv. Büyük J nin varlığında, 1. aşama katsayıları için Ampirik Bayes tahminleri ortalama için alternatif metotlarla yapılan tahminlerden daha doğru olma eğilimindedirler. (Daha düşük ortalama hata karelerine sahiptirler.) Ampirik Bayes tahminlerinin avantajı γ, T ve σ^{2}'nin EÇO tahminlerine bağlı olmalarındandır [13].

Tüm bu açıklanan özellikler, 2. aşama birimin sayısı J'nin büyük olması durumunda geçerlidir. J yani 2.aşama birim sayısı (grup sayısı) küçük olduğu durumlarda ise aşağıda sıralanan durumlar meydana gelebilir:

i. γ ve T tahminleri küçük örneklerde hatalı olabilir. Ayrıca, her iki tahmin ve özellikle T deki varyans elemanları normal dağılıma sahip olmayabilirler. Bu durum istatistiksel çıkarsamalarda birçok hataya neden olabilir

ii. Eğer veri dengesiz ise, γ tahminleri ağırlıklandırılır, yani T'nin ML tahminlerinin fonksiyonları olurlar. Bu ağırlıklardaki tesadüfi hatalar , γ da belirsizliklere neden olur ve bu durumda EÇO yaklaşımını kullanan yazılımlar standart hataları hesaplamayabilir. Bu yüzden de γ için güven aralıkları olması gerektiğinden daha dar, anlamlılık testleri yanıltıcı olabilir. Küçük j verildiği durumda, çok fazla dengesiz veriler çok ciddi problemlere neden olur, çünkü veri setleri dengesiz olduğunda ağırlıklar oldukça değişken olur. Yardımcı veriler hakkındaki belirsizlik (T) odak parametreler(γ) hakkında yalnış çıkarsamalar üretebilir.

iii. T hakkındaki çıkarsamalar γ hakkındaki belirsizlikleri göz önünde bulundururlar. Çünkü β'nın AB tahminleri , γ, T ve σ^{2}'nin EÇO tahminlerine bağlıdır. Bu durumda, β için güven aralıkları daralır ve β için anlamlılık testleri yanıltıcı olabilir.

Tersi durumda, σ^2 hakkındaki EÇO çıkarsamaları genellikle küçük J yada dengesiz verilere karşı duyarlı değildir. Çünkü σ^2 tahmininin doğruluğu, 1. aşama birimlerinin toplam örnek genişliğine bağlıdır [14]. ($N = \sum n_j$)

Tam en çok olabilirlik yöntemi ile (2.45) deki en çok olabilirlik fonksiyonunu maksimum yapacak σ^2 ve T tahmin edilmeye çalışılır.

$$l_{ML}(\sigma^2, T) = -\frac{1}{2}\log|V| - \frac{N}{2}\log r^T V^{-1} r - \frac{N}{2}\left[1 + \log\frac{2\pi}{N}\right] \qquad (2.45)$$

ve fonksiyondaki $r = Y - A(A^T V^{-1} A)^T A^T V^{-1} Y$'dir.

$$A_j = X_j W_j = \begin{bmatrix} 1 & W_j & X_{1j} & \cdots & W_j X_{1j} \\ 1 & W_j & X_{2j} & \cdots & W_j X_{2j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & W_j & X_{n_j j} & \cdots & W_j X_{n_j j} \end{bmatrix}$$

Deviyans testinden yararlanılmak isteniyorsa TEÇO, SEÇO yöntemine göre daha çok tercih edilir [8].

2.2.5 Sınırlandırılmış En Çok Olabilirlik Yöntemi (Restricted Maximum Likelihood-REML)

TEÇO'nin eksik yanı, varyans ve kovaryans tahminlerinin sabit etkilerin nokta tahminine bağımlı olmasıdır. Basit regresyon modeli;

$$Y_j = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_Q X_{Qi} + r_i$$

i=1,2,...,n olmak üzere r_i, hatalar 0 ortalama ve σ^2 varyans ile normal dağılmaktadır. Q+1 tane regresyon katsayısı $\beta_0, \beta_1, ..., \beta_Q$ nun bilindiğini varsayalım. Daha sonra σ^2 ye ait en çok olabilirlik tahmini;

$$\sigma^2 = \sum r_i^2 / n \qquad (2.46)$$

şeklinde hesaplanabilir. Regresyon katsayılarının bilinmediğini ve tahmin edileceğini varsayalım. Kalıntılar;

$$\hat{r}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - .... - \beta_Q X_{Qi} \qquad (2.47)$$

şeklindedir.

$\hat{\beta}_q$ sıradan en küçük kareler tahminidir. σ^2' ye ait sıradan yansız tahminci ise

$$\sigma^2 = \sum \hat{r}_i^2 /(n-Q-1) \qquad (2.48)$$

ile hesaplanır.

Paydadaki (n-Q-1), (q+1) tane regresyon parametresi tahmininde kullanılan serbestlik derecesi düzeltmesidir. q küçük olduğunda, düzeltmenin çok küçük etkisi olur. Ortalama hata karelerini minimize edecek düzeltme , (n-q-1) den çok (n-Q+1) dir. Q büyüdüğünde , (3.68) deki eşitliğin kullanımı σ^2 nin tahmininde ciddi sapmalara sahip olur. Sapma negatif olabilir, bu yüzden σ^2 tahmini oldukça küçük olacaktır, güven aralığı daralacaktır ve bu da yanıltıcı hipotez testlerine neden olacaktır. (2.46) ve (2.48) arasındaki fark TEÇO ve SEÇO arasındaki kesin farkı göstermektedir. RML, FML'nin γ ya göre integralinin alınmış halidir. Hiyerarşik lineer modeller için, TEÇO yada SEÇO yöntemine göre yapılmış varyans- kovaryans tahminleri arasındaki fark basit cebirsel formlarla açıklanamaz. Varyans bileşenleri için SEÇO tahminleri sabit etkiler hakkındaki belirsizlikleri düzeltir fakat TEÇO düzeltmez. İki aşamalı model için TEÇO ve SEÇO σ^2 için, birbirine yakın sonuçlar verirken T'nin tahmininde önemli farklar ortaya çıkar. 2. aşama birimlerinin sayısı J büyük olduğunda, iki yöntem oldukça yakın sonuçlar verirler. J küçük olduğunda, TEÇO varyans tahminleri $\hat{\tau}_{qq}$, SEÇO' dan daha küçük olur.

Sınırlandırılmış en çok olabilirlik yöntemi ile (2.49) daki en çok olabilirlik fonksiyonunu maksimum yapacak σ^2 ve T tahmin edilmeye çalışılır.

$$l_{ML}(\sigma^2, T) = -\frac{1}{2}\log|V| - \frac{1}{2}\log\left|A^T V^{-1} A\right| - \frac{(N-p)}{2}\log r^T V^{-1} r - \frac{(N-p)}{2}\left[1 + \log\frac{2\pi}{(N-p)}\right] \qquad (2.49)$$

ve fonksiyondaki $r = Y - A(A^T V^{-1} A)^T A^T V^{-1} Y$ ve $p = rank(A)$'dır.

2.2.6 EM Algoritması Kullanılarak Yapılan ML Tahminleri

Büyük örneklemlerde, varyans kovaryans bileşenleri tahminleri için en çok olabilirlik ve sınırlandırılmış en çok olabilirlik yöntemi birbirine yakın sonuçlar verir. Örneklem küçük

olduğunda ise, en çok olabilirlik fonksiyonunu maksimize edecek tahminleri bulmak için bazı yardımcı algoritmalara ihtiyaç duyulabilir. 2 aşamalı çok aşamalı modellerde EO tahminleri için genellikle EM (Expectation-Maximization) algoritması tahmin edilmektedir. EM algoritması, en çok olabilirlik tahminlerini hesaplarken 2. aşama tesadüfi parametrelerini kayıp veri olarak adlandırır [15].

Kayıp veri, u_j nin gözlendiği varsayımı altında, en çok benzerlik yöntemi ile tüm parametre (γ, T, σ^2) tahminlerinin hesaplanması daha kolay olacaktır. γ tahmini, birleştirilmiş modelin ($Y_j = \gamma + u_j + r_{ij}$) her iki tarafından u_j nin çıkarılması ile elde edilebilir. Daha sonra γ için sıradan en küçük kareler tahmini aşağıdaki hali alır; [16]

$$\hat{\gamma} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}(Y_{ij} - u_j) / \sum_{j=1}^{n_j} n_j \tag{2.50}$$

varyans bileşenleri için en çok olabilirlik tahminleri daha basittir;

$$\hat{\tau} = \sum_{j=1}^{J} u_j^2 / J \tag{2.51}$$

$$\hat{\sigma}^2 = \sum_{j=1}^{J}\sum_{i=1}^{n_j}(Y_{ij} - \hat{\gamma} - u_j)^2 / \sum_{j=1}^{J} n_j \tag{2.52}$$

2. aşama tesadüfi parametrelerinin kayıp veri olarak alındığı durumda, parametre tahmini için kullanılacak istatistikler " tamamlanmış veri yeterli istatistikleri (complete data sufficient statistics " olarak adlandırılır. Bu değerler, Y verildiği durumda tahmin edilebilir ve parametre ilk tahminleridir. Y verildiği durumda, her u_j nin, u_j^* ortalama ve V_j^* varyans ile normal dağıldığı varsayımı altında, parametre ilk tahminleri ($\gamma^{(0)}, \tau^{(0)}, \sigma^{2(0)}$) üst indis $^{(0)}$ ile gösterilirler.

$$\begin{aligned} u_j^* &= \lambda_j^{(0)}(\bar{Y}_j - \gamma^{(0)}), \\ \mathrm{V_j^*} &= \tau^{(0)}(1 - \lambda_j^{(0)}) \end{aligned} \tag{2.53}$$

Bu tahminler aşağıdaki gibi elde edilebilir.

$$E[\sum_{j=1}^{J}\sum_{i=1}^{n_j}(Y_{ij} - u_j) \Big| Y, \gamma^{(0)}, \tau^{(0)}, \sigma^{2(0)}] = \sum_{j=1}^{J}\sum_{i=1}^{n_j}(Y_{ij} - u_j^*) \tag{2.54}$$

ve

$$E[\sum_{j=1}^{J} u_j^2 \Big| Y, \gamma^{(0)}, \tau^{(0)}, \sigma^{2(0)}] = \sum_{j=1}^{J} (u_j^{*2} + V_j^*),$$
$$E[\sum_{j=1}^{J}\sum_{i=1}^{n_j} (Y_{ij} - \hat{\gamma} - u_j)^2 \Big| Y, \gamma^{(0)}, \tau^{(0)}, \sigma^{2(0)}] = \sum_{j=1}^{J}\sum_{i=1}^{n_j} [(Y_{ij} - \not{\gamma}^{(0)} - u_j^*)^2 + V_j^* \quad (2.55)$$

Algoritmadaki "E adımında", Y verisi ve parametre tahminleri verildiği durumdaki şartlı beklenen değer aracılığıyla parametrelere ait ilk tahminler yapılır. (2.55) " M adımında" ise E adımında elde edilen tahminler, tam verinin verildiği durumdaki EÇO tahminleri için kullanılan formüllerde yerine konulur. Dempster, Laird, ve Rubin her bir EM algoritmasının olasılıkları büyüttüğünü ve bununda tahminleri lokal maksimuma yakınsattığını ispatlamışlardır.

