LOG-LİNEER MODELLER İLE LOJİSTİK REGRESYONUN KARŞILAŞTIRMASI

Gülsüm Merve OKTAY

YÜKSEK LİSANS TEZİ
İSTATİSTİK

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

HAZİRAN 2012

ANKARA

Gülsüm Merve OKTAY tarafından hazırlanan "LOG-LİNEER MODELLER İLE LOJİSTİK REGRESYONUN KARŞILAŞTIRMASI" adlı bu tezin Yüksek Lisans tezi olarak uygun olduğunu onaylarım.

Prof. Dr. Semra ERBAŞ …….…….
Tez Danışmanı, İstatistik Anabilim Dalı

Bu çalışma, jürimiz tarafından oy birliği ile İstatistik Anabilim Dalında
Yüksek Lisans tezi olarak kabul edilmiştir.

Prof. Dr. Bedriye SARAÇOĞLU …….…….
Ekonometri Anabilim Dalı, G.Ü.

Prof. Dr. Semra ERBAŞ …….…….
İstatistik Anabilim Dalı, G.Ü.

Yrd. Doç. Dr. Meltem EKİZ …….…….
İstatistik Anabilim Dalı, G.Ü

Tarih: 13/06/2012

Bu tez ile G.Ü. Fen Bilimleri Enstitüsü Yönetim Kurulu Yüksek Lisans derecesini onamıştır.

Prof. Dr. Bilal TOKLU …….…….

Fen Bilimleri Enstitüsü Müdürü

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

Gülsüm Merve Oktay

LOG-LİNEER MODELLER İLE LOJİSTİK REGRESYONUN KARŞILAŞTIRMASI

(Yüksek Lisans Tezi)

Gülsüm Merve OKTAY

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Haziran 2012

ÖZET

Sosyal araştırmalara olan ilginin artmasıyla birlikte kategorik veri analizi daha önemli hale gelmiştir. Bu çalışmada kategorik veri analizi yöntemlerinden Lojistik Regresyon, Lojit Model ve Log-lineer modeller ele alınmış ve çeşitli açılardan istatistiksel özellikleri karşılaştırılmıştır. Araştırmalarda Lojistik Regresyonu ya da Log-lineer modelleri seçme kriterleri incelenerek TÜİK Hanehalkı 2009 Bütçe Anket verileri üzerine iki uygulama yapılmıştır. İlk uygulamada Lojistik regresyon ile Log-lineer modellerin en son aşamada modele seçtiği değişkenler karşılaştırılmıştır. İkinci uygulamada ise Lojit model ile Log-lineer modeller arasındaki ilişki ele alınarak, iki ve üç boyutlu olumsallık tablolarında lojit model parametre tahminlerinin, log-lineer model parametre tahminlerine eşitliği gösterilmiştir.

Bilim Kodu **:205.1.066**
Anahtar Kelimeler **:Kategorik veri, olumsallık tablosu, lojistik regresyon, lojit model, log-lineer modeller**
Sayfa Adedi **:131**
Tez Yöneticisi **:Prof. Dr. Semra ERBAŞ**

THE COMPARISON OF LOG-LINEAR MODELLING AND LOGISTIC REGRESSION

(MSc. Thesis)

Gülsüm Merve OKTAY

GAZI UNIVERSITY
INSTITUTE OF SCIENSES AND TECHNOLOGY
June 2012

ABSTRACT

With the growing interest in social researchs, categorical data analysis has become more important. In this study logistic regression, lojit models and log-linear models of categorical data analysis methods are discussed and compared with the statistical properties differently. Researchs' selection criteria for lojistic regression or log-linear models were examined and the two applications were applied on the Turkish Statistical Institute 2009 Household Budget Survey data. In the first application, variables of the model in the final stage were compared by log-linear models and logistic regression. In the second application, the relationship between the log-linear models and logit models were discussed and for two and three-way contingency tables logit model parameter estimates are shown to be equal to the log-linear model parameter estimates.

Science Cod : **205.1.066**
Key Words :**Categorical Data, contingency tables, logistic regression, lojit model, log-linear models**
Page Number :**131**
Manager of Thesis :**Prof. Dr. Semra ERBAŞ**

TEŞEKKÜR

Danışmanlığı süresince çalışmamın her aşamasında görüş ve katkılarıyla yardımcı olan, bilimsel alanda hiçbir şeyi esirgemeyen saygı değer hocam, sayın Prof. Dr. Semra ERBAŞ'a ve çalışmalarım sırasında göstermiş oldukları her türlü destek ve anlayıştan ötürü sevgili aileme en içten teşekkürlerimi sunarım.

İÇİNDEKİLER

Sayfa

ÖZET iv
ABSTRACT v
TEŞEKKÜR vi
İÇİNDEKİLER vii
ÇİZELGELERİN LİSTESİ x
ŞEKİLLERİN LİSTESİ xiii
SİMGELER VE KISALTMALAR xiv
1. GİRİŞ 1
1.1. Literatür Araştırması 2
1.2. Kategorik Verinin Tanımı 4
1.3. Kategorik Verinin Dağılımı 5
1.3.1. Binom dağılımı 6
1.3.2. Çokterimli dağılım 6
1.3.3. Poisson dağılımı 7
2. OLUMSALLIK TABLOLARI 8
2.1. Odds Oranı 9
2.2. Örnekleme Planları 11
2.2.1. Prospektif çalışma 12
2.2.2. Retrospektif çalışma 13
2.2.3. Diğer örnekleme planları 15
2.3. Maksimum Olabilirlik Teoremi ve Olabilirlik Oran Testi 16
2.3.1. Özellikleri 16
2.3.2. Olabilirlik fonksiyonu 17
2.3.3. İki boyutlu tablolarda maksimum olabilirlik tahmini 19
2.3.4. Olabilirlik oran G^2 testi 21
3. İKİ BOYUTLU TABLOLARDA LOG-LİNEER MODELLER VE LOJİSTİK REGRESYON 25
3.1. Log-Lineer Modeller 25
3.1.1. Log-lineer modellerin türetilmesi 25

3.1.2. İki boyutlu olumsallık tablolarında lambda parametrelerinin elde edilmesi 27
3.1.3. İki boyutlu olumsallık tablolarında hipotezler 30
3.1.4. Log-lineer modellerde odds oranı 31
3.2. Lojistik Regresyon 31
3.2.1. Lojistik model ve lojistik fonksiyon 32
3.2.2. Parametre tahmin yöntemleri 34
3.2.3. Lojistik regresyonda Hosmer-Lemeshow uyum iyiliği testi 37
3.2.4. Lojistik regresyonda katsayıların önem testi 37
3.3. İki Boyutlu Tablolarda Lojit Model ve Log-Lineer Modellerin Denkliği 39
3.3.1. İki boyutlu tablolarda katsayıların anlamlılığının testi 41
4. ÜÇ BOYUTLU TABLOLARDA LOG-LİNEER MODELLER VE LOJİSTİK REGRESYON 44
4.1. Bağımsızlık Modelleri 45
4.2. Üç Boyutlu Tablolarda Hiyerarşik Log-Lineer Modeller 50
4.2.1. Log-lineer modellerin testi 53
4.3. Lojistik Regresyonda Model Seçme Yöntemleri 56
4.4. Uygun Modelin Seçiminde Kullanılan Akaike Bilgi Kriteri (AIC) 57
4.5. Üç Boyutlu Tablolarda Lojit Model ve Log-Lineer Modellerin Denkliği 58
5. LOJİSTİK REGRESYON İLE LOG-LİNEER MODELLERİN ÖZELLİKLERİNİN KARŞILAŞTIRILMASI 61
5.1. Yapısal ve Örnekleme Sıfırları 61
5.2. Log-Lineer Modellerin Uygulamaları 62
5.3. Lojistik Regresyonun Uygulamaları 63
5.4. Lojistik Regresyonun Göreli Avantajları 64
5.5. Log-Lineer Modellerin Göreli Avantajları 65
6. HANEHALKI BÜTÇE ANKETİ ÜZERİNE BİR UYGULAMA 71
6.1. Hanehalkı Bütçe Anketi 71
6.2. Değişkenler 71
6.3. Log-lineer Modeller ile Lojistik Regresyonun Karşılaştırılması 73
6.3.1. Log-lineer modellerin uygulaması 74

6.3.2. Lojistik regresyon uygulaması 79
6.4. Log-Lineer Modeller ile Lojit Modelin Karşılaştırması 81
6.4.1. İki boyutlu tablolarda lojit model ve log-lineer modellerin denkliğinin gösterilmesi…….…….....83
6.4.2. Üç boyutlu tablolarda lojit model ve log-lineer modellerin denkliğinin gösterilmesi…….…….....87
7. SONUÇ VE ÖNERİLER…… 98
KAYNAKLAR.....100
EKLER……..103
EK- 1. Asgari ücrete göre olumsallık tablosu…….....104
EK-2. Asgari ücrete göre seçilen en son lojistik regresyon modelinin parametre ..... tahminleri 125
EK-3. Asgari ücrete göre seçilen en son lojit modelin parametre tahminleri 127
EK-4. Delta metodu teorisi 128
ÖZGEÇMİŞ 131

ÇİZELGELERİN LİSTESİ

Çizelge **Sayfa**

Çizelge 2.1. İki yönlü tabloda marjinal olasılıkların gösterimi……9

Çizelge 2.2. Odds'un 1'den büyük ya da küçük olduğu duruma göre olasılık değerinin değişimi……9

Çizelge 2.3. Olasılık değerlerine göre odds ve log odds aralıkları……10

Çizelge 2.4. Prospektif çalışma……...13

Çizelge 2.5. Prospektif çalışmaya ait olasılıklar tablosu……13

Çizelge 2.6. Retrospektif çalışma……...14

Çizelge 2.7. Retrospektif çalışmaya ait olasılıklar tablosu……15

Çizelge 2.8. Çarpımsal çokterimli dağılım için I*J tablosu……20

Çizelge 3.1. İki değişkenli log-lineer model için model parametreleri ve serbestlik dereceleri……...……...31

Çizelge 4.1. Üç boyutlu olumsallık tablosu……....44

Çizelge 4.2. Tam bağımsızlık modeli için Pearson χ^2 ve Olabilirlik oran G^2 test istatistikleri……46

Çizelge 4.3. Satırların, sütun ve tabakadan bağımsız olduğu model için Pearson χ^2 ve Olabilirlik oran G^2 test istatistikleri……....47

Çizelge 4.4. Tabakalar verildiğinde satır ve sütunların bağımsız olduğu model için Pearson χ^2 ve Olabilirlik oran G^2 test istatistikleri…….....48

Çizelge 4.5. Tüm oddsların eşit kabul edildiği model için Pearson χ^2 ve Olabilirlik oran G^2 test istatistikleri……...……..……. 50

Çizelge 4.6. Hiyerarşik log-lineer……...51

Çizelge 4.7. Kısa gösterimi ile log-lineer modeller……....52

Çizelge **Sayfa**

Çizelge 4.8. Model (1)'in Model (4)'e eşitliği için Pearson χ^2 ve Olabilirlik Oran G^2 test istatistikleri……...…….54

Çizelge 4.9. Üç boyutlu tablolarda lambda parametrelerinin serbestlik derecesi….....55

Çizelge 4.10. Log-lineer –Lojit ilişkisi……...60

Çizelge 5.1. Log-lineer modeller ve lojistik regresyon'un istatistiksel özelliklerinin karşılaştırılması……....67

Çizelge 5.2. Log-lineer modeller ile lojistik regresyonun bağımlı ve bağımsız değişkenler arasındaki birlikteliği test etmedeki farklılıkları……..…69

Çizelge 6.1. Hanehalkı fertlerinin anket ayında bitirmiş oldukları yaş grubu……..…72

Çizelge 6.2. Diploması alınan öğretim kurumu ……72

Çizelge 6.3. Asgari ücrete göre verinin genel durumu……...75

Çizelge 6.4. Log-lineer modellerde en son aşamada seçilen model……...75

Çizelge 6.5. Log-lineer modellerde en son aşamada seçilen modelin uyum İyiliği testi ……..……76

Çizelge 6.6. 15-24 yaş aralığında lise mezunu bekar bayanların haftalık çalışma süresi ve ücretlerinin olumsallık tablosu……...76

Çizelge 6.7. 15-24 yaş aralığında lise mezunu bayanların medeni durumu ve ücretlerinin olumsallık tablosu……..77

Çizelge 6.8. 25-34 yaş aralığındaki erkeklerin eğitim ve ücretlerinin olumsallık tablosu…….....78

Çizelge 6.9. Asgari ücret sınırına göre log-lineer modeller ve lojistik regresyonun model seçimleri……..81

Çizelge 6.10. Asgari ücret sınırına göre log-lineer modeller ve lojit model seçimleri ……..……...82

Çizelge 6.11. Eğitim* Aylık Ücret olumsallık tablosu……...84

Çizelge **Sayfa**

Çizelge 6.12. $\log(m_{ij}) = \lambda + \lambda^{U} + \lambda^{E} + \lambda^{U*E}$ doymuş log-lineer modeli için parametre tahminleri…….85

Çizelge 6.13. Bağımlı değişken "aylık ücret geliri" ve bağımsız değişken "eğitim" için lojit model parametre tahminleri……..……86

Çizelge 6.14. Eğitim* Aylık Ücret Geliri olumsallık tablosu…….87

Çizelge 6.15. Lojit- log lineer modeller eşitliği……..88

Çizelge 6.16. $\log(m_{ijk}) = \lambda + \lambda^{U} + \lambda^{Y} + \lambda^{E} + \lambda^{U*Y}$ modelinin parametre tahmini…….88

Çizelge 6.17. $\log(m_{ijk}) = \lambda + \lambda^{U} + \lambda^{Y} + \lambda^{E} + \lambda^{U*Y} + \lambda^{Y*E}$ modelinin parametre tahmini……...……...89

Çizelge 6.18. Bağımlı değişken "aylık ücret geliri" ve bağımsız değişken "yaş" için lojit model parametre tahminleri……...……..91

Çizelge 6.19. $\log(m_{ijk}) = \lambda + \lambda^{U} + \lambda^{Y} + \lambda^{E} + \lambda^{U*E}$ modelinin parametre tahmini……..…93

Çizelge 6.20. $\log(m_{ijk}) = \lambda + \lambda^{U} + \lambda^{Y} + \lambda^{E} + \lambda^{U*E} + \lambda^{Y*E}$ modelinin parametre tahmini …….94

Çizelge 6.21. Bağımlı değişken "aylık ücret geliri" ve bağımsız değişken "eğitim'' için lojit model parametre tahminleri…….……96

ŞEKİLLERİN LİSTESİ

Şekil **Sayfa**

Şekil 4.1. İç içe geçmiş hiyerarşik log-lineer modeller……..... 53

Şekil 5.1. Lojistik regresyona karşı log-lineer modelleri seçme akış şeması..... 70

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Kısaltmalar	Açıklama
MLE	Maksimum Olabilirlik Tahmini
TÜİK	Türkiye İstatistik Kurumu
S.D	Serbestlik Derecesi
EKK	En Küçük Kareler Yöntemi
STD.HATA	Standart Hata
OHK	Ortalama Hata Kare

1. GİRİŞ

Kategorik değişkenler, sadece sınırlı sayıda değerlerle ölçülen değişkenlerdir. Bu tür değişkenlerin analizinde "kategorik veri analizi" kapsamındaki istatistik yöntemlerinin uygulanması gerekmektedir [3]. Sosyal araştırmalara olan ilginin artmasıyla "kategorik veri analizi" daha çok önem kazanmıştır. Log-lineer modeller, lojistik regresyon ve lojit modeller kategorik verilerin analizinde en çok tercih edilen yöntemlerdir. Log-lineer modeller ile iki ya da daha fazla kategoriye sahip değişkenler arasındaki ikili ya da daha yüksek dereceli ilişkiler ve etkileşimler, bağımlı-bağımsız değişken ayrımı yapılarak veya yapılmadan saptanabilir [18]. Lojistik regresyon ise tüm veri tiplerine uygulanabilen, bağımlı değişken üzerinde bağımsız değişkenlerin etkisini ortaya koyan istatistiksel bir yöntemdir. Lojistik regresyonda bağımlı değişken nitel olmak üzere bağımsız değişkenler nicel ve nitel olabilir. Tüm bağımsız değişkenlerin kategorik değişken olması durumunda ise bir lojistik regresyon modeli "lojit" model olarak adlandırılır [1,17,42].