EÇO-EM yaklaşımının dezavantajı, odak parametreler hakkındaki çıkarsamaların bilinmeyen parametreler hakkındaki belirsizliği hesaba katmamasıdır. J küçük ve 1.aşama verileri oldukça dengesiz olduğu zaman bu problem çok önemlidir. SEÇO yaklaşımı kısmen bu sorunu çözebilir [17].

i. En çok olabilirliğe bağlı olarak varyans bileşenleri için güven aralıkları

T deki her bir varyans ve kovaryans bileşeni için standart hata tahmini hesaplanabilir. En çok olabilirlik yöntemi ile elde edilen bu tahminlerin kullanılmasında bazı problemler vardır. Problem, varyans tahminlerinin ($\hat{\tau}_{qq}$) örneklem dağılımının bilinmeyen bir derecede çarpıklığa sahip olmasıdır. Sonuç olarak simetrik güven aralıkları ve hipotez testleri oldukça yanıltıcı olabilir [18].

TEÇO ve SEÇO yöntemlerin deki eksiklikler, varyans ve kovaryans tahminleri için Bayes yaklaşımının kullanılması ile düzeltilebilir. Güven aralıkları ve hipotez testleri ile sabit etkiler hakkında sonuç çıkarma, varyans ve kovaryans parametrelerinin nokta tahminlerinin doğruluğuna bağlıdır. Örneğin sabit parametrenin çıktı olduğu modeldeki $\hat{\gamma}_{01}$ için 95% güven aralığı;

$$\hat{\gamma}_{01} \pm 1{,}96(V_{\hat{\gamma}_{01}})^{1/2} \quad (2.56)$$

şeklindedir ve eşitlikteki

$$V_{\hat{\gamma}_{01}} = Var(\hat{\gamma}_{01}) = [\sum \Lambda_j^{-1}(w_j - \overline{W}_\bullet^*)^2]^{-1} \quad (2.57)$$

ile hesaplanabilir.

τ_{00} yalnış tahmin edildiyse, güven aralıkları da kusurlu olacaktır. Fotiu's simülasyonu t değerinin çarpan olarak kullanıldığı güven aralıklarının daha uygun olduğunu bulmuştur. Yani güven aralıkları

$\hat{\gamma}_{01} \pm t_{0,025}(\hat{V}_{\hat{\gamma}_{01}})^{1/2} V_{\hat{\gamma}_{01}}$ şeklini alır.

2.3 Çok Aşamalı Modeller İçin Hipotez Testi

3. bölümde, Y_{ij}'lerin normal dağıldığı iki aşamalı modeller için bilinmeyen üç tip parametre 1. aşama tesadüfi parametreleri β lar, 2. aşama sabit parametreleri γ' lar ve varyans kovaryans parametreleri için tahmin prosedürlerini açıklanmıştır. Bu bölümde ise iki aşamalı modeller için hipotez testleri ele alınacaktır.

1. aşama regresyon modeli;

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_2 X_{2ij} + ... + \beta_{Qj} X_{Qij} + r_{ij} \quad (2.58)$$

$$= \beta_{0j} + \sum_{q=1}^{Q} \beta_{qj} X_{qij} + r_{ij} \qquad r_{ij} \sim N(0,\sigma^2)$$

şeklinde gösterilir.

(2.58) deki eşitlikte bazıları sabit, bazıları tesadüfi bazıları ise tesadüfi olmayan Q+1 tane parametre vardır. 2. aşama modelinde her bir β_{qj} aşağıdaki gibi modellenir (2. aşama açıklayıcı değişkenler W_{sj}=1,2,...,S_q olmak üzere)

$$\beta_{qj} = \gamma_{q0} + \gamma_{q1} W_{1j} + \gamma_{q2} W_{2j} + ... + \gamma_{qS_q} W_{S_q j} + u_{qj} \quad (2.59)$$

$$= \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qS} W_{Sj} + u_{qj}$$

Birleştirilmiş model ise (2.60) da verilmiştir.

$$Y_{ij} = \gamma_{00} + \sum_{q=1}^{Q} \gamma_{q0} X_{qij} + \sum_{s=1}^{S_q} \gamma_{0s} W_{sj} + \sum_{s=1}^{S_q} \sum_{q=1}^{Q} \gamma_{qs} W_{sj} X_{qij} + u_{0j} + \sum_{q=1}^{Q} u_{qj} X_{qij} + r_{ij} \quad (2.60)$$

Çok aşamalı modeller için test edilebilir hipotez tipleri Çizlege 2.1 de verilmiştir [9].

Çizelge 2.1 Çok aşamalı modeller için test edilebilir hipotez tipleri

Hipotez Tipleri	**Sabit Etkiler**	**1.Aşama Tesadüfi Katsayıları**	**Varyans Bileşenleri**
Tek Parametre için			
H_0	γ_h =0	β_{qj} =0	$\tau_{qq} = 0$
H_1	$\gamma_h \neq 0$	$\beta_{qj} \neq 0$	$\tau_{qq} \neq 0$
Birden fazla parametre için			
H_0	$C'\gamma$ =0	$C'\beta$ =0	T=T_0
H_1	$C'\gamma \neq 0$	$C'\beta \neq 0$	T≠ T_0

(2.59) da her β_{qj} için S_q+1 tane sabit etki vardır ve 2. aşama modeli için toplam tesadüfi etki $F = \sum_q (S_q + 1)$'dir. Her bir tesadüfi bileşen u_{qj} nin çok değişkenli normal dağılıma sahip olduğu varsayımı altında;

Var(u_{qj})= τ_{qq} (2.61)

Ve $Cov(u_{qj}, u_{q'j}) = \tau_{qq'}$'dir. (2.62)

Sabit etkiler (her bir γ_{qs}), 1. aşama tesadüfi katsayıları (her β_{qj}) ve varyans-kovaryans parametreleri için hipotez testleri formüle edilebilir. Bu testler tek parametre yada birden fazla parametre için olabilir. Çizelge (2.1) de, 6 çeşit hipotez tipi vardır. Bunlara karşılık gelen test istatistikleri ise Çizelge (2.2) de verilmiştir [9].

Çizelge 2.2 Hiyerarşik lineer modeller için genel hipotez testleri

Hipotez Tipleri	**Sabit Etkiler**	**1. Aşama Tesadüfi Katsayıları**	**Varyans Bileşenleri**
Tek Parametre için	t oranı	t oranı	Tek değişkenli X^2 yada z oranı
Birden Fazla Parametre için	Genel lineer hipotez testleri	Genel lineer hipotez testleri	Olabilirlik oran testi(X^2)

2.4 Sabit Etkiler İçin Hipotez Testleri

İ .Tek bir parametre için hipotez testleri:

H_0 hipotezi (4.6) daki gibidir:

$$H_0 : \gamma_{qs} = 0 \tag{2.63}$$

γ_{qs}'in 0 olması, 2. aşama açıklayıcı değişkenlerinin (Wsj), 2. aşama parametreleri üzerinde etkisini gösteren, katsayıların geçersiz olduğu anlamına gelir. Bu tip hipotezler için test istatistiği

$$t = \hat{\gamma}_{qs} / (\hat{V}_{\hat{\gamma}_{qs}})^{1/2} \tag{2.64}$$

ile hesaplanabilir. $\hat{\gamma}_{qs}$, γ_{qs}'in en çok olabilirlik tahmini, $\hat{V}_{\hat{\gamma}_{qs}}$ ise $\hat{\gamma}_{qs}$'in tahmin edilen örnek varyansıdır. Ve (J-S_q-1) serbestlik derecesine sahiptir.

ii. Birden fazla parametre için hipotez testleri:

j. gruptaki i. birim için bağımlı değişken, X_{ij}'nin bir fonksiyonu olarak aşağıdaki gibi yazılabilir:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \tag{2.65}$$

2.aşama modeli ise (2.66)' da verilmiştir:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{oj} \tag{2.66a}$$

$$\beta_{10} = \gamma_{10} + \gamma_{11} W_j + u_{1j} \tag{2.66b}$$

Eşitlikte, γ_{01}, γ_{11} in 0'a eşit olduğunu varsayan H_0 hipotezi test edilebilir. Alternatif hipotez ise W_j' nin modele dahil edilmesi gerektiğini ileri sürer.

Birden fazla parametrenin birlikte testi için (2.66)' daki eşitliğin matris notasyonu ile tekrar formüle edilmesine gerekir [9].

$$\begin{pmatrix} \beta_{0j} \\ \beta_{1j} \end{pmatrix} = \begin{pmatrix} 1 & W_j & 0 & 0 \\ 0 & 0 & 1 & W_j \end{pmatrix} \begin{pmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{11} \end{pmatrix} + \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \tag{2.67}$$

yani $\beta_j = W_j \gamma + u_j$ şeklinde gösterilebilir.

Bileşik hipotez testi ise;

$$H_0 : C^T \gamma = 0, \tag{2.68}$$

şeklindedir. $C^T = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ olmak üzere göre H_0 hipotezi (4.2) de verilmiştir.

$$H_0 : C^T \gamma = \begin{pmatrix} \gamma_{01} \\ \gamma_{11} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2.69}$$

$Var(\hat{\beta}_j) = \Lambda_j$ bilindiği durumda $\hat{\gamma}$ örnek varyansı:

$$Var(\hat{\gamma}) = (\sum W_j^T \Lambda_j^{-1} W_j)^{-1} = V_{\hat{\gamma}} \tag{2.70}$$

olarak yazılabilir.

Contrast vektör $C'\hat{\gamma}$ 'nın varyansı ise:

$$Var(C^T \hat{\gamma}) = C^T V_{\hat{\gamma}} C = (say) V_c \tag{2.71}$$

$V_{\hat{\gamma}}$ bilinmiyorsa tahmini aşağıdaki gibi yapılır;

$$\hat{V}_{\hat{\gamma}} = (\sum W_j^T \hat{\Lambda}_j^{-1} W_j)^{-1} \tag{2.72}$$

$H_0 = C^T \gamma = 0,$ hipotezi için, test edilen C^T'nin satır sayısı eşit serbestlik derecesinde X^2 dağılımına sahip yaklaşık test istatistiği:

$$H = \hat{\gamma}^T C \hat{\mathrm{V}}_{\mathrm{c}}^{-1} C^T \hat{\gamma} \qquad (2.73)$$

ile hesaplanabilir[9].

Buradaki örnekte C^T, 2 satıra sahiptir yani serbestlik derecesi 2'dir.

Birden fazla parametrenin birlikte test edilebilmesi için diğer bir yaklaşım ise olabilirlik oran (likelihood-ratio test- deviyans testi) testidir. Bu test ancak TEÇO tahminleri elde edildiğinde kullanılabilir. Test, iki modelden elde edilen tahmin ve sonuçların karşılaştırılmasına dayanmaktadır. Birinci model sabit etkilerin 0 olduğu modeldir ve H_0 hipotezini meydana getirir. Alternatif model ise, tüm etkileri içermektedir ve H_1 hipotezini meydana getirir. Her model için deviyans istatistikleri hesaplanır ve çok değişkenli hipotez testi için kullanılır. Deviyans modelin uygunluk ölçüsü olarak da kullanılır: yüksek deviyans, modelin veriye uygun olmadığını göstergesidir [9].

Birinci model için deviyans istatistiği $D_0 = -2\log\left(L_0\right)$ (2.74)

şeklindedir. L_0 ise, sıfır hipotezindeki en çok olabilirlik tahminleri için olabilirlik değeridir. D_1 ise, alternatif modeldeki en çok olabilirlik tahminleri için olabilirlik değeridir:

$$D_1 = -2\log\left(L_1\right) \qquad (2.75)$$

Sıfır hipotezi altında, deviyans farkı $D_0 - D_1$, tahmin edilen parametre sayıları arasındaki farka eşit serbestlik dereceli χ^2 dağılımına sahiptir. Büyük deviyans farkı, veri için birinci modelin çok basit olduğunu gösterir. Daha önce anlattığımız genel lineer hipotez testleri deviyans testine göre bazı avantajlara sahiptir. Bunlar[9]:

i. Genel lineer hipotez testleri kullanılırken sadece bir model (alternatif model) için tahminlerin hesaplaması yeterlidir. Oysa deviyans testinde tüm modeller için tahminlerin yapılmasını gerektirir.

ii. 2. genel lineer hipotez testleri ile parametreler arasındaki farklılıklar test edilebilir oysa deviyans testinin bu tür bir test için yeniden düzenlemesi gereklidir.

iii. SEÇO yöntemi kullanıldığında, deviyans testinden yararlanılamaz.

Sabit etkiler hakkındaki hipotez testleri için en çok olabilirlik oran testine başvurulduğunda, varyans kovaryans bileşenlerinin spesifikasyonları, indirgenmiş ve alternatif model için aynı olmalıdır. Bu test, sadece modeldeki varyans kovaryans elemanlarının testi için değil sabit etkilerle varyans kovaryans elemanlarının birlikte testi içinde kullanılabilir [9].

2.5 Birinci Aşama Tesadüfi Parametreleri İçin Hipotez Testleri

Tesadüfi parametre vektöründeki 2 yada daha fazla bileşenin karşılaştırılması tek parametre ve birden fazla parametre için hipotez testleri olmak üzere iki başlık altında ele alınabilir.

i. Tek bir parametre için hipotez testleri:

Tek bir gruba özel regresyon katsayısına ilişkin hipotez;

$$H_0 : \beta_{qj} = 0 \qquad (4.19)$$

olarak yazılabilir.

Bu hipotez tahmin edilen parametrenin kendi standart hatasına bölünmesi ile elde edilen z istatistiği ile sınanır. Tahmin olarak araştırmacı ampirik Bayes tahminini yada sıradan en küçük kareler tahminini kullanabilir.

Ampirik Bayes tahmini β_{qj}

$$z = \beta^*_{qj} / V^{*1/2}_{qqj} \qquad (2.76)$$

ile test edilir. Eşitlikte;

β^*_{qj}; ampirik Bayes tahmini,

$V^{*1/2}_{qqj}$, β^*_{qj} katsayılarına ait artçıl varyans kovaryans matrisinin (posterior dispersion) q. köşegen elemanıdır ve z yaklaşık olarak normal dağılmaktadır.

SEÇOY'de TEÇOY'e göre artçıl varyans (posterior varyans) daha büyük ve gerçeğe daha yakındır. Bu durum özellikle 2.aşama birim sayısı küçük olduğu zaman geçerlidir. TEÇOY de sabit etkilerin bilindiği varsayılır, bu yüzdende TEÇO sabit etkilerle ilgili herhangi bir

belirsizlik yansıtmaz. Bu durum ancak 2. aşama birimleri, J'ler yeterince büyük olduğu zaman geçerlidir.

β tahminleri sıradan en küçük kareler tahmincileri ile hesaplandığında, t testleri yanıltıcı olabilir.

ii. Birden fazla parametre için hipotez testleri:

β, [J(Q+1)$\times$1] boyutlu tesadüfi parametre vektörü β için hipotez;

$$H_0 : C^T \beta = 0 \tag{2.77}$$

şeklinde oluşturulabilir.

Eğer β'lar ampirik bayes yöntemi ile tahmin ediliyorsa, test istatistiği

$$H_{EB} = \beta^{*T} C (C^T V^* C)^{-1} C^T \beta^* \tag{2.78}$$

olarak yazılabilir.

V^*, parametrelere ait [J(q+1)*J(Q+1)[boyutlu varyans-kovaryans matrisidir.

Eğer β'lar sıradan en küçük kareler yöntemi ile tahmin ediliyorsa, test istatistiği (2.77) de verilen genel lineer hipotez istatistik ortalamaları ile hesaplanabilir [9].

$$H_{EKK} = \hat{\beta}^T C (C^T \hat{V} C)^{-1} C^T \hat{\beta} \tag{2.79}$$

$\hat{V}$, blok diyagonal matristir ve (Q+1)*(Q+1) için blok:

$$\hat{V}_j = \hat{\sigma}^2 (X_j^T X_j)^{-1} \tag{2.80}$$

ile hesaplanabilir[9].

2.6 Varyans ve Kovaryans Bileşenleri İçin Hipotez Testi

i. Tek bir parametre için hipotez testleri:

Hiyerarşik analizlerde, araştırmacılar, 1. aşama parametrelerinin sabit mi, tesadüfi mi ya da tesadüfi olarak değişmeyen yapıda mı olduğuna varyans bileşenleri için yapılan hipotez testleri yardımı ile karar verilebilir[9].

Tesadüfi varyasyonlar için test edeceğimiz hipotez:

$H_o : \tau_{qq} = 0$ dir. (2.81)

şeklinde kurulabilir ve $\tau_{qq} = Var(\beta_{qj})$

Eğer sıfır hipotezi red edilirse, araştırmacılar β_q da tesadüfi varyasyon olduğuna karar verirler. Tüm gruplar (en azından büyük bir kısmı) sıradan en küçük kareler tahminleri için yeterli veriye sahipse, en basit ve gerekli test, $H_o : \tau_{qq} = 0$ hipotezinin testidir. $\hat{V}_{qqj}$, $\hat{V}_j = \hat{\sigma}^2 (X_j^T X_j)^{-1}$ matrisinin q. diyagonal elemanıdır.

$$\beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj} \quad (2.82)$$

modeli altında, (J-Sq-1) serbestlik derecesi ile yaklaşık χ^2 dağılan istatistik

$$\sum_j (\hat{\beta}_{qj} - \hat{\gamma}_{q0} - \sum_{s=1}^{S_q} \hat{\gamma}_{qs} W_{sj})^2 / \hat{V}_{qqj} \quad (2.83)$$

şeklinde hesaplanır.

$H_o = \tau_{qq} = 0$ hipotezi, tahminin standart hatasını temel alan z istatistiği ile de test edilebilir.

$$z = \hat{\tau}_{qq} / [Var(\hat{\tau}_{qq})]^{1/2} \quad (2.84)$$

Bu oran, en çok olabilirlik tahminlerinin büyük örnek teorisi altında yaklaşık normal dağılmaktadır. Çoğu zaman, özellikle τ_{qq} sıfıra çok yakın olduğunda, normallik yaklaşımı oldukça zayıflar. Teknik olarak, en çok olabilirlik, mod etrafında simetrik değildir (asimptotik normallik başarılamamıştır) ve simetrik güven aralıklarına dayalı testlerde yanıltıcı olabilir [19].

ii. Birden fazla parametre için hipotez testleri:

Varyans ve kovaryans için genel hipotez testinin temelini, en çok olabilirlik oran testi oluşturmaktadır. Bu test, MLF ve MLR tahminleri altında elde edilebilir. Bu test için Ho hipotezi;

$$H_0 : T = T_0 \quad (2.85)$$

Alternatifi ise;

$H_1 : T = T_1$ dir. (2.86)

To, T_1'in indirgenmiş halidir. Örneğin, T_1 deki q.satır ve sütunun (bazı satır ve sütunların) geçersiz olması gibi. Yazılan hipotezin testi, ilk olarak iki model için tahminleri yapar, D_0 ve D_1 deviyanslarını hesaplar ve sonra test istatistiğini hesaplar:

$H = D_0 - D_1$ (2.87)

Bu istatistik m serbestlik derecesi ile yaklaşık χ^2 dağılır. m, iki modelden elde edilen varyans ve kovaryans tahminlerinin sayısı arasındaki farktır. Varyans ve kovaryanslar için en çok olabilirlik oran testi kullanıldığında, karşılaştırılacak modeller, sabit etkilerin spesifikasyonu bakımından aynı olmalıdır. (Aynı açıklayıcı değişkenlere sahip olmalıdır.)