Bu çalışma yedi bölümden oluşmaktadır. İlk bölümde kategorik veri analizinin tarihsel gelişim süreci hakkında bilgi verilmiş, kategorik değişkenin tanımı yapılarak sahip olduğu dağılımlar ele alınmıştır.

İkinci bölümde kategorik değişkenlerin özetlenmesinde kullanılan olumsallık tablosunun yapısı ele alınmış, log-lineer modeller ve lojistik regresyonda parametreleri tahmin etmek için kullanılacak metodlar açıklanmıştır.

Üçüncü ve dördüncü bölümde sırasıyla iki yönlü ve üç yönlü olumsallık tabloları için log-lineer modeller ve lojistik regresyonun istatistiksel varsayımları, parametre tahmin yöntemleri açıklanmış ayrıca lojit modellerle log-lineer modellerin ilişkisi incelenmiştir. Beşinci bölümde her iki modelin kullanım alanları ve birbirlerine göre avantaj ve dezavanjları ele alınmıştır.

Altıncı bölüm ise ülkemizdeki gelir dağılımını belirleyen TÜİK 2009 Hanehalkı Bütçe Anketi üzerine iki uygulamadan oluşmaktadır. İlk uygulamada log-lineer

modeller ile lojistik regresyon'un en son aşamada modele seçtiği değişkenler karşılaştırılmış, ikinci uygulamada ise iki ve üç boyutlu tablolarda lojit model parametre tahminlerinin log-lineer model parametre tahminlerine eşitliği gösterilmiştir.

1.1. Literatür Araştırması

Log-lineer modeller ve lojistik regresyon yöntemleri üzerine çok geniş bir literatür bulunmaktadır.

19. Yüzyılın başlarında 2*2 tabloda değişkenler arasındaki ilişkiyi özetlemek için çapraz sınıflama kullanılırken, yüzyılın sonuna doğru kategorik veri analizinde gelişme kaydedilmiştir. Örneğin, iki boyutlu çapraz sınıflamalarda değişkenlerin bağımsızlığının testi için kullanılan ki-kare testine karşılık, Karl Pearson bir ya da daha çok kategoriye sahip değişken içeren çapraz sınıflama üzerine çalışmıştır. Aynı dönemde Goodman ve Kruskal (1954, 1959, 1963) değişkenler arasındaki ilişkinin ölçülmesine yoğunlaşırken, Lancester(1950), S.N. Roy ve öğrencileri (1950) kategorik veri analizi için parametrelerin tahmin yöntemleri üzerine çalışmışlardır. 1960'lı yıllara gelindiğinde Roy, Kastenbaum ve Birch, Darroch, Good ve Goodman çok boyutlu olumsallık tablolarının analizi üzerine çalışmalarını yoğunlaştırmıştır. Çapraz tablolara olan ilginin ve bilgisayar kullanımının artmasıyla birlikte gelişmeler devam etmiştir. Yapılmış olan bazı çalışmalar aşağıdaki gibi özetlenebilir.

Kategorik veriler için log-lineer modeller, Bunker [1969], Bishop, Fienberg, Holland [1975], Mosteller [1968] tarafından hazırlanan National Halothane çalışmasında kullanılmıştır. Bu çalışma genel olarak hastalıkların altında yatan nedenleri ortaya koyan istatistiksel bir çalışmadır. Bu çalışmada halothane anestezi ve ameliyat sonrası hepatik nekroz arasındaki ilişkiyi ele alır. Buradaki asıl amaç halothane ile diğer genel anestezikleri karşılaştırmaktır. Bu çalışmaya benzer olarak Cox [1966] ve Berksan [1944, 1946] tarafından lojistik regresyon kullanılarak Framigham çalışması gerçekleştirilmiştir. Framigham çalışması William B. Kannel (1923-2011) tarafından ele alınmış olup geçen yüzyılın ilk 10 tıbbi gelişmesi arasında yer almaktadır. Dr.

Kannel bu çalışmada kardiyovasküler (kalp-damar sistemi) hastalık riskini tahmin etmek için birden fazla faktörü ele almıştır. Kandaki yüksek şeker ve kolesterol miktarı, kalpteki ritim bozukluğu, hastanın daha önce kalp krizi geçirmiş olması, obezite, sigara kullanımı, yüksek kan basıncı bu faktörlerden bazılarıdır.

1970'li yıllarının ortalarına gelindiğinde probit model, lojit model, log-lineer modeller ve lojistik regresyon gibi kategorik veri tekniklerinin ekonomide, sağlık araştırmalarında, sosyolojide, pazarlama alanında artarak kullanıldığı fakat davranışlara dayalı araştırmalarda bu tekniklerin pek kullanılmadığı görülmektedir [7].

Upton (1978) log-lineer modellerde değişkenler arasındaki ilişkileri dikkate alırken, Agresti (1984, 1990), Andersen (1991), Bishop, Fienberg, Holland (1975), Everitt (1977), Fienberg (1980), Haberman (1974), Plackett (1981), Read ve Cressie (1988), Santner ve Duffy (1989), Cox ve Snell (1989), Hosmer ve Lemeshow (1989) lojistik regresyona odaklanmışlardır.

Log-lineer modeller'in lojistik regresyon ve lojit modellerle ilişkisini başta Agresti (1995) olmak üzere Knoke ve Burke (1975), Christensen (1997), Laval (2003) kitaplarında detaylı bir şekilde ele almışlardır. Bu kitapların yanında çeşitli makaleler, çalışmalar yapılmıştır. Bu çalışmalardan birinde David Knoke (1975) doğrusal olasılık modeli ile log-lineer modellerin istatistiksel sonuçlarını karşılaştırmıştır. Eğer bağımlı değişken iki değer alıyorsa ve bağımlı değişken için belirlenen olayın gerçekleşme olasılığı 0,25 ile 0,75 arasındaysa doğrusal olasılık modeli ile log-lineer modellerin sonuçlarının birbirine yakın olacağı; fakat bağımlı değişkenin gerçekleşme olasılığı 0'a ya da 1'e yakınsa sonuçlarda farklılık olabileceği iddia edilmiştir. Ayrıca bu makalede odds ve log oddslara göre çeşitli yorumlar yapılmıştır. Bir diğer karşılaştırma Tansey R. ve diğerleri (1996) tarafından gerçekleştirilmiştir. Bu çalışmada ise log-lineer modeller ve lojistik regresyon'un istatistiksel varsayımları, kullanım alanları, bağımlı ve bağımsız değişkenler arasındaki ilişkileri test etmede birbirlerine göre üstünlükleri ele alınmıştır. Makalede her iki modelin özelliklerinin birbirinden farklı olduğu belirtilerek, hangi modelin

kullanılması gerektiğine; araştırmanın amacına ve kullanılan değişkenlerin ölçüm düzeylerine göre karar verilmesi gerektiği sonucuna varılmıştır.

Timothy E. Hanson, Wesley O. Johnson ve Ian A. Gardner (2000) da domuz ve sığırlarda ortaya çıkan üç çeşit hastalığın teşhisinde kullanılan kan testlerini ele almışlardır. Bu testlerin birbirine olan bağımlılıkları lojistik regresyon ve log-lineer modeller kullanılarak test edilmiştir. Üç hastalık için yapılan çalışmalarda iki hastalık için lojistik regresyon ve log-lineer modellerde aynı değişkenlerin modele alındığı fakat bir hastalık için farklı değişkenlerin modele alındığı belirtilmiştir. Bu çalışmada ayrıca log-lineer modellerin lojit modele eşit olduğu durumlardan bahsedilmiştir. Dr. Mohammad Alaya da kalp rahatsızlığı üzerinde diyabet, cinsiyet, sigara kullanımı, genetik rahatsızlıklar gibi çeşitli faktörlerin etkisini araştırmıştır. Log-lineer modeller ve lojistik regresyon birbirinin tamamlayıcısı olarak ele alınmış, log-lineer modeller veriyi analiz etmek için, lojistik regresyon da kalp rahatsızlığı üzerinde etkili olan bu faktörlerin parametrelerini tahmin etmek için kullanılmıştır.

1.2. Kategorik Verinin Tanımı

Genel olarak birimlerin ölçülebilen özelliklerine "değişken" adı verilir ve nicel -nitel değişkenler olarak sınıflandırılabilir. Nicel değişkenler doğrudan sayılarla ifade edilebilen (ücret, fiyat, uzunluk vb.) , nitel değişkenler ise doğrudan sayılarla ifade edilemeyen (cinsiyet, medeni durum, meslek vb.) değişkenlerdir. Nicel değişken belirli bir aralıkta tüm noktaları gerçel değer olarak alabiliyorsa sürekli, sadece belirli noktaları alabiliyorsa kesikli değişken olarak tanımlanır [3]. Birimlerden bilgi toplanmasında değişkenlerin nasıl ölçüldüğü önemlidir. Ölçek, değişken değerlerinin toplanma şeklidir. Sınıflama, sıralama, eşit aralıklı ve oranlama ölçme düzeyleri olmak üzere ölçüm düzeyleri dörde ayrılır. Sınıflama ve sıralama ölçme düzeyi nitel değişkenler için, eşit aralıklı ve oranlama ölçme düzeyi nicel değişkenler için kullanılır.

Kategorik değişkenler, sadece sınırlı sayıda değerlerle ya da kategorilerle ölçülen değişkenler olarak tanımlanmaktadır [2]. Bu tanımlama kategorik veriyi sürekli

değişkenlerden ayırır. Kategorik değişkenleri, kesikli nicel değişkenler ve nitel değişkenler olarak da değerlendirebiliriz.

Sosyal bilimlerle ilgili çalışmalarda kullanılan birçok değişken kategoriktir [3].

Örneğin bir ürünün tüketimi araştırmasında,
1-Bu ürünün kullanılıp kullanılmadığı,
2- Ne kadar sıklıkla kullanıldığı,
3- Üründen memnun olunup olunmadığı gibi birçok soru sorulabilir.

Burada 1. sorunun cevabı sınıflama düzeyinde kategorik değişken, 2. ve 3.soruların cevapları ise sıralama düzeyinde kategorik değişkendir. Bu gibi veriler olumsallık tablosu dediğimiz bir tabloda özetlenir. Klasik doğrusal regresyon modeli bu tip verilere uygun değildir. Bağımlı değişkenin sürekli olduğu ve tahmin edilen hataların normal dağıldığı durumda, yansız parametre tahminlerini sağlayan en küçük kareler tahmini kullanılamaz. Çünkü;

1- İki değer alan bağımlı değişken için tahmin edilen değerler 0-1 aralığının dışına düşebilir, bu da anlamsız sonuçlar doğurur.
2- Bağımlı değişken normal dağılıma sahip olmadığı için tahmin edilen hatalar da normal dağılıma sahip değillerdir. Bu nedenle bağımsız değişkenlerin katsayılarının tahmini için t testi kullanılamaz.
3- Tahmin edilen hatalar değişen varyansa sahiplerdir.
4- Bağımsız değişkenlerle modelin açıklanma gücünü belirten R^2 determinasyon katsayısı, uyum iyiliğinin ölçüsünde kuşkulu sonuçlar verir [1].

1.3. Kategorik Verinin Dağılımı

Sürekli değişkenlere uygulanan birçok analizde dağılım varsayımı bulunmaktadır. Bağımlı değişkenin sürekli olduğu durumda, regresyon modeli ve varyans analizi için normal dağılım varsayımı en önemli varsayımlardan biridir [3]. Bu bölümde kategorik değişkenler için üç dağılımı ele alacağız.

1.3.1. Binom dağılımı

İki mümkün sonucu olan (yani başarı/başarısızlık, evet/hayır, 1/0 vb) bir deneyde her denemenin, "başarı "olarak ifade edilen sonuca ulaşma olasılığı " p " , "başarısızlık" olarak ifade edilen sonuca ulaşma olasılığı "$1-p$" dir. Başarı sonucu elde edildiğinde x=1, başarısızlık sonucu elde edildiğinde x=0 değerlerini alan X rasgele değişkeni, Bernoulli rasgele değişkenidir [5]. Eğer Bernoulli denemelerini n kez birbirinden bağımsız olarak tekrarlarsak, o zaman x başarı sayısı binom dağılımına sahip olur. Matematiksel notasyon ile bir rassal değişken X, binom dağılımına sahipse $X \sim B(n,p)$ şeklinde ifade edilir.

X rassal değişkeninin dağılımın olasılık fonksiyonu Eş. 1.1'deki gibidir. $x = 0,1,2,\ldots,n$

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^x \tag{1.1}$$

Burada her deneme birbirinden bağımsızdır yani bir denemenin sonucu diğer bir denemeyi etkilemez. Bu denemeler "Bernoulli Denemeleri " olarak bilinir. Eğer n=1 olursa, bu Binom dağılım, yani B(1, p), gerçekte bir Bernoulli dağılımı ile aynıdır. Dağılımın beklenen değeri ve varyansı aşağıdaki gibidir. [3]

$$E(X) = np,$$
$$Var(X) = np(1-p) \tag{1.2}$$

1.3.2. Çokterimli dağılım

Çokterimli (multinomial) dağılım ikiden fazla (k tane) mümkün sonuçlu deneyler için uygulanan kesikli bir dağılımdır. Binom dağılımının genelleştirilmiş şeklidir.

n_i , $i.$ mümkün sonucun kaç kez meydana geldiğini göstersin $i = (1, 2, \ldots, k)$.

p_i her denemede $i.$ olayın meydana gelme olasılığıdır. Bu durumda $(n_1, n_2, ..., n_k)$, $(n, p_1, ..., p_k)$ parametreli çokterimli dağılım gösterir [3].

Dağılımın olasılık fonksiyonu $x_i \geq 0$, $x_1 + ... + x_k = n$ koşulları altında Eş. 1.3'deki gibidir.

$$p(n_1 = x_1, ..., n_k = x_k) = \frac{n!}{x_1! ... x_k!} p_1^{x_1} ... p_k^{x_k} \qquad (1.3)$$

Genel olarak her bir mümkün sonuç Binom dağılımı gösterir ve
$n_i \sim B(n, p_i)$ şeklinde ifade edilir. Dolayısıyla n_i'lerin beklenen değeri ve varyansı Binom dağılımındaki gibi hesaplanır.

$E(n_i) = np_i$,
$Var(n_i) = np_i(1-p)$ dir [3]. (1.4)

1.3.3. Poisson dağılımı

Bir olayın belirli aralıkta, x kez meydana gelme olasılığının hesaplanmasında kullanılan olasılık dağılımlarından biridir. Bu durumda olasılık fonksiyonu aşağıdaki gibidir [5].