BÖLÜM 3

İKİ AŞAMALI LOJİSTİK REGRESYON ANALİZİ

Çok aşamalı lojistik regresyon modelinde, iki ya da ikiden çok düzey içeren kesikli bağımlı değişken varlığında, birey veya birim aşamasından başlayarak, birimler, birimlerin meydana getirdiği gruplar, grupların meydana getirdiği topluluklar vb. şeklinde devam eden hiyerarşik yapının her aşaması için ayrı ayrı model kurulur. Daha sonra bu modeller birleştirilerek birleşik modele ulaşılır [7].

İki aşamalı modelde, 1. aşama birimleri "birimler", 2. aşama birimleri ise "gruplar" olarak tanımlanabilir. İki aşamalı lojistik regresyon analizinde, birinci aşamada bireyler ikinci aşamada ise bireylerin içinde bulunduğu gruplar için ayrı ayrı modeller kurulur. $i = 1,2,.....,n_j$ olmak üzere i indisi 1. aşama birimlerini; $j = 1,2,.....,J$ olmak üzere j indisi ise 2. aşama birimlerini yani grupları göstermektedir. n_j, j. gruptaki birim sayısını, J ise grup sayısını gösterir. Gruplardaki birim sayıları farklı olabilir ($j \neq k$ için $n_j = n_k$ olmak zorunda değildir) [9].

3.1 1. Aşama Modelleri

İki aşamalı lojistik regresyon analizinde, 1. aşama modeli:

i. Y bağımlı değişkenini tanımlayan ve Y'nin bir olasılık fonksiyonuna sahip olduğunu varsayan tesadüfi bileşen yani 1. aşama örnekleme modeli (level-1 sampling model)

ii. Modelin tahmincisi olarak kullandığımız bağımsız değişkenleri ifade eden sistematik bileşen yani 1. aşama yapısal modeli (level-1 structural model) ve

iii. Tesadüfi bileşenin beklenen değeri ile sistematik bileşen arasındaki fonksiyonel ilişkiyi gösteren bağlantı bileşeni yani 1. aşama bağlantı fonksiyonu (level-1 link function)

olmak üzere üç bileşenden oluşur[9].

3.1.1 1. Aşama Örnekleme Modeli

Y_{ij}, m_{ij} denemedeki başarı sayısı ve φ_{ij} de her denemedeki başarı olasılığı olmak üzere 1.aşama örnekleme modeli aşağıdaki gibi yazılabilir:

$$Y_{ij}\left|\varphi_{ij}\right. \sim \mathbf{B}inom(m_{ij},\varphi_{ij})$$

Yani ,Y_{ij}, m_{ij} deneme sayısı ve φ_{ij} de her denemedeki başarı olasılığı ile binom dağılımına sahiptir. Y_{ij} binom raslantı değişkeninin beklenen değeri olasılık dağılımının ortalaması $E(Y_{ij}\left|\varphi_{ij}\right.)=m_{ij}\varphi_{ij}$ ve varyansı $Var(Y_{ij}\left|\varphi_{ij}\right.)=m_{ij}\varphi_{ij}(1-\varphi_{ij})$'dir. Bernoulli raslantı değişkeni, Binom raslantı değişkeninin $m_{ij}=1$ için özel bir halidir [9].

3.1.2 1. Aşama Bağlantı Fonksiyonu

1. aşama örnekleme modeli binomial olduğunda, sistematik bileşenle tesadüfi bileşen arasındaki bağlantı fonksiyonu

$$\eta_{ij}=\log\left(\frac{\varphi_{ij}}{1-\varphi_{ij}}\right)$$

şeklinde yazılabilir. Bu tür fonksiyonlara logit fonksiyon denir. Yani model bir odds değerinin logaritmasıdır. Odds değeri $\left(odds=\frac{\varphi_{ij}}{1-\varphi_{ij}}\right)$, bir sonucun gerçekleşme olasılığının, gerçekleşmeme olasılığına oranıdır.

Sistematik bileşenle tesadüfi bileşen arasındaki bağlantı fonksiyonu bağımlı değişkenin sayma, ikili, sıralı veya çokterimli olmasına göre farklılıklar gösterir. İki aşamalı lojistik regreyon modelleri Çizelge 3.1'de verilmiştir [9].

Çizelge 3.1 İki Aşamalı Lojistik Regresyon Modelleri

Yanıt Değişkeni	1. Aşama Örnekleme Modeli	1. Aşama Beklenen Çıktı	1. Aşama Bağ Fonksiyonu
Sürekli $-\infty < \gamma_{ij} < \infty$	Normal	μ_{ij}	$\eta_{ij}=\mu_{ij}$ (identity)
İkili $Y_{ij} \in \{0,1\}$	Bernoulli	φ_{ij}	$\eta_{ij}=\log[\varphi_{ij}/(1-\varphi_{ij})]$ (logit)
Sayma $Y_{ij} \in \{0,1, \ldots M_{ij}\}$	Binomial	$m_{ij}\varphi_{ij}$	$\eta_{ij}=\log[\varphi_{ij}/(1-\varphi_{ij})]$ (logit)
Sayma $Y_{ij} \in \{0,1, \ldots\}$	Poisson	λ_{ij}	$\eta_{ij}=\log(\lambda_{ij})$ (log)
Sıralı Kategori $R_{ij} \in \{1, 2, \ldots, M\}$	Multinomial	φ^{*}_{mij}	$\eta^{*}_{mij}=[\varphi^{*}_{mij}/(1-\varphi^{*}_{mij})]$ (cumulative logit)
Nominal Kategori $R_{ij} \in \{1, 2, \ldots, M\}$	Multinomial	φ_{mij}	$\eta_{mij}=[\varphi_{mij}/(\varphi_{Mij})]$ (multinomial logit)

Tezde uygulama ile aynı doğrultuda olması açısından logit model üzerinden açıklamalar yapılmıştır.

3.1.3 1. Aşama Yapısal Modeli

İki değerli bağımlı değişkene sahip bir olasılık modeli logit model olarak aşağıdaki gibi ifade edilebilir:

$$\varphi_{ij} = \frac{1}{1+e^{-\eta_{ij}}} = \left[1+\exp\left(-\beta_0 - \sum_{j=1}^{m}\beta_j X_{ij}\right)\right]^{-1} \qquad (3.1)$$

Dolayısıyla; $\varphi_{ij} = \Pr\left(Y_{ij}=1\right)$ olarak ifade edilirse lojistik regresyon modeli

$$\eta_{ij} = \log\left(\frac{\varphi_{ij}}{1-\varphi_{ij}}\right) = \beta_{0j} + \sum_{q=1}^{Q}\beta_{qj}X_{qij} \qquad (3.2)$$

eşitliği ile verilebilir. 1. aşama modeli matris notasyonu ile de $\eta_j = X_j\beta_j$ şeklinde gösterilebilir [9].

(3.1) ve (3.2) numaralı eşitliklerde β_{qj}'ler 1.aşama katsayıları, X_{qij}'ler ise 1. aşama açıklayıcı değişkenleridir. $(q = 0,1,\ldots,Q)$ Lojistik regresyon analizinden farklı olarak, 1. aşama modelinde her gruba ait eğim ve sabit parametreleri farklılık göstermektedir. Bu yüzden çok aşamalı lojistik regresyon modelinin ikinci adımını, (2. aşama) grup

seviyesinde tanımlanan açıklayıcı değişken ile regresyon katsayılarındaki (β_j) değişimlerin tahmin edilmesi oluşturur [9].

3.2 2. Aşama Modelleri

Lojistik regresyon analizinden farklı olarak, 1. aşama modelinde her gruba ait eğim ve sabit parametreleri farklılık göstermektedir. Bu yüzden çok aşamalı lojistik regresyon modelinin ikinci adımını, (2. aşama) grup seviyesinde tanımlanan açıklayıcı değişken ile regresyon katsayılarındaki (β_j) değişimlerin tahmin edilmesi oluşturur.

$$\beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj} + u_{qj} \qquad (3.3)$$

(3.3) numaralı eşitlikte γ_{qs} 2. aşama katsayıları, W_{sj} grup aşaması (2. aşama) açıklayıcı değişkenleri ve u_{qj} ise 2. aşama tesadüfi etkisini göstermektedir. $(q = 0,1,....,Q)$

Çok aşamalı doğrusal modellerde olduğu gibi, u_{qj} 2. aşama tesadüfi etkilerinin 0 ortalama ve $Var(u_{qj}) = \tau_{qq}$ ile çok değişkenli normal dağılıma sahip oldukları varsayılır. Herhangi iki tesadüfi etki arasındaki kovaryans ise $Cov(u_{qj}, u'_{qj}) = \tau_{qq'}$'dir [9].

2. aşama açıklayıcı değişkenleri ile yapılan modellerden sonra, 1. aşama katsayıları β_{qj}'ler gruplar arası değişimlerden arındırılmış olur. 2. aşama modeli matris notasyonu ile $\beta_j = W_j \gamma_j + U_j$ şeklinde yazılabilir.

2. aşama modelinin 1. aşama modelinde yerine konması ile elde edilen çok aşamalı birleştirilmiş lojistik regresyon model ise $\eta_j = X_j W_j \gamma_j + X_j U_j$ şeklinde ifade edilebilir[9].

3.3 İki Aşamalı Lojistik Regresyon Analizinde Parametre Tahmini

Çok aşamalı lojistik regresyon modelinde parametre tahminlerinin bulunmasında kullanılacak yöntemler arasında Penalized Quasi-Likelihood yönteminin en iyi sonuç verdiği görülmüştür. Bu modelde maksimum benzerlik tahminleri Fisher'ın skorlama algoritması ile bulunabilmektedir.