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!} \qquad x = 0,1,2,...,n \quad (e = 2,718) \qquad (1.5)$$

Burada λ gerçekleşen ortalama olay sayısıdır ve λ=n.p dir. Poisson dağılımının aritmetik ortalaması ve varyansı aynı olup λ'ya eşittir [5].

$$E(x) = Var(x) = \lambda \qquad (1.6)$$

2. OLUMSALLIK TABLOLARI

Olumsallık tabloları genellikle kategorik değişkenler arasındaki ilişkilerin ortaya çıkarılmasında kullanılan tablolardır. Tablolar içerdikleri değişken sayılarına göre adlandırılırlar. Örneğin tablo tek değişkene sahipse, tek-yönlü (boyutlu); iki değişkene sahipse iki-yönlü olumsallık tabloları gibi.

Genel olarak tablolardaki satır değişkeni I, sütun değişkeni J ile gösterilir. Tablolarda yer alan değerler, değişken kategorilerinin veri kümesinde kaç kez tekrarlandığını gösteren frekanslardır [24]. Satır değişkeninin kategorisi i, sütun değişkeninin kategorisi j ile belirtilirse, tablolardaki frekanslar (f_{ij}) ile gösterilir. i.satır ve j. sütunda olayın gerçekleşme olasılığı ise p_{ij} ile gösterilmektedir. p_{ij}'nin dağılımı, satır ve sütun değişkenlerinin ortak dağılımıdır.

Marjinal olasılıklar ise, ortak olasılıkların toplanması sonucundan elde edilen satır ve sütun toplamlarıdır ve toplam birim sayısı n'in yüzde kaçının i. kategoride veya j. kategoride yer aldığını gösterir. "$p_{i.}$" satır değişkeni, "$p_{.j}$" sütun değişkeni için kullanılır. Burada alt indis ".", sabitlenen değişken üzerinden toplamları gösterir. Marjinal olasılıklar aşağıdaki gibi hesaplanır [11].

$$p_{i.} = \sum_j p_{ij}$$

$$p_{.j} = \sum_i p_{ij} \quad \text{ve} \quad \sum_i p_{i.} = \sum_j p_{.j} = \sum_i \sum_j p_{ij} = 1 \quad \text{dir.} \tag{2.1}$$

Bu marjinal olasılıklar sayesinde tek bir değişkenin bilgisi elde edilebilir. İki yönlü tablolar için marjinal olasılıklar Çizelge 2.1'deki gibidir. Burada $p_{1.} + p_{2.} = 1$ ve $p_{.1} + p_{.2} = 1$ dir.

Çizelge 2.1. İki yönlü tabloda marjinal olasılıkların gösterimi [11]

Değişkenler	Sütun değişkeni		
Satır değişkeni	J_1	J_2	Marjinal olasılıklar (satır için)
I_1	p_{11}	p_{12}	$p_{11} + p_{12} = p_{1.}$
I_2	p_{21}	p_{22}	$p_{21} + p_{22} = p_{2.}$
Marjinal olasılıklar (sütun için)	$p_{11} + p_{21} = p_{.1}$	$p_{12} + p_{22} = p_{.2}$	1

2.1. Odds Oranı

Odds oranı (üstünlük oranı) verilerin analizinde özellikle log-lineer modellerde ve lojistik regresyonda yaygın olarak kullanılan bir tekniktir. İlk önce odds'u tanımlayalım. Odds, bir olayın gerçekleşme olasılığının (p), gerçekleşmeme olasılığına ($1-p$) oranıdır.

$$odds = \Omega = \frac{p}{1-p} \tag{2.2}$$

Odds'un 1'den büyük ya da küçük olduğu duruma göre olasılık değeri Çizelge 2.2'deki gibi değişir.

Çizelge 2. 2. Odds'un 1'den büyük ya da küçük olduğu duruma göre olasılık değerinin değişimi

$odds = \Omega < 1$	$odds = \Omega > 1$
$\frac{p}{1-p} < 1$	$\frac{p}{1-p} > 1$
$p < (1-p)$	$p > (1-p)$
$2p < 1$	$2p > 1$
$p < 0,50$	$p > 0,50$

Görüldüğü üzere Ω ne kadar artarsa olayın gerçekleşme olasılığı da o kadar artar. Ω ne kadar sıfıra yaklaşırsa, olasılık da sıfıra yaklaşır. Bununla birlikte olasılık 1'e yaklaştıkça Ω sonsuza gider. Ayrıca bir olayın meydana gelmesinin oddsu belli olduğunda kolaylıkla olayın gerçekleşme olasılığını Eş. 2.3'deki gibi elde edebiliriz.

$$p = \frac{\Omega}{1+\Omega} \tag{2.3}$$

Çizelge 2.3. Olasılık değerlerine göre odds ve log odds aralıkları

Olasılık	Odds	Log odds
$0 < p < 0,50$	$0 < \Omega < 1$	$-\infty < \log\Omega < 0$
$0,50 < p < 1$	$1 < \Omega < \infty$	$0 < \log\Omega < \infty$

Odds oranı ise bir olayın bir grup için olasılıklar oranının, başka bir grubun olasılıklar oranına olan oranıdır. Yani ilk grup için hesaplanan olasılıkların oranı (Ω_1) ve ikinci grup için hesaplanan olasılıklar oranı (Ω_2) olduğunda odds oranı aşağıdaki gibi gösterilebilir.

$$\theta = \frac{\Omega_1}{\Omega_2} \tag{2.4}$$

Odds oranı, satırlar sütun, sütunlar satır olduğunda değişmez. Bu Eş. 2.4'ün yapısından kaynaklanmaktadır.

Çizelge 2.1 için odds oranı aşağıdaki gibidir.

$$\theta = \frac{(p_{11} / p_{12})}{(p_{21} / p_{22})} = \frac{p_{11}p_{22}}{p_{21}p_{12}} \tag{2.5}$$

Lemma 2.1: Satır ve sütunun bağımsız olduğu varsayımı altında, $\hat{p}_{ij} = \hat{p}_{i.}\hat{p}_{.j} = n_{i.}n_{.j}/(n_{..})^2$ dir böylece tahmin edilen odds oranı her zaman 1 olur.

İspat:

$$\theta = \frac{p_{11}p_{22}}{p_{21}p_{12}} = \frac{(p_{1.}p_{.1})(p_{2.}p_{.2})}{(p_{2.}p_{.1})(p_{1.}p_{.2})} = \frac{(\frac{n_{1.}}{n..}\frac{n_{.1}}{n..})(\frac{n_{2.}}{n..}\frac{n_{.2}}{n..})}{(\frac{n_{2.}}{n..}\frac{n_{.1}}{n..})(\frac{n_{1.}}{n..}\frac{n_{.2}}{n..})} = 1$$

Eğer bağımsızlığı göz ardı edersek, yine de odds oranının 1'e yaklaştığını gözlemleyebiliriz. Burada $\hat{p}_{ij} = n_{ij}/n_{..}$ olur [11].

$$\frac{\hat{p}_{11}\hat{p}_{22}}{\hat{p}_{12}\hat{p}_{21}} = \frac{n_{11}n_{22}}{n_{12}n_{21}} \qquad (2.6)$$

Eğer $1 < \theta < \infty$ ise, birinci satırın "başarı"sı, ikinci satırın "başarı"sından fazladır. $0 < \theta < 1$ durumunda ise tam tersi bir yorum geçerlidir [3].

2.2. Örnekleme Planları

Christensen'e göre iki türlü çalışma vardır. Bunlar "prospektif - ileriye yönelik" ve "retrospektif - geriye dönük" çalışmalardır. Prospektif çalışmalar çarpımsal çokterimli örneklemeye sahip çalışmalardır. Bu örnekleme planında bağımsız değişkenlerin her bir kategorisindeki gözlem sayıları her satırda sabitlenmiştir ($n_{ij.}$) ve her satır bağımsız çokterimli dağılıma sahiptir.

Log-lineer modeller ve Lojit modellerin birlikte ele alınmasındaki varsayımlardan bir tanesi de çarpımsal çokterimli örneklemenin varlığıdır. Bu yüzden çarpımsal çokterimli örnekleme tasarısının detaylarını "prospektif" ve "retrospektif" çalışmaları inceleyerek ele alacağız.

2.2.1. Prospektif çalışma

Bağımlı değişkenin kategorileri, bağımsız (açıklayıcı) değişkenlerin her bir kombinasyonu için aynı dağılıma sahipse ele alınan çalışma "prospektif "çalışma olarak adlandırılır. Christensen bu durumu bir örnek üzerinde anlatmıştır. Kalp krizi geçirme durumu için 250 deneğin her birinin; yaş, cinsiyet, kandaki kolesterol miktarı ve basınç düzeyi gibi bağımsız (açıklayıcı) değişkenler yardımıyla kalp krizi geçirip geçirmediği tahmin edilebilir. Bu ileriye yönelik çalışmada, bağımsız değişkenlerin her bir kombinasyonu, bir bireyin bağımlı değişkenin kategorilerine rasgele atanmasında etkilidir. Böylece birden fazla yığın oluşur ve her bir birey farklı yığından örneklendirilmiş gibi olur. Bu yüzden prospektif çalışmalar çarpımsal çokterimli örneklemeye sahiplerdir. Burada çok terime sahip olan kategoriler bağımlı değişkenin kategorileridir [1]. Satır toplamları yani bağımsız değişkenlerin kombinasyonlarını içeren her bir satırdaki toplam birey sayısı sabitlenmiştir. i , j açıklayıcı değişkenlerin bir setini göstersin. p_{ij1} kişinin kalp krizi geçirme, p_{ij2} kişinin kalp krizi geçirmeme olasılığı olsun.

Burada, $p_{ij1} + p_{ij2} = 1$'dir ve $p_{ij2} = 1 - p_{ij1}$ kişinin kalp krizi geçirmeme olasılığıdır. Dolayısıyla bağımlı değişken için olasılıkların oranı (p_{ij1} / p_{ij2}), odds oranıdır. Ayrıca beklenen hücre değerleri (m_{ijk}) için $m_{ij.} = n_{ij.}$ varsayılır. Tahminlerimiz $\hat{m}_{ij.} = \hat{n}_{ij.}$ kısıtını sağlamalıdır. O yüzden i j açıklayıcı değişkenleri tüm modellerde yer almak zorundadır ve $p_{ij1} / p_{ij2} = m_{ij1} / m_{ij2}$ varsayımı geçerlidir [1].

Örnek 2.1: Çarpımsal çokterimli örnekleme ve prospektif çalışma üzerine bir örnek Çizelge 2.4 ile gösterilmiştir. Burada tablodaki frekanslar örnek için türetilmiştir ve bağımsız değişkenlerin satır toplamları sabit tutulmuştur.

Çizelge 2.4. Prospektif çalışma [1]

		Daha önce Kalp krizi geçirme durumu		
	Cinsiyet	Evet	Hayır	Toplam
Yaş				
40 yaş üstü	Bayan	20	30	**50**
	Erkek	25	75	**100**
40 yaş altı	Bayan	40	20	**60**
	Erkek	20	20	**40**
Toplam		105	145	250

Olasılıklar, $p_{ijk} = n_{ijk} / n_{ij.}$ formülü ile elde edilir.

Çizelge 2.5. Prospektif çalışmaya ait olasılıklar tablosu

		Daha önce Kalp krizi geçirme durumu		
	Cinsiyet	Evet	Hayır	Toplam
Yaş				
40 yaş üstü	Bayan	0,40	0,60	0,20
	Erkek	0,25	0,75	0,40
40 yaş altı	Bayan	0,667	0,333	0,24
	Erkek	0,50	0,50	0,16
Toplam		0,42	0,58	1

$p_{11.} = p_{111} + p_{112} = 0,40 + 0,60 = 1$

$p_{12.} = p_{121} + p_{122} = 0,25 + 0,75 = 1$

$p_{21.} = p_{211} + p_{212} = 0,667 + 0,333 = 1$

$p_{22.} = p_{221} + p_{222} = 0,50 + 0,50 = 1$

Görüldüğü üzere $p_{ij1} + p_{ij2} = 1$'dir.

2.2.2. Retrospektif çalışma

Christensen'in bir diğer örneğinde ise kalp krizi geçirdiği bilinen 100 kişi ve kalp krizi geçirmediği bilinen 150 kişi ele alınır. Her bir birey, bağımsız değişkenlerin kategorilerine göre sınıflara ayrılabilir. Burada sadece iki yığın vardır: kalp krizi

geçirenler ve geçirmeyenler. Bu iki yığın için çokterimlilerin kategorileri, bağımsız (açıklayıcı) değişkenlerin farklı kategorileridir. Bu durumda çalışma geriye dönük "retrospektif" çalışma olarak adlandırılır ve bağımlı değişkenin kategorileri bağımsız değişkenlerin çokterimli durumları ile oluşur. Burada sadece bağımlı değişkenin iki kategorisinin birim sayıları sabittir (kalp krizi geçiren 100 kişi ve kalp krizi geçirmeyen 150 kişi) .

Bağımsız değişkenlere bağlı olarak her bir hücreye bireyin düşme olasılığı farklı olduğu için $p_{ij1} + p_{ij2} \neq 1$ dir, yani p_{ij1} / p_{ij2} oranı odds oranı değildir. Bu nedenle lojit model ve log-lineer modellerin birlikte ele alınmasında $\log(p_{ij1} / p_{ij2})$ hakkındaki varsayımları uygulamak zorlaşır ve $p_{ij1} / p_{ij2} = m_{ij1} / m_{ij2}$ olarak belirtilemez [1].

Örnek 2.2: Retrospektif çalışma üzerine bir örnek Çizelge 2.6 ile gösterilmiştir. Burada, bağımlı değişkenin kategorilerinin toplamları yani sütunlar sabit tutulur.

Çizelge 2.6. Retrospektif çalışma [1]

		Daha önce Kalp krizi geçirme durumu		
	Cinsiyet	Evet	Hayır	Toplam
Yaş				
40 yaş üstü	Bayan	20	60	80
	Erkek	70	15	85
40 yaş altı	Bayan	5	35	40
	Erkek	5	40	45
Toplam		**100**	**150**	250

Olasılıklar: $p_{ijk} = n_{ijk} / n_{..k}$ formülü ile elde edilir.

Çizelge 2.7. Retrospektif çalışmaya ait olasılıklar tablosu

		Daha önce Kalp krizi geçirme durumu		
	Cinsiyet	Evet	Hayır	Toplam
Yaş				
40 yaş üstü	Bayan	0,20	0,40	0,32
	Erkek	0,70	0,10	0,34
40 yaş altı	Bayan	0,05	0,234	0,16
	Erkek	0,05	0,266	0,18
Toplam		0,40	0,60	1

$p_{11.} = p_{111} + p_{112} = 0,20 + 040 \neq 1$

$p_{12.} = p_{121} + p_{122} = 0,70 + 0,10 \neq 1$

$p_{21.} = p_{211} + p_{212} = 0,05 + 0,234 \neq 1$

$p_{22.} = p_{221} + p_{222} = 0,05 + 0,266 \neq 1$

Görüldüğü üzere $p_{ij1} + p_{ij2} \neq 1$'dir ve odds oranı kullanışlı değildir.

2.2.3. Diğer örnekleme planları

Kategorik veri analizinde yığından veri elde etmede birden fazla örnekleme planı vardır. Bunlardan en önemlileri aşağıda sıralanmıştır.

Poisson: Bu örnekleme planında veri zamanın belli kısımlarında gözlemlenir, toplam birim sayısı n ve her bir hücredeki birim sayısı n_{ijk} poisson dağılımına sahiptir.

Çokterimli: Toplam birim sayısı n sabitlenmiştir ve her bir hücredeki birim sayısı n_{ijk} çokterimli dağılıma sahiptir.

Çarpımsal çokterimli: Bağımsız değişkenlerin her bir kategorisindeki gözlem sayıları her satırda sabitlenmiştir ($n_{ij.}$). Burada her satır bağımsız çokterimli dağılıma sahiptir.

Hipergeometrik: Hem satır toplamları hem sütun toplamları sabitlenmiştir.

Retrospektif çarpımsal binom: Sadece bağımlı değişkene ait olan sütun toplamları sabitlenir.

Yukarıda bahsi geçen örnekleme planlarından bir tanesini bir diğerine dönüştürebiliriz. Örneğin poisson örneklemesi, toplam sayı sabit tutularak uygulanırsa çokterimli örneklemeye; poisson ve çokterimli örnekleme satır toplamları sabitlenirse çarpımsal çokterimli örneklemeye ya da sütun toplamları sabitlenirse retrospektif örneklemeye dönüştürülebilir [1].

2.3. Maksimum Olabilirlik Teoremi ve Olabilirlik Oran Testi

Model parametrelerini tahmin etmek için en sık kullanılan yöntemler en küçük kareler (EKK) ve maksimum olabilirlik yöntemidir. Zaman ve maliyet kısıtlamaları nedeniyle çoğu çalışma yığından seçilen örneklerle yapılmaktadır. Bu yüzden yığın parametrelerinin tahmini istatistikteki en önemli konulardan biridir.

Maksimum olabilirlik yöntemi isminden de anlaşılacağı üzere, olasılık dağılımı belli olan istatistiksel bir model için örneğin karşımıza çıkması olabilirliğin en yüksek dağılımın parametresini tahmin etmek üzere, iyi özelliklere sahip tahmin ediciyi bize sağlar.

2.3.1. Özellikleri

En küçük kareler yöntemi (EKK) lineer regresyon, hata kareler toplamı, standart sapma gibi istatistiksel konularla yakından ilişkilidir ve maksimum olabilirlik yönteminin aksine dağılım varsayımlarına en alt seviyede ihtiyaç duyar. Gözlenen veriyi özetlemek için kullanışlıdır, fakat hipotezleri test etmede ya da güven aralıklarını oluşturmada varsayım temeli yoktur [16]. Maksimum olabilirlik yöntemi ise parametre tahmininde ve istatistiksel çıkarımlarda standart bir yaklaşımdır. Maksimum olabilirlik tahmin edicisi aşağıda verilen optimal özelliklere sahiptir.

1) Tutarlılık: Tahmin edicinin, tahmin edilen değerin olasılığına yakınsamasıdır.
2) Asimptotik normallik: Örnek çapı arttıkça, maksimum olabilirlik tahmin edicisinin dağılımı, normal dağılıma yaklaşır.

3) Etkinlik: Asimptotik olarak elde edilen parametre tahminlerinin en düşük varyansa sahip olmasıdır [16].

Bu özellikler en küçük kareler yöntemi için söylenemez. Çoğu istatistikçi en küçük kareler yöntemini parametre tahmini için genel bir yöntem olarak görmez ve istatistikteki çıkarım metodlarının çoğu maksimum olabilirlik yöntemine dayalı olarak geliştirilmiştir. Örneğin, maksimum olabilirlik yöntemi; χ^2, G^2, Bayesyen metodlar, rastgele etkilerin modellemesi, Akaike bilgi kriteri, Bayesian bilgi kriteri için bir önkoşuldur [16].

2.3.2. Olabilirlik fonksiyonu

$x_1, \ldots, x_n$; n bağımsız gözlemin örneği olsun. Bu örnekler bilinmeyen olabilirlik yoğunluk fonksiyonundan (f_0) elde edilirler. $f_0 = f(.|\theta_0)$ ile gösterilir. Burada θ_0 değeri bilinmez ve parametrelerin "gerçek değer" i olarak kabul edilir. Amaç, θ_0 değerine yakın olan $\hat{\theta}$ tahmin edicisini bulmaktır.

Maksimum olabilirlik yöntemini kullanmak için, ilk önce tüm gözlemlerin ortak yoğunluk fonksiyonu tanımlanmalıdır. Bağımsız ve aynı dağılımlı bir örnek için ortak yoğunluk fonksiyonu her bir gözlemin olasılık yoğunluk fonksiyonlarının çarpımına eşittir.

$$f(x_1, \ldots, x_n|\theta) = f(x_1|\theta) \ldots f(x_n|\theta) \tag{2.7}$$

Parametre değerlerinin bir seti verildiğinde, uygun bir olabilirlik yoğunluk fonksiyonu bazı verilerin daha olası olduğunu gösterebilir [16]. Eğer tam tersi düşünülürse, gözlenen veriler verildiğinde, tüm olabilirlik fonksiyonları arasından veriyi en olası üreten bir olabilirlik yoğunluk fonksiyonunu bulmak isteriz. Bu problemi çözmek için olabilirlik fonksiyonunu tanımlarız. Burada θ parametre

vektörü ve x veri vektörü rollerini değiştirirler. θ fonksiyon değişkeni serbestçe değiştiği için, bu dağılım fonksiyonu olabilirlik olarak isimlendirilir [39].

$$L(\theta|x_1,...,x_n) = f(x_1,...,x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \tag{2.8}$$

Burada olasılık yoğunluk fonksiyonu ile olabilirlik fonksiyonu arasında önemli bir fark vardır. Olasılık yoğunluk fonksiyonu sabitlenmiş bir parametre için verilerin olasılığını verirken, olabilirlik fonksiyonu sabitlenmiş bir veri seti için parametre değerlerinin olabilirliğini verir [16].

Olabilirlik fonksiyonunu maksimum yapmak yerine, bu fonksiyonun logaritmasını maksimumlaştırmak daha kolaydır. Çünkü logaritma tam anlamıyla artan bir fonksiyondur, olabilirliğin maksimumu ve log-olabilirliğin maksimumu aynı noktada meydana gelir. Bu yüzden maksimum olabilirlik tahmin edicisi olabilirlik ve log-olabilirlik fonksiyonlarını birbirinden ayırt edemez.

Log-olabilirlik fonksiyonu Eş. 2.9'daki gibidir.

$$\log L(\theta|x_1,...,x_n) = \sum_{i=1}^{n} \log f(x_i|\theta) \tag{2.9}$$

Diferansiyel fonksiyonun maksimumunun ya da minimumunun olması, bu noktalarda ilk türevlerinin sıfıra eşit olması ile belirtilir.

Log-olabilirlik fonksiyonun θ' ya göre birinci türevi alınır.

$$\frac{\partial \log L(\theta|x_1,...,x_n)}{\partial \theta_i} = 0 \tag{2.10}$$

Ve bulunan eşitlik aşağıdaki gibidir.

$$\hat{l}(\theta|x) = \frac{1}{n}\sum_{i=1}^{n} \log f(x_i|\theta) \tag{2.11}$$

Maksimum olabilirlik metodu θ_0'ı , $\hat{l}(\theta|x)$i maksimum yapan θ değerini bularak tahmin eder. Bulunan değer, θ_0' ın maksimum olabilirlik tahmin edicisidir. İkinci bir koşul da, $\log L(\theta|x)$'in maksimum olduğunu ya da minimum olmadığını ispatlamaktır. Maksimum olması için log-olabilirlik fonksiyonu konveks şekilde olmalıdır. Bu log-olabilirliklerin Eş. 2.12'deki gibi ikinci türevinin negatif olduğunu ispatlayarak belirtilir [16].

$$\frac{\partial^2 \log L(w|y)}{\partial w^2_i} < 0 \tag{2.12}$$

Ele alınan model birden fazla parametre içerdiğinde ve olasılık yoğunluk fonksiyonu lineer olmadığında maksimum olabilirlik tahminlerini elde etmek zorlaşır. Bu gibi durumlarda maksimum olabilirlik tahmini lineer olmayan optimizasyon algoritmaları ile elde edilirler. Buradaki amaç, log-olabilirliği maksimum yapan optimal parametreleri hemen bulmaktır. Tüm parametreleri araştırmaktan ziyade çok boyutlu parametre uzayının daha küçük alt setleri ile çalışılır. İteratif aşamalarla ilerlenir. Özellikle, her bir iterasyonda, bir önceki iterasyonda elde edilen parametrede küçük değişiklikler yapılarak model geliştirilmeye çalışılır [16].

2.3.3. İki boyutlu tablolarda maksimum olabilirlik tahmini

Birbirinden bağımsız olan I yığından alınan örneklerin çarpımsal çokterimli dağılımını ele alalım. Burada örneklerin her biri aynı J kategoriye bölünmüştür. I yığın I*J tablosunun satırlarını oluşturur. ($i = 1,2,...,I$; $j = 1,2,...,J$)

Çizelge 2.8. Çarpımsal çokterimli dağılım için I*J tablosu

		Kategoriler				
Örnekler	n_{ij}	1	2	…	J	Toplam
	1 2 . . I	n_{11} n_{21} . . n_{I1}	n_{12} n_{22} . . n_{I2}	… … … … …	n_{1J} n_{2J} . . n_{IJ}	$n_{1.}$ $n_{2.}$ . . $n_{I.}$
Toplam		$n_{.1}$	$n_{.2}$	…	$n_{.J}$	$n_{..}$

i.çokterimli örnekten $n_{i1}, n_{i2}, ..., n_{ij}$ verisini elde etme olasılığı Eş. 2.13'de verilmiştir [1].

$$\frac{n_{i.}!}{\prod_{j=1}^{J} n_{ij}!} \prod_{j=1}^{J} p_{ij}^{n_{ij}} \tag{2.13}$$

Her bir satır birbirinden bağımsız olduğu için, n_{ij} değerlerinin tümünü elde etme olasılığı Eş. 2.14'deki gibidir [1].

$$\prod_{i=1}^{I} \left[\frac{n_{i.}!}{\prod_{j=1}^{J} n_{ij}!} \prod_{j=1}^{J} p_{ij}^{n_{ij}} \right] \tag{2.14}$$

Eğer p_{ij} olasılıklarını bilirsek, n_{ij} değerlerinin herhangi bir setini elde etme olasılığını bulabiliriz. p_{ij} olasılıklarını bilmiyoruz fakat n_{ij} değerlerini gözlemlediğimiz için biliyoruz. Eğer Eş. 2.14'ü p_{ij}'lerin fonksiyonu olarak düşünürsek, p için olabilirlik fonksiyonu Eş. 2.15'e eşit olur.

$$L(p) = \prod_{i=1}^{I} \left[\frac{n_{i.}!}{\prod_{j=1}^{J} n_{ij}!} \prod_{j=1}^{J} p_{ij}^{n_{ij}} \right] \tag{2.15}$$

Burada $p = (p_{11}, p_{12}, ..., p_{IJ})$'dir. Amacımız p'yi yani $L(p)$'yi maksimum yapan $\hat{p}$ tahmin değerini bulmaktır. $\hat{p}$, p'nin maksimum olabilirlik tahmini (MLE) olarak adlandırılır.

Çarpımsal çokterimli dağılım için, log-olabilirlik fonksiyonu Eş. 2.16'da verilmiştir.

$$\log L(p) = \sum_{i=1}^{I} \left[\log(n_{i.}!) - \sum_{j=1}^{J} \log(n_{ij}!) + \sum_{j=1}^{J} n_{ij} \log p_{ij} \right] \tag{2.16}$$

Daha önce de belirtildiği gibi amacımız p'nin fonksiyonunu maksimumlaştırmaktır bu yüzden $p$ye bağlı olmayan ifadeleri ihmal edebiliriz. Dolayısıyla, $l(p) = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} \log(p_{ij})$ 'yi ele almak yeterlidir. Burada $\hat{p}_{ij} \equiv n_{ij} / n_{i.}$ dir ve $p = \hat{p}$ olduğunda amacımıza ulaşırız. Beklenen hücre frekansı $m_{ij} = n_{i.} p_{ij}$ olduğu için maksimum olabilirlik yöntemi sayesinde m_{ij} değerlerini elde edebiliriz. Tahmin edilen $\hat{p}_{ij}$ ler ile m_{ij} 'nin MLE'si $\hat{m}_{ij} = n_{i.} \hat{p}_{ij} = n_{ij}$ olur [1].

2.3.4. Olabilirlik oran G^2 testi

Olabilirlik oran istatistiği, iki model arasından hangisinin veriye daha uygun olduğunu test eder. Test, olabilirlik oranına dayanır. Bu oran iki farklı setteki parametrelerin değişimine bağlı olan olabilirlik fonksiyonlarının oranıdır.

Hipotezlerimiz aşağıdaki gibi kurulur.

$H_0 : \theta = \theta_0$

$H_1 : \theta = \theta_1$

Olabilirlik oranı basitçe Eş. 2.17'deki gibidir ve "G^2" ile gösterilir.

$$G^2 = \frac{L(\theta_0 | x)}{L(\theta_1 | x)} = \frac{f(x|\theta_0)}{f(x|\theta_1)} \tag{2.17}$$

Olabilirlik oranı ayrıca $H_0 : p_{1j} = p_{2j} = ... = p_{IJ}$'ın doğru olup olmadığının test edilmesinde de kullanılır $(J = 1,2,..., j)$. $\pi_j = p_{1j} = p_{2j} = ... = p_{IJ}$ olacak şekilde Eş. 2.16'daki log-olabilirlik fonksiyonu aşağıdaki gibi tekrar yazılabilir.

$$\log L(p) = \sum_{i=1}^{I} [\log(n_{i.}!) - \sum_{j=1}^{J} \log(n_{ij}!) + \sum_{J=1}^{J} n_{ij} \log \pi_j] \tag{2.18}$$

p_{ij}'leri içermeyen ifadeleri eşitlikten çıkardığımızda $\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} \log \pi_j$ ifadesini maksimumlaştırmaya çalışırız. Dolayısıyla maksimum olabilirlik tahminleri $\pi_j = \hat{p}_{ij}^{(0)} = \hat{\pi}_j = n_{.j} / n..$ olur. Burada $\hat{p}^{(0)}, H_0$ hipotezinin doğruluğu altında elde edilen tahmindir. Beklenen değer $m_{ij} = n_{i.} p_{ij}$ olduğu için $H_0 : p_{1j} = p_{2j} = ... = p_{Ij}$ hipotezinin doğruluğu altında beklenen değer tahmini $\hat{m}_{ij}^{(0)} = n_{i.} \hat{p}_{ij}^{(0)} = n_{i.} n_{.j} / n..$ olur [1].

Eğer p_{ij}'lere bir kısıt getirilirse veriyi elde etmedeki olabilirliğimiz azalır. Eğer bu kısıt olabilirliği çok azaltıyorsa, p_{ij}'ler üzerindeki bu kısıtın geçerli olmadığı sonucuna varırız. Olabilirlikteki bu azalma, kısıtlamaya tabi olan $L(p)$'nin maksimumunun, kısıt olmayan tüm modelin $L(p)$'nin maksimumuna bölümüne

bakarak ölçülür. Eğer oran çok küçülürse, p_{ij}'ler üzerinde geçerli olan kısıtlama reddedilir.