Penalized Quasi-Likelihood (PQL)

PQL yöntemi doğrusal olmayan regresyon modellerinde kullanılabilir.

1. aşama modeli:

$$Y_{ij} = \varphi_{ij} + e_{ij} \tag{3.4}$$

$$E(e_{ij}) = 0 \qquad \text{var}(e_{ij}) = \varphi_{ij}(1-\varphi_{ij})$$

şeklindedir ve doğrusallaştırılan bu model birinci sıradan s. iterasyon için Taylor serisine açıldığında

$$\varphi_{ij} \approx \varphi_{ij}^{(s)} + \frac{d\varphi_{ij}}{d\eta_{ij}}(\eta_{ij} - \eta_{ij}^{(s)}) \tag{3.5}$$

elde edilir ve eşitlikte $\frac{d\varphi_{ij}}{d\eta_{ij}} = \varphi_{ij}(1-\varphi_{ij}) = w_{ij}$ 'ye eşittir.

(3.5) nolu denklemi (3.4) de yerine koyduğumuzda

$$\frac{Y_{ij} - \varphi_{ij}^{(s)}}{V_{ij}^{(s)}} + \eta_{ij}^{(s)} = \eta_{ij} + \frac{e_{ij}}{V_{ij}^{(s)}} * \tag{3.6}$$

elde edilir. (3.6) eşitliğini $\frac{Y_{ij} - \varphi_{ij}^{(s)}}{V_{ij}^{(s)}} = Y_{ij}^{*(s)}$ ve $\frac{e_{ij}}{V_{ij}^{(s)}} = \varepsilon_{ij}$ olacak şekilde yeniden düzenlediğimizde $Y_{ij}^{*(s)} = \eta_{ij} + \varepsilon_{ij}$ şeklinde sıradan 2 aşamalı hiyerarşik lineer model denklemi elde ederiz[9].

Bu denklemdeki 2. aşama varyans-kovaryans bileşenleri ve sabit etkilerin maksimum benzerlik tahmininde Fisher algoritmasından yararlanılır.

Bu algoritmadan yararlanarak 2. aşama birimlerine göre modeli

$d_j = Z_j U_j + \varepsilon_{ij}$, $V_j = Z_j' T \sigma_j^2 Z$ ve $Y_j^{*(s)} \sim N_{nj}(X_j\gamma_j, V_j)$ olmak üzere

$$Y_{ij}^{*(s)} = X_j\gamma_j + d_j, \qquad d_j \sim N(0, V_j) \tag{3.7}$$

şeklinde ifade edebiliriz.

$$\log[L(Y_j^{*(s)}; \gamma_j, T, \sigma_v^2] = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|V_j| - \frac{1}{2}d_j' V_j d_j \tag{3.8}$$

şeklindeki log benzerlik fonksiyonunun çözümü istenen maksimum benzerlik tahminlerini verecektir [9].

3.3.1 Varyans-Kovaryans Bileşenlerinin (Tesadüfi Etkiler) Tahmini

Varyans kovaryans bileşenleri için Fisher algoritması [9]

$$\omega^{(s+1)} = \omega^{(s)} + [Q(\omega)]^{-1}[M(\omega)] \tag{3.9}$$

Şeklindedir ve eşitlikte φ $\mathrm{M}(\omega_j) = \frac{1}{2}\left[\frac{\partial vec(V_j)}{\partial \omega'}\right](V_j^{-1} \otimes V_j^{-1}) vec(d_j d_j' - V_j)$ (3.10)

ve

$$\mathrm{Q}(\omega_j) = \frac{1}{2}\left[\frac{\partial vec(V_j)}{\partial \omega'}\right](V_j^{-1} \otimes V_j^{-1})\frac{\partial vec(V_j)}{\partial \omega'} \tag{3.11}$$

3.3.2 Sabit Etkilerin Tahmini (γ)

Fisher algoritması sabit etkiler için

$$\gamma^{(s+1)} = \gamma^{(s)} + \left[X_j' V_j^{-1} X_j\right]^{-1}\left[X_j' V_j^{-1} d_j\right] \tag{3.12}$$

şeklindedir [9].

3.4 Hipotez Testleri

i.Sabit etkiler için hipotez testleri:

Sabit etkiler için tekil parametre testinde $H_0 : \gamma_{qs} = 0$ şeklinde, grup aşamasındaki açıklayıcı değişkenlerin 1. aşama parametreleri üzerinde etkisinin olmadığı anlamına gelen hipotez test edilebilir.

Test istatistiği $J - s_q - 1$ serbestlik derecesi ile

$$t = \gamma_{qs} / (\hat{V}_{(\hat{\gamma}_{qs})})^{1/2} \tag{3.13}$$

Şeklindedir [9].

ii. Varyans kovaryans bileşenleri için hipotez testleri:

Çok aşamalı lojistik regresyon analizinde 1. aşama parametrelerinin 2. aşamada sabit mi, tesadüfi mi yoksa tesadüfi olarak değişmeyen yapıda mı modelleneceğine yani 2. aşama model spesifikasyonuna varyans kovaryans bileşenleri için yapılan hipotez testleri ile karar verilebilir.

$\tau_{qq} = \mathrm{var}(\beta_{qj})$ olmak üzere hipotezimiz $H_o : \tau_{qq}$ şeklindedir.

H_o hipotezini $J - s_q - 1$ serbestlik derecesi ile yaklaşık χ^2 dağılan

$$\chi^2 = \left(\sum_j \left[\hat{\beta}_{qj} - \hat{\gamma}_{q0} - \sum_{s=1}^{S_q} \hat{\gamma}_{qs} W_{sj} \right] \right) \Big/ \hat{V}_{qqj} \qquad (3.14)$$

istatistiği ile test edebiliriz[9].

BÖLÜM 4

UYGULAMA

Uygulama Türkiye'de bankacılık sistemi içinde özel bir yere sahip kredi kartı kullanımı üzerine gerçekleştirilmiştir. Kredi kartı kullanımı günümüz koşullarında hızla artarak ve toplumu tüketime özelliklede karşılığını ödeyemeyecekleri tüketime teşvik etmektedir. Çalışmada Türkiye'de geniş bir müşteri portföyüne sahip özel bir bankanın müşterileri hakkında oluşturduğu veri tabanından yararlanılmıştır. Amaç Türkiye'de kredi kartı kullanan bireylerin temel özelliklerini ortaya çıkarmak ve bu temel özelliklerin kredi kartı kullanımı üzerindeki etkilerinin banka şubalerine göre farklılık gösterip göstermediğini ortaya koymaktır. Çok aşamalı lojistik regresyon analizinden yararlanarak, müşterilere ait karakteristiklerin modele dahil edilmesiyle elde edilen özel hiyerarşik modellerin karşılaştırılmasıyla kredi kartı kullanımının müşteriler ve banka şubeleri bazında açıklayacak en iyi model kurulmaya çalışılmıştır. Uygulamada HLM 7.03 programı kullanılmıştır. Uygulamaya ilk olarak koşulsuz model ile başlanılır.

4.1 Veri ve Değişkenlerin Tanımları

Özel bir bankanın rassal olarak seçilen 20 ildeki 65 banka şubesinden müşterilere ait olan veriler talep edilmiş bunlardan 14 il ve 62 banka şubesinden dönüş olmuştur. 3 banka şubesi 4 ya da daha az müşteriye ait dönüş olduğu için analiz dışında kalmıştır. Çünkü çok aşamalı yapı için gruplar içindeki birim sayısının minumum 5 olması gerekmektedir. Şubelerden minumum 5 maximum 15 müşteriye ait veri alınmıştır. Ortalama müşteri sayısı 8'dir. Çok aşamalı modellerde, gruplardaki birim sayısından çok grup sayısı önemlidir. Yapılan çalışmalar, standart hataları doğru tahmin edebilmek

için grup sayısının en az 50 grupladaki birim sayısının da en az 5 olması gerektiğini göstermektedir [20].

Veri seti banka şubeleri ve banka şubelerinin müşterilerden meydana gelen 2 aşamalı hiyerarşik bir yapıya sahiptir. Banka müşterileri 1. aşama birimlerini, 2. aşama birimlerini ise banka şubeleri meydana getirmektedir. 1. aşamada 399 banka müşterisi, 2. aşamada ise farklı illerde bulunan 59 banka şubesi bulunmaktadır. Müşterilerin % 78,2'si kredi kartı kullanmaktadır.

Çizelge 4.1 Kredi kartı kullanımı

	Frekans	Yüzde	Geçerli Yüzde	Kümülatif Yüzde
Hayır	87	21,8	21,8	21,8
Evet	312	78,2	78,2	100,0
Toplam	399	100,0	100,0	

Bağımlı değişkenimiz $y = \begin{cases} 1, \text{kredi kartı kullanmak} \\ 0, \text{kredi kartı kullanmamak} \end{cases}$, şeklinde değerler alan dikotom bir değişken olmak üzere uygulamada çok aşamalı lojistik regresyon analizinden yararlanılmıştır. Bireylerin cinsiyetleri, medeni durumları, meslekleri, gelirleri, yaşları ve sosyal güvenceleri 1. aşama açıklayıcı değişkenlerini oluşturmaktadır. 2. Aşamada ise bu değişkenlerin kredi kartı kullanımı üzerindeki etkisinin banka şubelerine göre rastsal olarak değiştiği bir model kurulmasına karar verilmiştir.

4.2 Betimsel Istatistikler

Veriler özel bir bankanın Türkiye'nin 14 farklı ilinde bulunan 59 şubesinin 399 müşterisine ait olan verilerdir. Şubelerin bulunduğu iller Çizelge 4.2' de verilmiştir. En fazla müşteri sırasıyla İstanbul, Ankara ve İzmir illerindendir.

Çizelge 4.2 Banka şubelerinin bulunduğu iller ve müşteri sayıları

	Frekans	Yüzde	Geçerli Yüzde	Kümülatif Yüzde
Adana	9	2,3	2,3	2,3
Ankara	99	24,8	24,8	27,1
Antalya	8	2,0	2,0	29,1
Bursa	18	4,5	4,5	33,6
Denizli	7	1,8	1,8	35,3
Diyarbakir	6	1,5	1,5	36,8
Erzurum	8	2,0	2,0	38,8
Gaziantep	6	1,5	1,5	40,4
İstanbul	151	37,8	37,8	78,2
İzmir	59	14,8	14,8	93,0
Kayseri	5	1,3	1,3	94,2
Konya	7	1,8	1,8	96,0
Samsun	8	2,0	2,0	98,0
Trabzon	8	2,0	2,0	100,0
Toplam	399	100,0	100,0	

Çalışma dahilindeki banka müşterilerin %79.9'u erkek, %20.1'i kadın müşterilerden meydana gelmektedir.

Çizelge 4.3 Banka müşterilerinin cinsiyete göre dağılımı

	Frekans	Yüzde	Geçerli Yüzde	Kümülatif Yüzde
Erkek	319	79,9	79,9	79,9
Kadın	80	20,1	20,1	100,0
Toplam	399	100,0	100,0	

Müşterilerin yaşlara göre dağılımına bakıldığında, %28,6 ile en fazla 35-44 yaş aralığında müşteri bulunduğu, %26,3 ile 25-34 yaş aralığındaki müşterilerin ikinci sırayı aldığı görülmektedir. 55 yaş ve üstü müşteriler ise %9,3 ile en küçük müşteri grubunu oluşturmaktadır.

Çizelge 4.4 Banka müşterilerinin yaşa göre dağılımı

	Frekans	Yüzde	Geçerli Yüzde	Kümülatif Yüzde
18 - 24	62	15,5	15,5	15,5
25 - 34	105	26,3	26,3	41,9
35 - 44	114	28,6	28,6	70,4
45 - 54	81	20,3	20,3	90,7
55 yaş ve üzeri	37	9,3	9,3	100,0
Toplam	399	100,0	100,0	

399 müşterinin %32,1'i bekar iken %67,9'unun evli olduğu görülmektedir.

Çizelge 4.5 Banka müşterilerinin medeni duruma göre dağılımı

	Frekans	Yüzde	Geçerli Yüzde	Kümülatif Yüzde
Bekar	**128**	**32,1**	**32,1**	**32,1**
Evli	**271**	**67,9**	**67,9**	**100,0**
Toplam	**399**	**100,0**	**100,0**	

Çizelge 4.6' da müşterilerin %67,4'ünün eğitim durumunun lise ve üniversite mezunlarından oluştuğu görülmektedir.

Çizelge 4.6 Banka müşterilerinin eğitim durumuna göre dağılımı

	Frekans	Yüzde	Geçerli Yüzde	Kümülatif Yüzde
İlköğretim / 1 yıl	2	0,5	0,5	0,5
İlköğretim / 2 - 4 yıl	1	0,3	0,3	0,8
İlköğretim / 5 yıl	39	9,8	9,8	10,5
Orta öğretim / 6 - 7 yıl	9	2,3	2,3	12,8
Orta öğretim / 8 yıl	33	8,3	8,3	21,1
Lise / 9 -10 yıl	15	3,8	3,8	24,8
Lise / 11 yıl	144	36,1	36,1	60,9
Üniversite / 12 - 14 yıl	18	4,5	4,5	65,4
Üniversite / 15 yıl	125	31,3	31,3	96,7
Lisansüstü / 17 yıl	13	3,3	3,3	100,0
Toplam	399	100,0	100,0	

Müşterilerin %17,3'lük bir bölümü küçük esnaflardan, %15'i nitelikli elemanlardan ve %13'lük bir bölümü masabaşı çalışan memurlardan oluşmaktadır. Tezin uygulamasında modelleme yapılırken meslek değişkeni işsiz olanlar 0, diğer meslek grupları 1 olarak kodlanmış ve meslek değişkeni olarak analizlere alınmıştır.

Çizelge 4.7 Banka müşterilerinin meslek durumuna göre dağılımı

	Frekans	Yüzde	Geçerli Yüzde	Kümülatif Yüzde
Ev Hanımı	10	2,5	2,5	2,5
Emekli	41	10,3	10,3	12,8
İşsiz	4	1,0	1,0	13,8
Öğrenci	23	5,8	5,8	19,5
Rant geliri ile geçinen	1	0,3	0,3	19,8
Nitelikli serbest meslek sahibi	12	3,0	3,0	22,8
0 – 5 çalışanlı tüccar	18	4,5	4,5	27,3
6 – 20 çalışanlı tüccar	5	1,3	1,3	28,6
1 – 9 çalışanlı şirket	4	1,0	1,0	29,6
10 – 25 çalışanlı şirket	1	0,3	0,3	29,8
25 + çalışanlı şirket	2	0,5	0,5	30,3
Büyük Esnaf	8	2,0	2,0	32,3
Küçük Esnaf	69	17,3	17,3	49,6
Çiftçi 5'den az çalışanlı	1	0,3	0,3	49,9
Fikir işiyle uğraşanlar	1	0,3	0,3	50,1
Sanatçılar	2	0,5	0,5	50,6
Özelde üst düzey yönetici	3	0,8	0,8	51,4
Kamuda üst düzey yönetici	3	0,8	0,8	52,1
Özelde 10'dan az çalışanlı yönetici	3	0,8	0,8	52,9
Kamuda 10'dan az çalışanlı yönetici	1	0,3	0,3	53,1
Kamuda 10'dan fazla çalışanlı yönetici	1	0,3	0,3	53,4
Nitelikli elemanlar	60	15,0	15,0	68,4
Memurlar(Masa başı çalışanlar)	52	13,0	13,0	81,5
Yardımcı hizmetliler	16	4,0	4,0	85,5
Fikir çalışanları	6	1,5	1,5	87,0
Vasıflı işçiler	46	11,5	11,5	98,5
Vasıfsız işçiler	6	1,5	1,5	100,0
Toplam	399	100,0	100,0	

Çalışma kapsamındaki banka müşterilerin %34,8'i kiracı, %63,7'si ev sahibidir.

Çizelge 4.8 banka müşterilerinin mülk sahipliğine durumuna göre dağılımı

	Frekans	Yüzde	Geçerli Yüzde	Kümülatif Yüzde
Kiracıyım / Kiracıyız	139	34,8	34,8	34,8
Ev sahibiyim / Ev sahibiyiz	254	63,7	63,7	98,5
Diğer	6	1,5	1,5	100,0
Toplam	399	100,0	100,0	

Çalışma kapsamındaki banka müşterilerinin %68,2'lik bölümünün 1001-3000 TL'lik gelire sahip olduğu görülmektedir. 0-1500 TL gelir sahipleri %43,6'lik bir bölümü oluşturmaktadır.

Çizelge 4.9 banka müşterilerinin aylık gelir miktarına göre dağılımı

	Frekans	Yüzde	Geçerli Yüzde	Kümülatif Yüzde
0 - 1500 TL	174	43,6	43,6	43,6
1501 - 2000 TL	89	22,3	22,3	65,9
2001 - 3000 TL	79	19,8	19,8	85,7
3001 - 5000 TL	36	9,0	9,0	94,7
5000 TL ve üzeri	21	5,3	5,3	100,0
Total	399	100,0	100,0	

Çalışma kapsamındaki müşterilerin %50,9 unun SSK'ya bağlı olduğu görülmekte, ikinci sırada %24,6 ile emekli sandığına bağlı olanlar yer alırken, %17,3 ile Bağ-Kur'lular azınlığı oluşturmaktadır. %6 lık bir dilimin sosyal güvencesi yoktur.

Çizelge 4.10 Banka müşterilerinin sosyal güvence sahibi olma durumuna göre dağılımı

	Frekans	Yüzde	Geçerli Yüzde	Kümülatif Yüzde
Sosyal güvencesi yok	24	6,0	6,0	6,0
SSK	203	50,9	50,9	56,9
Bağ - Kur	69	17,3	17,3	74,2
Emekli sandığı	98	24,6	24,6	98,7
Yeşil Kart	2	0,5	0,5	99,2
Özel sigorta	3	0,8	0,8	100,0
Total	399	100,0	100,0	

Kredi kartı kullanan müşterilerin %81,7'si erkek, %18,3'ü kadındır.

Çizelge 4.11 Kredi kartı kullanımının cinsiyete göre dağılımı

			Kredi kartı kullanıyor musunuz?		Toplam
			Hayır	Evet	
Cinsiyet	**Erkek**	Sayı	64	255	319
		Toplam %	73,6%	81,7%	79,9%
	Kadın	Sayı	23	57	80
		Toplam %	26,4%	18,3%	20,1%
	Toplam	Sayı	87	312	399
		Toplam %	100,0%	100,0%	100,0%

Kredi kartı kullanan müşterilerin %72,1'i evli iken, % 27,9'u bekar müşterilerden meydana gelmektedir.

Çizelge 4.12 Kredi kartı kullanımının medeni duruma göre dağılımı

			Kredi kartı kullanıyor musunuz?		Toplam
			Hayır	Evet	
Medeni durum	Bekar	Sayı	41	87	128
		Toplam %	47,1%	27,9%	32,1%
	Evli	Sayı	46	225	271
		Toplam %	52,9%	72,1%	67,9%
	Toplam	Sayı	87	312	399
		Toplam %	100,0%	100,0%	100,0%

Kredi kartı kullanan banka müşterilerinin %43,6'lik en büyük dilimi 1.500 TL'den az gelire sahiptir. Kredi kartı kulananların en küçük dilimi ise %6,1 ile 5.000TL ve üzerinde aylik gelire sahiptir.