Kısıtlama getirilmiş $H_0 : p_{1j} = p_{2j} = ... = p_{Ij}$ hipotezinin doğruluğu altında log-olabilirliğin maksimum değeri

$$\log L(\hat{p}^{(0)}) = \sum_{i=1}^{I}\left[\log(n_{i.}!) - \sum_{j=1}^{J}\log(n_{ij}!) + \sum_{j=1}^{J} n_{ij}\log(n_{.j}/n_{..}) \right] \text{'dir.} \qquad (2.19)$$

Tüm modelin log-olabilirliğin maksimumu ise

$$\log L(\hat{p}) = \sum_{i=1}^{I}\left[\log(n_{i.}!) - \sum_{j=1}^{J}\log(n_{ij}!) + \sum_{j=1}^{J} n_{ij}\log(n_{ij}/n_{i.}) \right] \text{'dir.} \qquad (2.20)$$

Log-olabilirliklerin farkı,

$$\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log(n_{.j}/n_{..}) - \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log(n_{ij}/n_{i.}) \quad \text{olur.} \qquad (2.21)$$

Bu eşitliği "-2" ile çarptığımızda Eş. 2.22'de belirtilen olabilirlik oran test istatistiğini elde ederiz [1].

$$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J} \hat{m}_{ij}\log\left(\frac{\hat{m}_{ij}}{\hat{m}_{ij}^{(0)}}\right) \qquad (2.22)$$

Burada kısıtlama olmayan modelin beklenen değeri m_{ij}'nin maksimum olabilirlik tahmini $\hat{m}_{ij} = n_{ij}$, "H_0'ın doğru olduğu" kısıtlaması getirilen modelin beklenen değeri m_{ij}'nin maksimum olabilirlik tahmini $\hat{m}_{ij}^{(0)} = n_{i.}n_{.j}/n..$ 'dir.

Log olabilirlik farkı eşitliğini "-2" sayısı ile çarpmamızın nedeni, büyük çaplı örnekler için $G^2 \sim \chi^2((I-1)(J-1))$ yaklaşımını sağlamaktır. H_0 hipotezi olabilirlik oranının küçük değerleri için reddedilebilir fakat logaritması alınıp "-2" ile çarpıldıktan sonra G^2'nin büyük değerleri için de reddedilebilir hale gelir. Yani büyük çaplı örnekler için α seviyesinde $G^2 > \chi^2(1-\alpha,(I-1)(J-1))$ olduğunda H_0 hipotezi reddedilir.

Genel olarak belirtirsek olabilirlik oran istatistiği G^2, χ^2 dağılımına sahiptir ve serbestlik derecesi, iki modelin serbestlik derecelerinin farkına eşittir.

3. İKİ BOYUTLU TABLOLARDA LOG-LİNEER MODELLER VE LOJİSTİK REGRESYON

3.1. Log-Lineer Modeller

Log-lineer modeller ele alınan tüm değişkenler kategorik olduğunda özellikle veriyi analiz etmek için tasarlanmış çok değişkenli istatistiksel tekniklerden biridir.

Log-lineer modeller ile

1) İki ya da daha fazla kategoriye sahip değişkenler arasındaki ikili ya da daha yüksek dereceli ilişkiler ve etkileşimler, bağımlı-bağımsız değişken ayrımı yapılarak veya yapılmadan saptanabilir [18].

2) Her bir bağımsız değişkenle her bir bağımlı değişken arasındaki ilişkinin derecesi, geriye kalan bağımlı ve bağımsız değişkenlerin etkisi sabitlenerek ölçülür [7].

Log-lineer modellerin tercih edilmesinin diğer bir nedeni de çok değişkenli olumsallık tablolarına uygulanan ki-kare bağımsızlık testlerinin uygulanmasındaki zorluktur. Değişkenlere ait sınıf sayıları arttıkça, sınıfların yer aldığı satır ve sütunları birbirleriyle karşılaştırmak oldukça güçleşmektedir. Bu durumda çok değişkenli olumsallık tablolarının satır ve sütun sayılarına kısıtlama getirmeyen, aynı tablo üzerinde ki-kareye oranla çeşitli hipotezlerin test edilmesine olanak sağlayan log-lineer modeller tercih edilir [14].

3.1.1. Log-lineer modellerin türetilmesi

$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}$ ANOVA modelini ele alalım. Parametrelerin sembollerini değiştirerek model Eş. 3.1'deki gibi tekrar yazılabilir. Burada hata terimi olan e_{ijk}'ler bağımsız $N(0,\sigma^2)$ olarak dağılırlar.

$$y_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} + e_{ijk} \qquad (3.1)$$

$$i = 1,\ldots,I \;,\; j = 1,\ldots,J \;,\; k = 1,\ldots,K$$

Eş. 3.1'de verilen ANOVA modelinde y_{ijk}' lar bağımsızdır ve $y_{ijk} \sim N(m_{ij}, \sigma^2)$ dağılır.

Beklenen değer için,

$$m_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \quad \text{'dir.} \qquad (3.2)$$

Amacımız m_{ij}'nin yapısını açıklamaktır ve m_{ij}'nin maksimum olabilirlik tahmin edicisi $\bar{y}_{ij.}$'dir [1].

$\hat{m}_{ij} \sim N(m_{ij}, \sigma^2 / K)$ dağılımına sahiptir ve her bir hücrede K sayıda gözlem vardır. σ^2 'nin tahmin edicisi ortalama hata kare (OHK) dir ve $\hat{m}_{ij}$ 'den bağımsızdır. Ancak OHK , σ^2' nin maksimum olabilirlik tahmin edicisi değildir.

Eğer m_{ij}'ye modelde hiç etkileşim olmadığı kısıtı getirilirse Eş. 3.2, Eş. 3.3'e dönüşür.

$$m_{ij} = u + u_{1(i)} + u_{2(j)} \qquad (3.3)$$

Ve m_{ij} 'nin tahmin edicisi Eş. 3.4'deki gibidir.

$$\hat{m}_{ij} = \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) \qquad (3.4)$$

Burada m_{ij}'nin maksimum olabilirlik tahmini ve dolayısıyla σ^2'nin maksimum olabilirlik tahmini değişecektir.

Daha önce de belirtildiği gibi biz m_{ij}'nin yapısı ile ilgileniyoruz. Buna rağmen Eş. 3.2 ve Eş. 3.3'teki gibi lineer modelleri değerlendirmek yerine

$$\log(m_{ijk}) = \lambda + \lambda_{1(i)} + \lambda_{2(j)} \tag{3.5}$$

$$\log(m_{ijk}) = \lambda + \lambda_{1(i)} + \lambda_{2(j)} + \lambda_{12(ij)} \tag{3.6}$$

gibi log-lineer modelleri ele alacağız. Analizimiz yine m_{ij}'nin maksimum olabilirlik tahminine ve olabilirlik oran testine dayanır. Buna rağmen bazı farklılıklar mevcuttur. Burada n_{ij}'ler çokterimli ya da çarpımsal çokterimli dağılıma sahiplerdir [1].

3.1.2. İki boyutlu olumsallık tablolarında lambda parametrelerinin elde edilmesi

Bir birimin A değişkeninin $i.$ satırına ve B değişkeninin $j.$ sütununa düşme olasılığı p_{ij} olsun. A ile B değişkenleri bağımsız ise;

$p_{ij} = p_{i.}p_{.j}$'dir. $(i = 1,2,...,I \ , \ j = 1,2,...,J)$ (3.7)

Eş. 3.7'de çarpımsal olarak ifade edilen modelin logaritması alınarak log-lineer modeller elde edilir [13].

$$\log p_{ij} = \log p_{i.} + \log p_{.j} \tag{3.8}$$

Beklenen frekans m_{ij} olmak üzere

$$m_{ij} = np_{ij} = np_{i.}p_{.j} \quad \text{dir.} \tag{3.9}$$

n_{ij} gözlenen frekans, n örneklem hacmini göstermek üzere, m_{ij}'nin maksimum olabilirlik tahmin edicisi Eş. 3.10'daki gibidir.

$$\hat{m}_{ij} = \frac{n_{i.} n_{.j}}{n} \tag{3.10}$$

Eş. 3.11'de beklenen frekanslar, satır ve sütun etkilerinin doğal logaritmalarının toplamı şeklinde ifade edilmektedir. Eş. 3.9 ile düzenlenen Eş. 3.11'deki model, iki değişkenli frekans tabloları için genel log-lineer model eşitliğidir [13].

$$\log m_{ij} = \log n + \log p_{i.} + \log p_{.j} \tag{3.11}$$

Benzer şekilde satır ve sütunlar için marjinal beklenen frekanslar aşağıdaki gibidir.

$$m_{i.} = np_{i.}$$
$$m_{.j} = np_{.j} \tag{3.12}$$

e tabanında logaritmaları alındığında:

$$\log m_{i.} = \log n + \log p_{i.}$$
$$\log m_{.j} = \log n + \log p_{.j} \quad \text{olur.} \tag{3.13}$$

Eş. 3.13 aşağıdaki şekilde de ifade edilebilir.

$$\log p_{i.} = \log m_{i.} - \log n$$
$$\log p_{.j} = \log m_{.j} - \log n \tag{3.14}$$

Eş. 3.14'de elde edilen çift taraflı logaritmik toplamlar Eş. 3.11'de yerine yazıldığında Eş. 3.15 elde edilir [13].

$$\log m_{ij} = \log m_{i.} + \log m_{.j} - \log n \tag{3.15}$$

Eş. 3.15'de her $i = 1, 2, ..., I$ ve $j = 1, 2, ..., J$ için eşitliğin her iki tarafından toplam alındığında Eş. 3.16 elde edilir.

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\log m_{ij} = J\sum_{i=1}^{I}\log m_{i.} + I\sum_{j=1}^{J}\log m_{.j} - IJ\log n \tag{3.16}$$

$$\lambda_i = \log m_{i.} - \frac{1}{I}\sum_{i=1}^{I}\log m_{ij}$$

$$\lambda_j = \log m_{.j} - \frac{1}{J}\sum_{j=1}^{J}\log m_{ij}$$

$$\lambda_0 = \frac{1}{I}\sum_{i=1}^{I}\log m_{i.} - \frac{1}{J}\sum_{j=1}^{J}\log m_{.j} + \log n \tag{3.17}$$

olmak üzere $\lambda_i, \lambda_j, \lambda_0$ şeklinde üç parametre tanımlanıp, Eş. 3.16'da yerine konursa aşağıdaki eşitlik elde edilir.

$$\log m_{ij} = \lambda_0 + \lambda_I + \lambda_J \tag{3.18}$$

Bu model *bağımsız* log-lineer model olarak adlandırılır. İki değişken arasında etkileşim olmadığında geçerli olan log-lineer modele aittir [13].

Değişkenler arasındaki etkileşim anlamlı ise,

$$\log m_{ij} = \lambda_0 + \lambda_I + \lambda_J + \lambda_{IJ} \tag{3.19}$$

olarak yazılır ve bu model, *doymuş* log-lineer model olarak adlandırılır [11].

A ve B değişkenleriyle bağımsız model aşağıdaki gibi de gösterilebilir.

$$\log m_{ij} = \lambda_0 + \lambda_I{}^A + \lambda_J{}^B \tag{3.20}$$

Burada,

λ_0 : beklenen frekansların doğal logaritmasının ortalaması

$\lambda_I{}^A$: A kategorik değişkeninin beklenen frekans üzerindeki ana etkisini

λ_j^B: B kategorik değişkeninin beklenen frekans üzerindeki ana etkisini göstermektedir.

A ve B değişkenleriyle *doymuş* log-lineer model ise aşağıdaki gibi ifade edilir. Burada λ_{ij}^{AB}= satır ve sütun değişkeninin i inci ve j inci düzeyinin beklenen frekans üzerindeki etkisini ifade eder [14].

$$\log m_{ij} = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \tag{3.21}$$

3.1.3. İki boyutlu olumsallık tablolarında hipotezler

A ve B kategorik değişkenleri arasında etkileşim olup olmadığını incelemek, başka bir ifadeyle ikili etkileşimin anlamlılığını test etmek için Eş. 3.22'deki hipotez kurulur.

$$H_0 : \lambda_{ij}^{AB} = 0 \qquad (i = 1,2,...,I \quad j = 1,2,...,J) \tag{3.22}$$

Etkileşimin anlamlılığını test etmek için $(I-1)(J-1)$ serbestlik dereceli Pearson Ki-kare ve Olabilirlik Oran test istatistikleri kullanılır [35]. İki değişkenli logaritmik doğrusal model için model parametreleri ve serbestlik dereceleri Çizelge 3.1'de gösterilmiştir.

Çizelge 3.1. İki değişkenli log-lineer model için model parametreleri ve serbestlik dereceleri

Parametreler	S.D
λ_i^A	$(I-1)$
λ_J^B	$(J-1)$
λ_{iJ}^{AB}	$(I-1)(J-1)$
λ_0	1
Toplam	IJ

3.1.4. Log-lineer modellerde odds oranı

Bölüm 2.1'de odds oranı için yapılan açıklamalar, log-lineer modellerde $\log m_{ij}$'ler için de geçerlidir. Bildiğimiz gibi, odds oranının 1'e eşit olması etkileşimin sıfıra eşit olduğunu belirtir ve odds oranları beklenen değer terimleri ile de yazılabilir.

Çarpımsal çokterimli örnekleme için, $m_{ij} = n_{i.} p_{ij}$ ve çokterimli örnekleme için $m_{ij} = n_{..} p_{ij}$ dir.

Her bir durum için odds oranı aşağıdaki gibi hesaplanır.

$$\theta = \frac{p_{ij} p_{i'j'}}{p_{ij'} p_{i'j}} = \frac{m_{ij} m_{i'j'}}{m_{ij'} m_{i'j}} \tag{3.23}$$

3.2. Lojistik Regresyon

Lojistik regresyon kategorik veri analizinde bağımlı değişken nitel olduğunda en çok tercih edilen istatistiksel yöntemlerden biridir. Lojistik regresyon; bağımlı değişkenin ikili, üçlü ve çoklu kategorilerinde bağımsız değişkenlerle sebep-sonuç ilişkisini inceler [10].

1970'den beri istatistikçiler lojistik regresyonun iki uygulamasını ele almışlardır.

1- Bağımlı değişken için grup üyelerinin tahmini

2- Bir olayın olma olasılığını, olmama olasılığını ve odds oranlarının belirlenmesi [7].

3.2.1. Lojistik model ve lojistik fonksiyon

Lojistik regresyona ihtiyaç duyulmasındaki esas neden bağımlı değişkenin sürekli dağılmadığı durumda normallik varsayımının sağlanamamasıdır. Lojistik regresyon analizi normal dağılım ve süreklilik varsayımları gibi önkoşul gerektirmeyen bir regresyon yöntemidir [37].

Lojistik regresyonda bağımlı değişken nitelken, bağımsız değişkenler hem nicel hem nitel olabilir. Bağımsız değişkenlerin tümünün kategorik değişken olması durumunda lojistik regresyon "lojit" model adını almaktadır. Ayrıca lojistik regresyon bağımlı değişkenin sahip olduğu kategori sayısına göre adlandırılır. Eğer bağımlı değişken iki değer alıyorsa "binary (iki değerli) lojistik regresyon", ikiden fazla değer alıyorsa "multinomial (çoklu) lojistik regresyon" adını alır.

Lojistik fonksiyon aşağıdaki gibidir. Burada $Z = \sum_{k=0}^{p} \beta_k x_{ik}$ 'dir.

$$f(z) = \frac{1}{1+e^{-z}} \qquad (3.24)$$

$f(z)$, bağımsız değişken x'in tanım aralığında 0-1 aralığının dışına çıkmamaktadır [8].

Herhangi bir regresyon probleminde en önemli nicelik, bir bağımsız değişkenin değeri verildiğinde sonucun ortalama değeridir. Bu nicelik "koşullu ortalama" diye adlandırılır ve "$E(Y/X)$" şeklinde gösterilir. $E(Y/X)$,"Verilen X değeri için, Y'nin beklenen değeri" diye okunur [10].