Çizelge 4.13 Kredi kartı kullanımının ailenin aylık gelirine göre dağılımı

Ailenizin aylık toplam geliri?		Kredi kartı kullanıyor musunuz		Toplam
		Hayır	Evet	
0- 1500 YTL	Sayı	47	127	174
	Toplam %	54,02%	40,07%	43,6%
1501 - 2000 YTL	Sayı	19	70	89
	Toplam %	21,8%	22,4%	22,3%
2001 - 3000 YTL	Sayı	15	64	79
	Toplam %	17,2%	20,5%	19,8%
3001 - 5000 YTL	Sayı	4	32	36
	Toplam %	4,6%	10,3%	9,0%
5000 YTL ve üzeri	Sayı	2	19	21
	Toplam %	2,3%	6,1%	5,3%
Toplam	Sayı	87	312	399
	Toplam %	100,0%	100,0%	100,0%

4.3 İki Aşamalı Lojistik Regresyon Analizi

4.3.1 1.Model: Koşulsuz Model

Hiyerarşik lojistik regresyon modelinin özel bir hali olan koşulsuz model tesadüfi etkili tek yönlü ANOVA modeli olarak da adlandırılır. Burada koşulsuz kelimesi her iki aşamada da modele açıklayıcı değişken dahil edilmediğini vurgulamak için kullanılmıştır. Bu model ile amaç Türkiye'de kredi kartı kullanımının banka şubelerine arasında anlamlı farklılık gösterip göstermediğini ortaya çıkarmaktır.

1.aşama yani bireyler düzeyindeki model: $\eta_{ij} = \beta_{0j}$

2. aşama ya da iller düzeyindeki model ise $\beta_{0j} = \gamma_{00} + u_{0j}$; $u_{oj} \sim N(0, \tau_{oo})$

şeklindedir. Burada γ_{00}, banka şubeleri ait kredi kartı kullanımının ortalama log-odds değeridir. Tesadüfi etkili tek yönlü ANOVA modeline göre, kredi kartı kullanımının log-odds değeri 1,27 [$\hat{\gamma}_{00} =$ 1,27 (s.e= 0,132529, t= 9,598, p<0,001)]olarak tahmin edilmiştir.

Çizelge 4.14 Koşulsuz Model- tek yönlü ANOVA modeli sonuçları

Sabit etkiler	Katsayılar	Standart hata	t oranı	Serbestlik derecesi	p-değeri
γ_{00}	1,272040	0,132529	9,598	50	<0,001

Buna karşılık gelen odds değeri 3,56 ($e^{\eta_{ij}}$)ve karşılık gelen olasılık ise ($\frac{1}{1+e^{-\eta_{ij}}}$) 0,781'dir. Bu sonuca göre, bireylerin ortalama kredi kartı kullanma olasılığı 0,781'dir diyebiliriz. Koşulsuz modele göre, $\hat{\tau}_{00} = 0.14868$ (s.e=0,3855, $\chi^2 = 79.49$,p=0,005) olarak tahmin edilmiştir. $H_0 : \tau_\beta = 0$ hipotezi red edilir yani banka şubeleri arasında kredi kartı kullanımı açısından anlamlı farklılıklar vardır.

4.3.2 2. Model: Koşullu Model

Banka şubeleri arasında kredi kartı kullanımı açısından anlamlı farklılıklar olduğuna karar verdikten sonra, bu farklılıkları açıklamak amacıyla koşullu modellere geçeriz. Koşullu modelle kredi kartı kullanımın müşteriler açısından açıklayabilecek 1. aşama açıklayıcı değişkenlerini modele dahil ederiz. Yani müşterilerin yaşları, cinsiyetleri, meslekleri, medeni durumları, gelirleri ve sosyal güvence sahibi olma durumlarının kredi kartı kullanımı üzerinde etkili olup olmadığını araştırabiliriz.

1.aşama yani müşteriler düzeyindeki model:

$$\eta_{ij} = \beta_{0j} + \beta_{1j} cinsiyet + \beta_{2j} Med - Dur + \beta_{3j} Meslek + \beta_{4j} Gelir + \beta_{5j} konut_mul + \beta_{6j} sosyal_guv$$

Şeklinde kurulmuştur.

Çok aşamalı modellemenin varsayımlarından biri, 1. aşamadaki parametrelerin banka şubelerine göre farklılık göstermesidir. İllerdeki ortalama kredi kartı kullanımının log odds değeri ve 1. aşama açıklayıcı değişkenlerinin etkilerini 2. aşamada modelleriz.

Müşterilerin yaşları, cinsiyetleri, meslekleri, medeni durumları, gelirleri ve sosyal güvence sahibi olma durumlarının kredi kartı kullanımı üzerindeki etkilerinin banka şubelerine göre anlamlı farklılıklar gösterdiği;

2.aşama yada banka şubeleri düzeyindeki model ise:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + u_{2j}$$
$$\beta_{3j} = \gamma_{30} + u_{3j}$$
$$\beta_{4j} = \gamma_{40} + u_{4j}$$
$$\beta_{5j} = \gamma_{50} + u_{5j}$$
$$\beta_{6j} = \gamma_{60} + u_{6j}$$

şeklindedir.

Çizelge 4.14'te koşullu çok aşamalı lojistik regresyon modelinden elde edilen sonuçlar verilmiştir. Bu sonuçlara göre, müşterilerin kredi kartı kullanımını cinsiyetleri, medeni

durumları, meslekleri ve gelirleri etkilemektedir ($p_{\hat{\gamma}_{q0}} < 0,05$). Konut mülkiyeti ve sosyal güvence sahibi olmak ise müşterinin kredi kart kullanımı üzerinde etkili değildir.

Çizelge 4.15 Koşullu model için sonuçlar

Sabit etkiler	Katsayılar	Standart hata	t oranı	Serbestlik derecesi	p-değeri
Sabit γ_{00}	1.5638	0.5983	2.617	52	0.012
Cinsiyet γ_{10}	0.983145	0.214670	4.580	50	0.001
Med-Dur γ_{20}	0.920055	0.207569	4.433	50	0.001
Meslek γ_{30}	0.132948	0.0059870	2.224	50	0.031
Gelir γ_{40}	0.1833	0.0069	2.649	50	0.011
Konut Mul γ_{50}	0.169436	0.110544	1.533	50	0.132*
Sosyal_guv γ_{60}	0.210820	0.193287	1.091	50	0.281*

Modelin varyans-kovaryans bileşenleri incelendiğinde ise (Çizelge 4.16), eklenen 1. Aşama açıklayıcı değişkenlerinin kredi kartı kullanımı üzerinde etkisi olmadığı gibi aynı zamanda kredi kartı kullanımının banka şubelerine göre görülen anlamlı farklılığı da bozmuştur.

Çizelge 4.16 Koşullu model için varyans-kovaryans bileşen tahminleri

Sabit etkiler	Katsayılar	Standart hata	χ^2 değeri	Serbestlik derecesi	p-değeri
Sabit u_{00}	1,98421	3,9371	38,31515	32	>0,500
Cinsiyet u_{10}	0,56280	0,31675	19,81018	32	>0,500
Med-Dur u_{20}	0,23971	0,05746	50,032264	32	>0,500
Meslek u_{30}	0,22127	0,04896	38,14013	32	>0,500
Gelir u_{40}	0,3101	0,09617	32,77968	32	>0,500
Konut Mul u_{50}	0,22673	0,015141	40,68531	32	>0,500
Sosyal_guv u_{60}	0,28335	0,08029	25,737248	32	>0,500

Modelde anlamsız olan iki değişken modelden çıkarılarak koşullu model yeniden kurulur.

4.3.3 3. Model (Koşullu Model)

1. aşama modeli:

1. aşama yani müşteriler düzeyindeki model;

$$\eta_{ij} = \beta_{0j} + \beta_{1j} cinsiyet + \beta_{2j} Med - Dur + \beta_{3j} Meslek + \beta_{4j} Gelir$$

2.aşama yada banka şubeleri düzeyindeki model ise;

$\beta_{0j} = \gamma_{00} + u_{0j}$
$\beta_{1j} = \gamma_{10}$
$\beta_{2j} = \gamma_{20}$
$\beta_{3j} = \gamma_{30}$
$\beta_{4j} = \gamma_{40}$

Şeklinde kurulur.

Analiz sonuçlarına bakıldığında, kredi kartı kullanımını müşterilerin cinsiyetleri, medeni durumları, meslekleri ve gelirleri etkilemektedir.

Çizelge 4.17 Koşullu son model için sonuçlar

Sabit etkiler	Katsayılar	Standart hata	t oranı	Serbestlik derecesi	p-değeri
Sabit γ_{00}	1,40187	0,592779	2,2364	50	0,022
Cinsiyet γ_{10}	0,501381	0,27789	1,804	50	0,004
Med-Dur γ_{20}	0,95331	0,215084	4,432	50	0,001
Meslek γ_{30}	0,178753	0,058701	3,045	50	0,004
Gelir γ_{40}	0,196813	0,070578	2,789	50	0,007

2.aşama için ortalama kredi kartı kullanımı banka şubelerine göre farklılık gösterdiği fakat cinsiyet, medeni durum, meslek ve gelirin kredi kartı kullanımı üzerindeki etkisinin banka şubelerine göre farklılık göstermediği model anlamlı çıkmıştır.

Çizelge 4.18 Koşullu son model için varyans-kovaryans bileşen tahminleri

Sabit etkiler	Katsayılar	Standart hata	χ^2 değeri	Serbestlik derecesi	p-değeri
Sabit u_{00}	0,28553	0,1853	73,17	50	0,000

Modele göre, $\hat{\gamma}_{10} = 0,50$ olarak tahmin edilmiştir. Bu durum diğer koşullar sabit kabul edildiğinde, bireylerin cinsiyeti ile kredi kartı kullanımları arasında pozitif yönlü bir ilişki olduğunu göstermektedir. Erkeklerin kredi kartı kullanımlarının odds değeri, kadınların 1,648 katıdır [$e^{0.50} = 1,648$].

Çok aşamalı lojistik regresyon modelinden elde edilen sonuçlara göre evli bireylerin bekarlara göre kredi kartı kullanma olasılığı 2,52 kat daha fazladır [$\hat{\gamma}_{30} = 0,92331, p = 0,00, e^{0.95331} = 2.52$].

Tabloya bakıldığında kredi kartı sahibi olma ile meslek arasında pozitif yönlü bir ilişki olduğu görülmektedir [$\hat{\gamma}_{30} = 0,17; p = 0.004$, $e^{0.17} = 1,18530$].

Geliri yüksek olan müşterilerin kredi kartı kullanma olasılığı da yüksektir [$\hat{\gamma}_{40} = 0,196813; p = 0.007; e^{0.19} = 1.2175$].