İki değer alan (binary) y_i değişkeninin doğrusal olasılık modeli aşağıdaki gibidir ve 0 ile 1 arasında olasılık değerleri alır.

$$E(y_i) = P(y_i = 1 | X_i) = \sum_{k=0}^{p} \beta_k x_{ik} \text{ 'dir.} \quad (3.25)$$

Burada $X_i, (x_{i1}, x_{i2}, ..., x_{ip})$ şeklinde p tane x_{ik} bağımsız değişkeni içeren satır vektörüdür. Lojistik regresyonda kesikli veya sürekli olabilen ve $[-\infty, \infty]$ aralığında değer alabilen bağımsız x_{ik} değişkenleri ile yalnızca 0 ile 1 değeri alan y_i değişkeni arasındaki ilişki ele alındığından bu eşitlik her zaman gerçekleşmemektedir. Çözüm olarak $P(y_i = 1 | X_i)$ olasılık değerlerini $[-\infty, \infty]$ aralığına dönüştürmek gerekir [36].

İlk önce $\frac{P(X_i)}{1 - P(X_i)}$ dönüşümü ile değişim aralığı $[0, \infty]$ aralığına, daha sonra da logaritması alınarak $[-\infty, \infty]$ aralığına dönüştürülür. Bu dönüşüm işlemine "lojit dönüşüm" denilmektedir. Dolayısıyla doğrusal olasılık modeli Eş. 3.26'ya dönüşür.

$$\log it(P_i) = \ell(X_i) = E(y_i | X_i) = \log \frac{P(X_i)}{1 - P(X_i)} = \sum_{k=0}^{p} \beta_k x_{ik} \quad (3.26)$$

Bazen $P(X_i)$ kısaca P_i olarak da gösterilebilmektedir. Bu eşitlik "lojit dönüşüm modeli" olarak adlandırılır. Burada x_i bilinen bir tahmin edici, β_k'lar bilinmeyen parametrelerdir. $P(X_i)$ olasılığı da Eş. 3.27 ile elde edilir ve $\frac{P(X_i)}{1 - P(X_i)}$ odds oranıdır.

$$P(X_i) = \frac{e^{\sum_{k=0}^{p} \beta_k x_{ik}}}{1 + e^{\sum_{k=0}^{p} \beta_k x_{ik}}} = \frac{e^{\ell(x_i)}}{1 + e^{\ell(x_i)}} \quad (3.27)$$

Eş. 3.27 ile belitilen eşitliğe "lojistik fonksiyon" denir [36] . Bağımlı değişken y_i için hata terimi $\varepsilon_i \sim binom(0, P(X_i)(1-P(X_i))$ dağılır.

Lojit dönüşüm sayesinde odds'un logaritması $\log it(P_i)$, sadece X_i ye göre değil, yığın parametre tahminleri β_k'lara göre de doğrusallaşır. Özetle;

1) P_i, 0'dan 1'e giderken, $\log it(P_i)$ de $-\infty$'dan $+\infty$'a doğru değişir. Yani olasılıklar 0 ile 1 arasında değer alırken, $\log it(P_i)$ sınırlı değildir [9].

2) $\log it(P_i)$, X_i'e göre doğrusalken olasılıklar değildir.

3) Lojit dönüşüm modelinde β_k , X_i teki bir birim değişmeye karşılık log odds oranının nasıl değiştiğini belirtir.

4) P_i olasılığının tahminini (Bkz. Eş. 3.28) elde edebilmek için önce (β_k) bilinmeyen parametreleri tahmin edilmelidir.

$$\hat{P}_i = \frac{e^{\sum_{k=0}^{p} \hat{\beta}_k x_{ik}}}{1 + e^{\sum_{k=0}^{p} \hat{\beta}_k x_{ik}}} \tag{3.28}$$

3.2.2. Parametre tahmin yöntemleri

Lojistik modelin parametrelerinin tahmini için bu çalışmada maksimum olabilirlik tahmin edicisi kullanılacaktır. Diğer yöntemler ise "yeniden ağırlıklandırılmış iteratif en küçük kareler yöntemi" ve "minimum lojit ki-kare" yöntemidir [19].

Basit lojistik regresyon için maksimum olabilirlik tahmini:

Basit lojistik regresyon modeli yalnızca bir tane bağımsız değişkenin bulunduğu modeldir.

Lojit dönüşüm modeli: $lojit(P_i) = \ell(X_i) = \log(\frac{P(X_i)}{1-P(X_i)}) = \beta_0 + \beta_1 x_{i1}$

Lojistik fonksiyon: $P(X_i) = \frac{e^{\ell(x_i)}}{1+e^{\ell(x_i)}} = \frac{e^{\beta_0+\beta_1 x_{i1}}}{1+e^{\beta_0+\beta_1 x_{i1}}}$ (3.29)

şeklinde yazılabilir [19].

Parametrelerin maksimum olabilirlik tahmin edicilerinin, fonksiyonu maksimum yapan değerleri bulacak şekilde seçildiği Bölüm 2.3'de belirtmiştir. Elde edilen tahminler, gözlenen verilere çok yakın değerlere sahiptir.

$P(x)$ ifadesi, verilen herhangi bir x için y 'nin 1'e eşit olma koşullu olasılığını yani [$P(x) = P(y=1|x)$] 'yi,

$1-P(x)$ ifadesi, verilen herhangi bir x için y 'nin 0'a eşit olma koşullu olasılığını yani $\left[1-P(x) = P(y=0|x)\right]$ 'yi göstermektedir.

(x_i, y_i) çiftinin $y_i = 1$ olduğunda olabilirlik fonksiyonuna katkısı $P(x_i)$ iken , $y_i = 0$ olduğunda olabilirlik fonksiyonuna katkısı $1-P(x_i)$ kadardır. Diğer bir ifadeyle Eş. 3.30'daki gibidir.

$$P(y_i / x_i) = P(x_i)^{y_i}(1-P(x_i))^{1-y_i} \quad (3.30)$$

Olabilirlik fonksiyonu da Eş. 3.31'deki gibidir.

$$L(y|x,\beta) = \prod_{i=1}^{n} P(x_i)^{y_i}[1-P(x_i)]^{1-y_i} \quad \text{dir.} \quad (3.31)$$

Log-olabilirlik fonksiyonu Eş. 3.32'de verilmiştir [25].

$$\log L(y|x,\beta) = \sum_{i=1}^{n} \left\{ \left[y_i \log P(x_i) + (1-y_i) \log n[1-P(x_i)] \right] \right\} \quad (3.32)$$

İstatistiksel paket programlar yardımıyla iteratif yöntemler uygulanarak bu fonksiyonu maksimum yapan parametre tahminleri elde edilir.

Çoklu lojistik regresyon için maksimum olabilirlik tahmini:

Uygulamada genelde birden fazla bağımsız değişkenin olduğu modeller ele alınır. Lojistik regresyonda bağımsız değişkenler hem nitel hem nicel özelliklere sahip olabilir. Bu özelliğiyle başlıca tercih sebebidir.

$x = \left(x_1, x_2, ..., x_p \right)$ şeklinde p tane bağımsız değişken için lojit dönüşüm modeli ve lojistik fonksiyon aşağıdaki gibidir.

Lojit dönüşüm modeli: $\log it(P_i) = \ell(X_i) = \log_e \dfrac{P(X_i)}{1-P(X_i)} = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}$ (3.33)

Lojistik fonksiyon: $P(X_i) = \dfrac{e^{\ell(x_i)}}{1+e^{\ell(x_i)}}$ (3.34)

Yine aynı şekilde istatistiksel paket programlar yardımıyla iteratif yöntemler uygulanarak Eş. 3.33'deki fonksiyonu maksimum yapan parametre tahminleri elde edilir.

Eğer çoklu lojistik regresyonda sınıflama ya da sıralama düzeyinde ölçülmüş bağımsız değişkenler varsa bu değişkenler için kukla değişken kullanılmalıdır. x in r tane sınıfı varsa, r-1 tane kukla değişkeni modelde olmak zorundadır.

3.2.3. Lojistik regresyonda Hosmer-Lemeshow uyum iyiliği testi

Hosmer-Lemeshow uyum iyiliği testi, lojistik regresyon modeli için uyum iyiliğinin istatistiksel bir testidir. Özellikle küçük çaplı örnekler ve sürekli dağılan bağımsız değişken içeren modeller için daha güçlü bir uyum iyiliği testidir. Hosmer ve Lemeshow gözlemleri, tahmin edilen olasılıklarına göre 10 eşit gruba bölmeyi önermiştir. Test (n-2) serbestlik dereceli χ^2 dağılır. Büyük χ^2 değerleri yani küçük p değerleri zayıf uyumu, küçük χ^2 değerleri iyi uyumu belirtir [25].

$$G_{HL}^2 = \sum_{j=1}^{10} \frac{(O_j - E_j)^2}{E_j(1 - E_j / n_j)} \tag{3.35}$$

Burada,

n_j = j. gruptaki gözlemlerin sayısıdır.

$0_j = \sum y_{ij}$ = j. gruptaki olayların sayısıdır.

$E_j = \sum \hat{p}_{ij}$ = j. gruptaki olayların sayısıdır.

3.2.4. Lojistik regresyonda katsayıların önem testi

Hem basit hem de çoklu regresyonda katsayılar hesaplandıktan sonra katsayıların önemliliği Wald testi ve olabilirlik oran testi ile test edilir. Bu çalışmada SPSS sonuçlarına göre Wald testi kullanılacaktır [8].

Wald testi

Wald testi, Abraham Wald tarafından önerilen parametrik bir testtir. Değişkenler arasındaki bir ilişki, örneklerden tahmin edilen parametrelerin olduğu istatistiksel bir model ile açıklandığında; Wald testi bu parametrelerin gerçek değerlerini test etmek için kullanılabilir. Örneğin sosyal sınıfı ve sahip olduğu ayakkabı sayısı belli olan veri için, sosyal sınıf ile ayakkabı sayısı arasında anlamlı bir ilişki olup olmadığı

araştırılmak istensin. β, orta sınıf insanlara kıyasla yüksek sınıf insanların ayakkabı sayısındaki ortalama artış olsun, o zaman Wald testi β'nın sıfır olup olmadığını belirlemek için kullanılabilir. Burada β, yüksek ve orta sınıf insanların sahip olduğu ayakkabı sayısındaki hipotetik (varsayımsal) farklılıktır. Örneğin sağlık araştırmaları için sigaranın akciğer kanseri yapma riski R ile gösterilsin. Wald testi R=1 olup olmadığını test etmek için de kullanılabilir [40].

Wald testi için, β parametresinin maksimum olabilirlik tahmini $\hat{\beta}$, önerilen β_0 ile karşılaştırılır. İkisi arasındaki farkın yaklaşık normal dağıldığı varsayılır.

Wald istatistiği: $\frac{\left(\hat{\beta}-\beta_0\right)^2}{\text{var}(\hat{\beta})}$'dir. (3.36)

Farkın karesi χ^2 dağılımı ile karşılaştırılır. Alternatif olarak, fark Normal dağılım ile de karşılaştırılabilir. Bu durumda Wald testi $\frac{\hat{\beta}-\beta_0}{sh(\hat{\beta})}$'dir. Burada $sh(\hat{\beta})$,maksimum olabilirlik tahmininin standart hatasıdır. SPSS paket programında lojistik regresyon için tahmin edilen parametreler B ile gösterilmektedir ve katsayıların testi Wald B testi olarak bilinmektedir.

Bölüm 2.3.4'de anlatılan olabilirlik oran testi ile Wald testi genelde çok benzer sonuçlara sahiptir, nadiren de olsa farklı sonuçlar çıkabilmektedir. Uygulamada olabilirlik oran testi, Wald testine göre daha çok tercih edilir ancak biz bu çalışmada SPSS sonuçlarına göre Wald testini dikkate alacağız. Bunun nedenlerinden biri, sigara içenlerin akciğer kanseri olma riski R için olabilirlik oran testi; R, log R ya da R'nin herhangi bir monotimik dönüşümünde aynı sonucu verir. Fakat Wald testinde elde edilen sonuçlar aynı değildir çünkü R'nin ya da Log R'nin standart hataları arasında basit bir ilişki yoktur. Diğer bir sebep ise, Wald testinde iki yaklaşım kullanılırken (standart hatalar ve χ^2) , olabilirlik oran testinde sadece χ^2 yaklaşımı vardır [20-41].

3.3. İki Boyutlu Tablolarda Lojit Model ve Log-Lineer Modellerin Denkliği

Bu bölümde iki boyutlu tablolar için lojit modeller ile log-lineer modellerin ilişkisi incelenecektir. Bu modeller ayrı gibi görünseler de birbiri ile ilişkilidir. Log-lineer modellerin yorumlanmasında lojit modeller temel alınabilmektedir. Log-lineer modellerde değişkenlerden birinin iki kategorili olması durumunda modelin parametreleri lojit model yardımıyla kolayca yorumlanmaktadır [3].

y_i bağımlı değişkeninin 1 ve 2 değerlerini aldığı kabul edilsin. Buna göre y_i değişkeninin 1 değerini alma olasılığı p, 2 değerini alma olasılığı 1-p dir ve Bernoulli dağılımına sahiptir. Y, k tane kategoriye ayrılarak incelendiğinde ve her bir kategoride yer alan birim sayısı N_i (i=1,2,…,k) dikkate alındığında ise Binom dağılımına sahip olur. Basit lojistik regresyonu ele alalım, her biri $y_i \sim B(N_i, p_i)$ dağılan bağımsız gözlemlerimiz olsun. Burada $E(y_i) = m_i = N_i.p_i$ ve odds $\frac{p_i}{1-p_i}$'dir. Lojit model log odds için lineer bir yapı belirtir ve Eş. 3.37'de belirtilmiştir.

$$lojit(P_i) = \log(\frac{P_i}{1-P_i}) = \beta_0 + \beta_1 x_i \qquad i = 1,...,I \qquad (3.37)$$

Log oddslar için lojit model, log-lineer modeller ile değerlendirilebilir. Örnekleme planı Bölüm 2.2'de de belirtildiği gibi çarpımsal çokterimlidir. Çarpımsal çokterimli örnekleme altında aşağıdaki varsayım geçerli olur.

$y_i \sim Bin(N_i, p_i)$ dağılımına sahip olduğundan $(N_i - y_i) \sim B(N_i, 1-p_i)$ dağılımına sahiptir. İki yönlü tablo için , $y_i \equiv n_{i1}$ ve $N_i - y_i \equiv n_{i2}$ dir. Tüm i ler için $N_i \equiv n_{i.}$ kabul edilir ve beklenen değerler için benzer tanımlamalar ile $m_i = N_i p_i = n_{i.} p_{i1} = m_{i1}$ ve $m_{i2} = N_i(1-p_i)$ olur. Böylece, $\frac{p_i}{1-p_i} = \frac{m_{i1}}{m_{i2}}$ varsayımı sağlanır. Burada $p_i \equiv p_{i1}$ ve $1-p_i \equiv p_{i2}$'dir [1].

Eş. 3.37 için log-lineer versiyon,

$$\log(m_{ij}) = \lambda_{1(i)} + \lambda_{2(j)} + \eta_j x_i \qquad (3.38)$$

şeklindedir. $\eta_j x_i$ terimindeki x_i bilinen fakat η_j bilinmeyen parametredir. Bu terim, $\lambda_{12(ij)}$ gibi etkileşim terimidir çünkü, *i*, *j* alt simgelerinin ikisini de içermektedir. Lojit model (Eş. 3.37) ile log-lineer model (Eş. 3.38) arasındaki ilişki çarpımsal çokterimli örnekleme varsayımı altında aşağıdaki gibidir:

$$\begin{aligned}
\log\left(\frac{p_i}{1-p_i}\right) &= \log\left(\frac{m_{i1}}{m_{i2}}\right) \\
&= \log(m_{i1}) - \log(m_{i2}) \\
&= [\lambda_{1(i)} + \lambda_{2(1)} + \eta_1 x_i] - [\lambda_{1(i)} + \lambda_{2(2)} + \eta_2 x_i] \\
&= [\lambda_{2(1)} - \lambda_{2(2)}] + [\eta_1 x_i - \eta_2 x_i] \\
&\equiv \beta_0 + \beta_1 x_i \qquad [1] \qquad (3.39)
\end{aligned}$$

Burada $\beta_0 = [\lambda_{2(1)} - \lambda_{2(2)}]$ ve $\beta_1 \equiv [\eta_1 - \eta_2]$'dir. Maksimum olabilirlik yöntemi ile parametreler tahmin edilir.

İki yönlü tablolar için olabilirlik fonksiyonu $L(p)$ Eş.3.40'daki gibidir.

$$L(p) = \prod_{i=1}^{I} \left[\frac{n_{i.}!}{\prod_{j=1}^{J} n_{ij}!} \prod_{j=1}^{J} p_{ij}^{\,n_{ij}} \right] \qquad p = (p_{11}, p_{12}, \ldots, p_{IJ}) \qquad (3.40)$$

Eş. 3.33 ile elde edilen olasılıklar olabilirlik fonksiyonunda yerine koyulursa aşağıdaki denklem elde edilir.

$$L(\beta_0,\beta_1)=\prod_{i=1}^{I}[\frac{n_{i.}!}{\prod_{j=1}^{2} n_{ij}!}\left\{\frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}\right\}^{n_{i1}}\left\{\frac{1}{1+\exp(\beta_0+\beta_1 x_i)}\right\}^{n_{i2}}] \tag{3.41}$$

Bulunan $\hat{\beta}_0$ ve $\hat{\beta}_1$ tahminleri ile

$\hat{p}=\frac{\exp(\hat{\beta}_0+\hat{\beta}_1 x_i)}{1+\exp(\hat{\beta}_0+\hat{\beta}_1 x_i)}$ elde edilebilecektir. (3.42)

Olasılıkların tahminleriyle $\hat{m}_{ij}$ beklenen değerler de elde edilebilir. ($\hat{m}_{ij}=n_{i.}\hat{p}_{ij}$)

3.3.1. İki boyutlu tablolarda katsayıların anlamlılığının testi

$$\log(m_{ij})=\lambda_{1(i)}+\lambda_{2(j)}+\eta_j x_i \tag{3.43}$$

$$\log(m_{ij})=\lambda+\lambda_{1(i)}+\lambda_{2(j)}+\lambda_{12(ij)} \tag{3.44}$$

Eş. 3.43 ile Eş. 3.44'ü karşılaştırmak için olabilirlik oran G^2 ve Pearson χ^2 istatistikleri kullanılır.

Eş. 3.44'deki *doymuş* model için beklenen hücre sayıları n_{ij} lere eşittir. Bu durumda,

$$\begin{aligned} G^2 &= 2\sum_{i=1}^{I}\sum_{j=1}^{2} n_{ij}\log(\frac{n_{ij}}{\hat{m}_{ij}}) \\ &= 2\sum_{i=1}^{I}[n_{i1}\log(n_{i1}/\hat{m}_{i1})+n_{i2}\log(n_{i2}/\hat{m}_{i2})] \\ &= 2\sum_{i=1}^{I}[y_i\log(y_i/N_i\hat{p}_i)+(N_i-y_i)\log((N_i-y_i)/N_i(1-\hat{p}_i))] \end{aligned} \tag{3.45}$$

Ve Pearson χ^2 test istatistiği,

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{2}\frac{(n_{ij}-\hat{m}_{ij})^2}{\hat{m}_{ij}}$$

$$= \sum_{i=1}^{I}[\frac{(y_i - N_i\hat{p}_i)^2}{N_i\hat{p}_i} + \frac{[(N_i - y_i) - N_i(1-\hat{p}_i)]^2}{N_i(1-\hat{p}_i)}]$$

$$= \sum_{i=1}^{I}[\frac{(y_i - N_i\hat{p}_i)^2}{N_i\hat{p}_i} + \frac{(y_i - N_i\hat{p}_i)^2}{N_i(1-\hat{p}_i)}]$$

$$= \sum_{i=1}^{I}\frac{(y_i - N_i\hat{p}_i)^2}{N_i\hat{p}_i(1-\hat{p}_i)}$$ olarak elde edilir [1]. (3.46)

Eş. 3.37'deki lojit modelden yararlanılarak olabilirlik oran G^2 ve Pearson χ^2 testlerinin serbestlik derecesi "N- 2" dir diyebiliriz. Çıkarılan iki sayısı, eğim ve sabit parametreyi temsil eder. Eş. 3.43'e göre de serbestlik derecesi
I*2 – I(satır etkisi ve ana etki için)- 1(sütun etkisi için) -1 (etkileşim etkisi için) 'dir.

Her iki test istatistiği de, katsayıların anlamlılığının testinde kullanılabilir fakat χ^2 dağılımına sahip büyük çaplı örnekler için n_{ij} 'lere ihtiyaç vardır. Genellikle, bir lojistik regresyon, Eş. 3.44'deki gibi doymuş modele karşı χ^2 testi, sadece örnek çapı N_i büyük olduğunda uygundur. Bununla birlikte Eş. 3.43 doymuş model kabul edilerek azaltılmış modellere karşı test edilebilir [1].

Eş. 3.37'deki model ile Eş. 3.38'deki model eşit olduğu için,

$$\log(\frac{p_i}{1-p_i}) = \beta_0 \qquad (3.47)$$

azaltılmış modelini Eş. 3.43'deki modele karşı test etme , $H_0 : \beta_1 = 0$ ı test etmeye eşittir.

Eş. 3.47'deki model'in $\hat{\beta}_0$ tahmini ile $\hat{p}_i = e^{\hat{\beta}_0}/(1+e^{\hat{\beta}_0})$ elde edilir ve $H_0 : \beta_1 = 0$ hipotezi altında beklenen değer $\hat{m}_{ij}^{(0)} = n_{i.}\hat{p}_{ij}$ dir [1]. Beklenen gözlemler ile test istatistiği aşağıdaki gibidir.

$$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{2}\hat{m}_{ij}\log(\frac{\hat{m}_{ij}}{\hat{m}_{ij}^{(0)}}) \tag{3.48}$$

4. ÜÇ BOYUTLU TABLOLARDA LOG-LİNEER MODELLER VE LOJİSTİK REGRESYON

Genel olarak üç boyutlu tabloların gözlem sayıları n_{ijk} olarak belirtilir.

Çizelge 4.1. Üç boyutlu olumsallık tablosu

		(k)=[3]									
		1			2			…	k		
	(j)=[2]	1	…	J	1	…	J	….	1	…	J	Toplam
(i)=[1]	1	n_{11}	…	n_{1j1}	n_{112}	…	n_{1j2}	…	n_{11K}	…	n_{1JK}	$n_{1..}$
	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	I	n_{I1}	…	n_{IJ1}	n_{I12}	…	n_{IJ2}	…	n_{I1K}	…	n_{IJK}	$n_{I..}$
	Toplam	$n_{..1}$			$n_{..2}$				$n_{.k}$			$n_{...}$

$i = 1, 2, ..., I \quad j = 1, 2, ..., J \quad k = 1, 2, ..., K$ olmak üzere marjinal toplamlar aşağıdaki gibidir.

$$n_{ij.} = \sum_{k=1}^{K} n_{ijk} \ , \ n_{i.k} = \sum_{j=1}^{J} n_{ijk} \ , \ n_{.jk} = \sum_{i=1}^{I} n_{ijk}$$

$$n_{i..} = \sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk} = \sum_{j=1}^{J} n_{ij.} = \sum_{k=1}^{K} n_{i.k}$$

$$n_{.j.} = \sum_{i=1}^{I}\sum_{k=1}^{K} n_{ijk} \ , \ n_{..k} = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ijk}$$

$$n_{...} = \sum_{ijk} n_{ijk} \qquad (4.1)$$

Örneğin $n_{ij.}$ iki boyutlu I*J marjinal tablosunu tanımlar. p_{ijk} , m_{ijk} tahminleri için benzer notasyonlar kullanılabilir.

4.1. Bağımsızlık Modelleri

İki boyutlu tablolarda sadece satır ve sütun bağımsızlık modeli ile ilgilenilirken, üç boyutlu tablolarda satır, sütun ve tabaka boyutlarının kombinasyonları ile oluşan 8 bağımsızlık modeli ile ilgilenilir. Tüm bağımsızlık modelleri için maksimum olabilirlik tahminleri elde edilebilir. Bu modellerden ilerleyen bölümlerde de bahsedileceği için ayrıntı olarak ele alınacaktır.

(0) Tam bağımsızlık modeli: Satır, sütun ve tabakanın bağımsız olduğu modeldir.

$$M^{(0)} : p_{ijk} = p_{i..}p_{.j.}p_{..k} \tag{4.2}$$

Bu model için p_{ijk} 'nın maksimum olabilirlik tahmini:

$$\hat{p}^{(0)}_{ijk} = \hat{p}_{i..}\hat{p}_{.j.}\hat{p}_{..k}$$
$$= (n_{i..}/n_{...})(n_{.j.}/n_{...})(n_{..k}/n_{...}) \text{ 'dir.} \tag{4.3}$$

$m_{ijk} = n_{...}p_{ijk}$ olduğu için m_{ij} 'nın maksimum olabilirlik tahmini:

$$\hat{m}^{(0)}_{ijk} = n_{...}\hat{p}^{(0)}_{ijk}$$
$$= n_{i..}n_{.j.}n_{..k}/n^2_{...} \quad \text{olur.} \tag{4.4}$$

$M^{(0)}$ modelinin uyumluluğunun testi için test istatistikleri aşağıda verilmiştir.

Çizelge 4.2. Tam bağımsızlık modeli için Pearson χ^2 ve Olabilirlik oran G^2 test istatistikleri

Pearson χ^2	Olabilirlik oran G^2
$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{(n_{ijk} - \hat{m}_{ijk}{}^{0})^2}{\hat{m}_{ijk}{}^{0}}$	$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk}\log(n_{ijk}/\hat{m}_{ijk}^{(0)})$

Eğer hesaplanan test istatistiği $\chi^2_{(1-\alpha,IJK-I-J-K+2)}$'den büyükse, α seviyesinde $M^{(0)}$ modeli reddedilir. Herhangi bir kısıtlama olmadan m_{ijk}' nın maksimum olabilirlik tahmini $\hat{m}_{ijk} = n_{ijk}$ dir. Bu yüzden n_{ijk}, G^2 istatistiği için formülde kullanılır.

<u>Bir faktörün diğer iki faktörden bağımsız olması:</u>

(1) Satırların, sütun ve tabakadan bağımsız olduğu model: $M^{(1)} : p_{ijk} = p_{i..}p_{.jk}$

(2) Sütunların, satır ve tabakalardan bağımsız olduğu model: $M^{(2)} : p_{ijk} = p_{.j.}p_{i.k}$

(3) Tabakaların sütun ve satırdan bağımsız olduğu model: $M^{(3)} : p_{ijk} = p_{..k}p_{ij.}$

Bu modellerin üçü de aslında tam bağımsızlık $M^{(0)}$ modelini kapsar. Eğer $M^{(0)}$ doğruysa, bu üç modelin hepsi doğrudur.

$M^{(1)}$ modeli için p_{ijk}'nın maksimum olabilirlik tahmini:

$$\hat{p}_{ijk}^{(1)} = \hat{p}_{i..}\hat{p}_{.jk} = (n_{i..}/n_{...})(n_{.jk}/n_{...}) \qquad (4.5)$$

m_{ijk}' nın maksimum olabilirlik tahmini:

$$\hat{m}_{ijk}^{(1)} = n_{...}\hat{p}_{ijk}^{(1)}$$

$$= n_{i..} n_{.jk} / n_{...} \tag{4.6}$$

Örneğin $M^{(1)}$ modeli için test istatistikleri aşağıdaki gibidir. Serbestlik derecesi (sd) $(I-1)(JK-1)$ olan χ^2 ile karşılaştırılır [1].

Çizelge 4.3. Satırların, sütun ve tabakadan bağımsız olduğu model için Pearson χ^2 ve Olabilirlik oran G^2 test istatistikleri

Pearson χ^2	Olabilirlik oran G^2
$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{(n_{ijk} - \hat{m}_{ijk}^{\ 1})^2}{\hat{m}_{ijk}^{\ 1}}$	$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk}\log(n_{ijk} / \hat{m}_{ijk}^{(1)})$

Koşullu bağımsızlık modelleri:

Bazı faktörlerin bazı seviyeleri verildiğinde, diğer iki faktör bağımsız olabilir. Üçüncü bir faktör verildiğinde diğer iki faktör bağımsız ise, ortaya üç model çıkar:

(4) Tabakalar verildiğinde satır ve sütunların bağımsız olduğu model:

$$M^{(4)} : p_{ijk} = p_{i.k} p_{.jk} / p_{..k} \tag{4.7}$$

(5) Sütunlar verildiğinde satır ve tabakanın bağımsız olduğu model:

$$M^{(5)} : p_{ijk} = p_{ij.} p_{.jk} / p_{.j.} \tag{4.8}$$

(6) Satırlar verildiğinde sütun ve tabakaların bağımsız olduğu model:

$$M^{(6)} : p_{ijk} = p_{ij.} p_{i.k} / p_{i..} \tag{4.9}$$

$M^{(4)}$ modeli için hesaplamaları ele alalım, diğer modeller de aynı yolla elde edilebilirler.

k. tabaka verildiğinde i. satır ve j. sütunun olasılığı Eş. 4.10'daki gibidir.

$$P(satır = i, sütun = j / tabaka = k) = \frac{P(satır = i, sütun = j, tabaka = k)}{P(tabaka = k)}$$

$$= p_{ijk} / p_{..k} \qquad (4.10)$$

Her bir tabaka için satır ve sütunun koşullu bağımsızlığı tüm i, j, k'lar için:

$$P(satır = i, sütun = j / tabaka = k) = P(satır = i / tabaka = k)P(sütun = j / tabaka = k)$$

$= (p_{i.k} / p_{..k})(p_{.jk} / p_{..k})$ 'dir. (4.11)

Eş. 4.10 ile Eş. 4.11'i eşitlersek $p_{ijk} = p_{i.k} p_{.jk} / p_{..k}$ olur [1] .

p_{ijk}'nın maksimum olabilirlik tahmini Eş. 4.12'deki gibidir.

$$\hat{p}_{ijk}^{(4)} = \hat{p}_{i.k} \hat{p}_{.jk} / \hat{p}_{..k}$$

$$= (n_{i.k} / n_{...})(n_{.jk} / n_{...}) / (n_{..k} / n_{...})$$

$$= n_{i.k} n_{.jk} / n_{..k} n_{...} \qquad (4.12)$$

$m_{ijk} = n_{...} p_{ijk}$ için maksimum olabilirlik tahmini Eş. 4.13'deki gibidir.

$$\hat{m}_{ijk}^{(4)} = n_{...} \hat{p}_{i.k} \hat{p}_{.jk} / \hat{p}_{..k}$$

$$= n_{i.k} n_{.jk} / n_{..k} \qquad (4.13)$$

Test istatistikleri (I-1)(J-1) serbestlik dereceli χ^2 ile karşılaştırılır.