Çok aşamalı lojistik regresyon modeline göre ortalama kredi kartı kullanma olasılığı banka şubelerine göre anlamlı farklılıklar göstermektedir. Müşterilerin kredi kartı kullanmalarının log odds değeri $\hat{\gamma}_{00} = 1,4187$ olarak tahmin edilmiştir ve karşılık gelen olasılık ise $\frac{1}{1+e^{-1.4187}} = 0,8051$'dir.

Yapılan çok aşamalı modeller sonucunda, Tesadüfi etkili ANOVA Modeli (1.model-koşulsuz model) ile, Müşterilerin ortalama kredi kartlı kullanma olasılıklarının banka şubelerine göre anlamlı farklılıklar gösterdiğine karar verilmiştir. Müşterilerin kredi kartı kullanmalarında etkili olan faktörler ve bu faktörlerin etkilerinin banka şubelerine

göre farklıklar gösterip göstermediği koşullu modellerde araştırılmıştır. Koşullu modeller sonucunda ise, kredi kartı kullanımını müşterilerin cinsiyetleri, medeni durumları, meslekleri ve gelirleri etkilediği bulunmuştur. Bu değişkenlerin etkileri banka şubelerine göre tesadüfi olarak değişmemekte sadece ortalama kredi kartı kullanma olasılığı banka şubelerine göre anlamlı farklılıklar göstermektedir.

BÖLÜM 5

SONUÇ VE ÖNERİLER

Çok aşamalı lojistik regresyon modelinde, bağımlı değişken en düşük aşamadan (birimler aşamasından) seçilen iki yada daha çok kategoriye sahip bir kesikli değişkendir. Bağımsız değişkenler ise tüm aşamalardan seçilebilir yani bağımsız değişken için herhangi bir kısıt yoktur. Hiyerarşik yapıya sahip verilerde aynı gruptaki birimlerin benzer karakteristiklere sahip olmaları (çevre, deneyim, demografik özellikler gibi,) bu birimlerden elde edilen gözlemlerin birbirine bağımlı olmasına neden olur. Bu bağımlılık ilişkisi de klasik lojistik regresyondan farklı olarak parametrelerin sabit değil gruplara göre değişen bir yapıya sahip olmasına neden olmaktadır. Çok aşamalı modellemede, parametrelerdeki bu değişim her gruba ait karakteristikler (açıklayıcı değişkenler) yardımı ile hiyerarşinin diğer aşamaları için kurulan modellerde açıklanmaya çalışılır.

Bu çalışmada da kredi kartı kullanımını etkileyen faktörler çok aşamalı lojistik regresyon analizi kullanılarak tespit edilmeye çalışılmıştır. 1. aşamayı müşteriler oluştururken, 2. aşamayı müşterilerin bağlı olduğu banka şubeleri oluşturmaktadır. Tezde, kredi kartı kullanımını etkileyen temel faktörler ortaya çıkarılmaya çalışılırken, bu faktörlerin kredi kartı kullanımı üzerindeki etkisinin banka şubelerine göre değişen bir yapısı olup olmadığı araştırılmıştır.

Yapılan iki aşamalı lojistik regresyon analizi sonucunda; banka müşterilerin ortalama kredi kartlı kullanma olasılıklarının banka şubelerine göre anlamlı farklılıklar gösterdiğine karar verilmiştir. Müşterilerin kredi kartı kullanmalarında etkili olan faktörler ve bu faktörlerin etkilerinin banka şubelerine göre farklıklar gösterip göstermediği koşullu modellerde araştırılmıştır. Bireylerin cinsiyeti ile kredi kartı

kullanımları arasında pozitif yönlü bir ilişki olduğu ve erkeklerin kredi kartı kullanım olasılıklarının bayanlarınkinden yüksek olduğu görülmüştür. Çok aşamalı lojistik regresyon modelinden elde edilen sonuçlara göre evli bireylerin bekarlara göre kredi kartı kullanma olasılığı daha fazladır. Kredi kartı sahibi olma ile meslek arasında pozitif yönlü bir ilişki olduğu da görülmektedir. Geliri yüksek olan müşterilerin kredi kartı kullanma olasılığı da yüksek çıkmıştır.

Genel olarak, koşullu modeller sonucunda, kredi kartı kullanımını müşterilerin cinsiyetleri, medeni durumları, meslekleri ve gelirleri etkilediği bulunmuştur. Bu değişkenlerin etkileri banka şubelerine göre tesadüfi olarak değişmemekte sadece ortalama kredi kartı kullanma olasılığı banka şubelerine göre anlamlı farklılıklar göstermektedir.

Küreselleşen dünyada, bankalar içinde rekabet ortamı hızla gelişmekte ve dönüşmektedir. Bankalar en büyük gelir kaynaklarını oluşturan kredi kartlarına yönelik pazaralama stratejilerine büyük yatırımlar yapmaktadır. Tezin sonuçlarının kredi kartı pazarlama stratejilerine ışık tutacağı düşünülmektedir. Çok aşamalı lojistik regresyon analizinin bankacılık sektörünü uygulanışı açısından da bundan sonraki çalışmalara ışık tutabilir.

KAYNAKLAR

[1] Tezergil S., (2009), Çok Aşamalı Lojistik Modellerin Finans Sektörüne Bir Uygulaması, Doktora Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

[2] Yılmaz, H., Budak, G. S., ve Başaran, B. (2013), "Kredi Kartı Kullanım Alışkanlıklarında Kategorik Değişkenler Arasındaki İlişkiler ve Bireylerin Davranışsal Eğilimleri: Bilecik Örneği", Uluslararası Yönetim İktisat ve İşletme Dergisi, 9(19): 31-49.

[3] Heck, R.H. ve Thomas, S. L., (2000), "An Introduction To Multilevel Modeling Techniques", Lawrence Erlbaum Assocşates, London.

[4] Obsorne, J.W., (2004), "Advantages Of Hierarchical Linear Modeling", Http://Ericae.Net/Pare/Getvn.Asp?V=7&N=1

[5] Moerbeek, M., Breukelen, G.J.Po. ve Berger, M.P.F., (2002), „ A Comparison Between Traditional Methods And Multilevel Regression For The Analysis Of Multicenter Intervention Studies", Journal Of Clinical Epidemiology, 56: 341-350.

[6] Rice, N. ve Jones, A., (1997), "Multilevel Models And Health Economics", Econometrics And Health Economics, 6: 561-575.

[7] Noyan, F. (2004), " Çok Aşamalı Modeller Ve Bir Uygulama", Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

[8] Sullivan L.M., Dukes, A. K. ve Losina, E., (1999), "Tutorial In Biostatistics An Introduction To Hierarchical Linear Modeling", Statistics In Medicine, 18: 855-888.

[9] Raudenbush, S.W. ve Bryk, A. S. (2002), "Hierarchical Linear Models: Applications And Data Analysis Methods (2nd Edition), Thousand Oaks, Ca: Sage Publications.

[10] Killip, S., Mahfound, Z. ve Pearce, K., (2004), "What Is An Intracluster Correlation Coefficient? Crucial Concepts For Primary Care Researchers", Annals Of Family Medicine, Vol.2, No.3,www.Annfammed.Org/Cgı/Reperint/213/2004.

[11] Zhang, F., Salvucci, S. ve Cohen, M.,(2000), " Multilevel Linear Regression Analysis Of Complex Survey Data" www.amstat.org/sections/ srms/Proceedings/papers/2000_029.pdf –

[12] Goldstein, H., (1986), "Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Squares", Biometrica,73: 43-56.

[13] Raudenbush, S. W., (1988), " Educational Applications Of Hierarchical Linear Models: A Review", Journal Of Educational Statistics, 13(2): 85-116.

[14] Longford, N., (1993), "Random Coefficient Models", Sage Publications, London.

[15] Dempster, A.P., Laird, N. M., & Rubin, D. B., (1977), "Maximum Likelihood From Incomplete Data Via The EM Algorithm", Journal Of The Royal Statistical Society, Series B,39: 1-8.

[16] Kreft, I & Leeuw, J. D., (1998)," Introducing Multilevel Modeling", Sage, Publications, London.

[17] Goldstein, H., (2003), "Multilevel Statistical Models", Edward Arnold, (3rd Edition), London.

[18] Hox, J., (2002), "Multilevel Analaysis, Techniques And Applications", Lawrence Erlbaum Associates, Mahwah, Nc.

[19] Maas, J. M. & Hox, J., (2004), "Influence Of Violations Assumptions On Multilevel Parameter Estimates And Their Standard Errors", Computational Statistics & Data Analysis 46: 427-440.

[20] Maas, C. J. and Hox, J. J., (2005), Sufficient sample sizes for multilevel modeling. Methodology 1: 86–92.

[21] Paccagnella, O., (2011), Sample size and accuracy of estimates in multilevel models: New simulation results. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences 7: 111–120.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Alp Abdullah ÇERÇİ

Doğum Tarihi ve Yeri : 21.04.1986, İstanbul

Yabancı Dili : İngilizce

E-posta : alp.cerci@hotmail.com

ÖĞRENİM DURUMU

Derece	**Alan**	**Okul/Üniversite**	**Mezuniyet Yılı**
Y. Lisans	Bankacılık ve Finans	İstanbul Ticaret Üniversitesi	2015
Lisans	İstatistik	Anadolu Üniversitesi	2011
	İktisat	Anadolu Üniversitesi	2014
Lise	Sayısal	Vefa Lisesi	2004

İŞ TECRÜBESİ

Yıl	**Firma/Kurum**	**Görevi**
2018-Halen	Türkiye Finans Katılım Bankası A.Ş.	Obi Müşteri Hiz. Yetkili Yardımcısı
2017-2018	MNG Kargo A.Ş.	Satış Analizi ve Planlama Yetkili Yardımcısı

2013-2016	Türkiye Finans Katılım Bankası A.Ş.	Müşteri Hizmetleri Yetkili Yardımcısı

YAYINLARI

Bildiri

1. Çerçi, A.A. ve Tekeli, F.N., (2018). “Multilevel Logistic Regression Analysis of Factors Influencıng The Usage of Credit Cards”, IWSC 2018 Conference, 8-9 March 2018, İstanbul.