Çizelge 4.4. Tabakalar verildiğinde satır ve sütunların bağımsız olduğu model için Pearson χ^2 ve Olabilirlik oran G^2 test istatistikleri

Pearson χ^2	Olabilirlik oran G^2
$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \frac{(n_{ijk} - \hat{m}_{ijk}{}^4)^2}{\hat{m}_{ijk}{}^4}$	$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk} \log(n_{ijk} / \hat{m}_{ijk}^{(4)})$

Koşullu odds oranı:

Tabakalar verildiğinde satır ve sütunların bağımsız olduğu modeli ele alalım. Tabaka değişkeninin kategorileri için ayrı ayrı odds oranları hesaplanabilir.
Tabaka değişkeninin k. kategorisi için koşullu odds oranı aşağıdaki gibidir.

$$\theta_{IJ(k)} = \frac{n_{11k} n_{22k}}{n_{12k} n_{21k}} \tag{4.14}$$

Eğer koşullu odds değeri 1'e eşitse tabakaların, satır ve sütundan bağımsız olduğunu söyleyebiliriz.

(7) Üç yönlü tablolarda son model:

Üç yönlü tabloların son modeli Bartlett (1935) tarafından ortaya atılmıştır. Odds oranları ile ifade edilir ve bu modelde tüm koşullu oddslar eşittir. Üç yönlü tablolarda odds oranları, bir faktörü sabitlenerek diğer iki faktör için odds oranını hesaplar.

Örneğin tabakalar sabitlendiğinde odds oranı: $p_{11k} p_{ijk} / p_{1jk} p_{i1k}$ 'dır ve her bir tabaka için bu odds oranları aynıdır [1]. $i = 2,..,I$, $j = 2,..,J$ $k = 2,..,K$ için son model Eş. 4.15'deki gibidir.

Son model: $M^{(7)} : \frac{p_{111} p_{ij1}}{p_{i11} p_{1j1}} = \frac{p_{11k} p_{ijk}}{p_{i1k} p_{1jk}}$ (4.15)

$\hat{p}_{ijk}^{(7)}$ ve $\hat{m}_{ijk}^{(7)}$ için basit bir formül yoktur. Maksimum olabilirlik tahminleri bilgisayar programlarındaki iteratif yöntemlerle elde edilir. Çizelge 4.5'de verilen test istatistikleri de (I-1)(J-1)(K-1) serbestlik dereceli χ^2 ile karşılaştırılır.

Çizelge 4.5. Tüm oddsların eşit kabul edildiği model için Pearson χ^2 ve Olabilirlik oran G^2 test istatistikleri

Pearson χ^2	Olabilirlik oran G^2
$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{(n_{ijk} - \hat{m}_{ijk}{}^{7})^2}{\hat{m}_{ijk}{}^{7}}$	$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk}\log(n_{ijk}/\hat{m}_{ijk}^{(7)})$

4.2. Üç Boyutlu Tablolarda Hiyerarşik Log-Lineer Modeller

Üç boyutlu tablolarda sürekli olan veriler için temel model, tüm etkileşimleri içeren üç yönlü varyans modelidir. "1" satırları,"2" sütunları ve "3" tabakayı temsil etmektedir.

$$y_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)} + e_{ijk} \quad (4.16)$$

Log-lineer modellerde de $\log(m_{ijk})$ için aynı model kullanılır. Üç boyutlu tablolar için doymuş model Eş. 4.17'de verilmiştir.

$$\log(m_{ijk}) = \lambda + \lambda_{1(i)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{12(ij)} + \lambda_{13(ik)} + \lambda_{23(jk)} + \lambda_{123(ijk)} \quad (4.17)$$

İlk önce bu modelin sekiz adet azaltılmış versiyonu ile ilgileneceğiz. Bu azaltılmış modellerin her biri, üç boyutlu tablolar için sekiz bağımsız odds oranı modellerinden birine karşılık gelir. Burada, bir odds oranı modeli sadece karşılaştırılabilen log-linear modeli geçerli olduğunda geçerlidir [1]. Hiyerarşik log-lineer modeller Çizelge 4.6'da verilmiştir.

Çizelge 4.6. Hiyerarşik log-lineer modeller [1]

Log-lineer model	Model numarası
$\log(m_{ijk}) = \lambda$	(-1)
$\log(m_{ijk}) = \lambda + \lambda_{1(i)} + \lambda_{2(j)} + \lambda_{3(k)}$	(0)
$\log(m_{ijk}) = \lambda + \lambda_{1(i)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{23(jk)}$	(1)
$\log(m_{ijk}) = \lambda + \lambda_{1(i)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{13(ik)}$	(2)
$\log(m_{ijk}) = \lambda + \lambda_{1(i)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{12(ij)}$	(3)
$\log(m_{ijk}) = \lambda + \lambda_{1(i)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{13(ik)} + \lambda_{23(jk)}$	(4)
$\log(m_{ijk}) = \lambda + \lambda_{1(i)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{12(ij)} + \lambda_{23(jk)}$	(5)
$\log(m_{ijk}) = \lambda + \lambda_{1(i)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{12(ij)} + \lambda_{13(ik)}$	(6)
$\log(m_{ijk}) = \lambda + \lambda_{1(i)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{12(ij)} + \lambda_{13(ik)} + \lambda_{23(jk)}$	(7)
$\log(m_{ijk}) = \lambda + \lambda_{1(i)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{12(ij)} + \lambda_{13(ik)} + \lambda_{23(jk)} + \lambda_{123(ijk)}$	(8)

$i = -1,...,8$ için log-lineer model (i) sadece ve sadece $M^{(i)}$ geçerli olduğunda geçerlidir. Örneğin model (7) sadece $\lambda_{123(ijk)}$ terimi doymuş modelden atıldığında geçerlidir. Kısa gösterimi ile log-lineer modeller Çizelge 4.7'de verilmiştir.

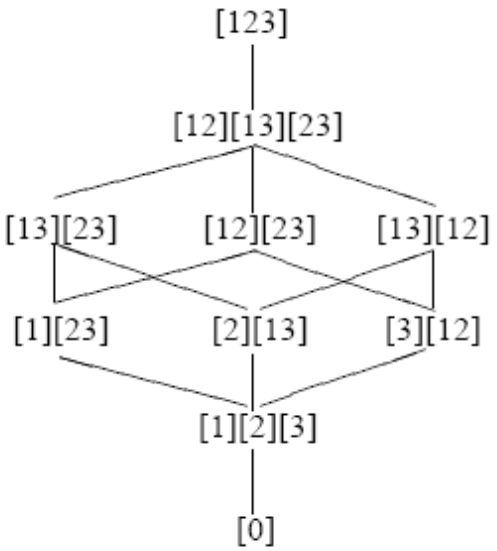
Kısaltmalar modellerin koşullu bağımsızlıklarının yorumları için de kullanılabilirler. Çizelge 4.1'de satırlar 1, sütunlar 2 ve tabaka 3 ile gösterilmek üzere, [1][2][3] modelinde satır, sütun ve tabaka birbirlerinden bağımsızdır. [1][23] modelinde satırlar, sütun ve tabakadan farklı dilimdedir yani "satırlar, sütun ve tabakadan bağımsızdır" diye yorumlanır. [13][23] modelinde ise satırlar ve sütunlar farklı dilimlerde fakat tabaka her iki dilimdedir. Yani tabaka verildiğinde satır ve sütun bağımsızdır. [12][13][23] için değişkenlerin ayrımı söz konusu değildir ve karşılığı

olan model için bağımsızlık terimlerinde hiçbir yorumlama yoktur. Parametreler maksimum olabilirlik tahmin yöntemi ile tahmin edilebilir [1].

Çizelge 4.7. Kısa gösterimi ile log-lineer modeller

Model	Kısa gösterim	Serbestlik derecesi
(-1) $\log(m_{ijk}) = \lambda$	[0]	-
(0) $\log(m_{ijk}) = \lambda_{1(i)} + \lambda_{2(j)} + \lambda_{3(k)}$	[1][2][3]	IJK-I-J-K+2
(1) $\log(m_{ijk}) = \lambda_{1(i)} + \lambda_{23(jk)}$	[1][23]	(I-1)(JK-1)
(2) $\log(m_{ijk}) = \lambda_{2(j)} + \lambda_{13(ik)}$	[2][13]	(J-1)(IK-1)
(3) $\log(m_{ijk}) = \lambda_{3(k)} + \lambda_{12(ij)}$	[3][12]	(K-1)(IJ-1)
(4) $\log(m_{ijk}) = \lambda_{13(ik)} + \lambda_{23(jk)}$	[13][23]	K(I-1)(J-1)
(5) $\log(m_{ijk}) = \lambda_{12(ij)} + \lambda_{23(jk)}$	[12][23]	J(I-1)(K-1)
(6) $\log(m_{ijk}) = \lambda_{12(ij)} + \lambda_{13(ik)}$	[12][13]	I(J-1)(K-1)
(7) $\log(m_{ijk}) = \lambda_{12(ij)} + \lambda_{13(ik)} + \lambda_{23(jk)}$	[12][13][23]	(I-1)(J-1)(K-1)
(8) $\log(m_{ijk}) = \lambda_{123(ijk)}$	[123]	0

Hiyerarşik modeller ayrıca iç içe geçmiş şekilde Şekil 4.1'de gösterilmiştir.



Şekil 4.1. İç içe geçmiş hiyerarşik log-lineer modeller [23]

4.2.1. Log-lineer modellerin testi

Regresyon ve varyans analizinde daha küçük olan modelin veriye uyup uymadığını görmek için model daha büyük modele karşı test edilir. Bu teknik aynı zamanda log-lineer modellerde de kullanılabilir [1].

Herhangi bir modeli r ile gösterirsek $M^{(r)}$ modelinin testi , $\hat{m}_{ijk}^{(r)}$ ve n_{ijk}'ye dayanır. $n_{ijk} = \hat{m}_{ijk}^{(r)}$ olduğu için tahminlerde n_{ijk}'ler kullanılır. $\hat{m}_{ijk}$'ler m_{ijk}'nın maksimum olabilirlik tahmin edicisidir. Bu tahmin edici m_{ijk}'ya hiçbir kısıtlama getirmeyen bir model yani doymuş model (model (8)) yardımıyla elde edilir [1].

$$\log(m_{ijk}) = \lambda + \lambda_{1(i)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{12(ij)} + \lambda_{13(ik)} + \lambda_{23(jk)} + \lambda_{123(ijk)} \qquad (4.18)$$

Yukarıdaki eşitlikte belirtilen doymuş model, tabloda her bir hücre için en az bir parametreye sahiptir. Böylece model veriye her zaman mükemmel uyar.

Genel olarak herhangi bir modeli kendisinden daha büyük olan modele karşı test edebiliriz. Örneğin, model (1) , model (4)'e karşı test edilebilir.

$M^{(1)}$: model (1): $\log(m_{ijk}) = \lambda + \lambda_{1(i)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{23(jk)}$

$M^{(4)}$: model (4) : $\log(m_{ijk}) = \lambda + \lambda_{1(i)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{13(ik)} + \lambda_{23(jk)}$

Çünkü model (4) model (1)'deki tüm parametrelerin yanında $\lambda_{13(ik)}$'yi de içermektedir. Diğer bir ifadeyle, model (1), model (4)'ün özel bir durumudur. Test basit olarak, $\lambda_{13(ik)}$'ün model (4) te gerekli olup olmadığının testidir yani $M^{(1)}$'in $M^{(4)}$'e karşı eşitliğinin testidir. [1][23]'ü [13][23]'e karşı test etmek için $\hat{m}_{ijk}^{(1)}$, $\hat{m}_{ijk}^{(4)}$tahminlerini, Pearson χ^2 ve olabilirlik oran G^2 testlerini kullanırız [1]. Hipotezler aşağıdaki gibidir:

$H_0 : \lambda_{13(ik)} = 0$

$H_1 : \lambda_{13(ik)} \neq 0$

Çizelge 4.8. Model (1)'in Model (4)'e eşitliği için Pearson χ^2 ve Olabilirlik oran G^2 test istatistikleri

Pearson x^2	Olabilirlik oran G^2
$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{(\hat{m}_{ijk}{}^4 - \hat{m}_{ijk}{}^1)^2}{\hat{m}_{ijk}{}^1}$	$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\hat{m}_{ijk}^{(4)}\log(\hat{m}_{ijk}^{(4)}/\hat{m}_{ijk}^{(1)})$

Bu bir satır-tabaka etkileşimi (λ_{13}) olmadığının testi olduğu için, serbestlik derecesi, varyans analizinde olduğu gibi, satır-tabaka etkileşimin serbestlik derecesi $((I-1)(K-1))$'ye eşittir.

Benzer olarak, model (1) : [1][23]'i, model (5): [12][23]'e karşı test edebiliriz çünkü model (5) ,model (1) i içerir. Fakat [1][23] modelini, [12][13] modeline karşı test

edilemez çünkü [1][23] modeli $\lambda_{23(jk)}$ terimini içerirken [12][13] modeli içermez. Böylece [12][13] modeli, [1][23] modelinden tam anlamıyla büyük bir model değildir. Bu durumda bu iki modelin karşılaştırılamayacağını söyleriz [1].

Üç boyutlu tablolarda lambda (λ) parametrelerinin serbestlik derecesi:

Bilindiği gibi bu testleri uygulayabilmek için, her bir modelle ilişkili olan serbestlik derecelerini belirlemeliyiz. Klasik varyans analizinde olduğu gibi bir modelin serbestlik derecesi, modeldeki her bir terimin serbestlik derecelerinin toplamına eşittir. Lambda (λ) parametrelerinin ve modellerin serbestlik derecesi varyans analizinde olduğu gibidir [1].

Çizelge 4.9. Üç boyutlu tablolarda lambda parametrelerinin serbestlik derecesi

Parametre	Serbestlik derecesi
λ	1
λ_1	(I-1)
λ_2	(J-1)
λ_3	(K-1)
λ_{12}	(I-1)(J-1)
λ_{13}	(I-1)(K-1)
λ_{23}	(J-1)(K-1)
λ_{123}	(I-1)(J-1)(K-1)

Kısa gösterimleriyle [1][23] modelini [13][23] modeline karşı test etmede serbestlik derecesi, [13][23] modelinin serbestlik derecesinden [1][23] modelinin serbestlik derecesinin çıkarımı ile elde edilir.

[13][23] modeli için serbestlik derecesi, 1 +(I-1)+(J-1)+(K-1)+(I-1)(K-1)+(J-1)(K-1) ve
[1][23] modeli için serbestlik derecesi, 1 +(I-1)+(J-1)+(K-1)+(J-1)(K-1) 'dir.

Test için serbestlik derecesi, serbestlik dereceleri arasındaki farktır: (I-1)(K-1).
Daha önce de belirtildiği gibi bu serbestlik derecesi, [13][23] modelinde olup [1][23] modelinde olmayan $\lambda_{13(ik)}$ nin serbestlik derecesidir.

4.3. Lojistik Regresyonda Model Seçme Yöntemleri

Lojistik regresyonda değişken seçiminde farklı yöntemler kullanılmaktadır. Bunlardan biri basit lojistik regresyon ile katsayıların tek tek testi, diğeri adımsal yöntemlerdir.

İlk yöntemde değişkenlerin önce teker teker basit lojistik regresyon analizleri ile önemlilikleri belirlenir ve önemli bulunan değişkenler çok değişkenli modele dahil edilir. Burada herhangi bir olumsallık tablosundaki hücrelerden birinin ya da daha fazlasının gözlenen frekansının sıfır olması durumuna dikkat edilmelidir. Değişken ve model seçimi için diğer bir yaklaşım da adımsal yöntemlerdir. Hem lojistik regresyonda hem de log-lineer modellerde kullanılır.

Adımsal Yöntemler

Adımsal yöntemler isminden de anlaşılacağı üzere adım adım ilerleyerek her adımda geçerli bir model varsayar ve bu modele, her seferinde bir terim ekleyerek ya da silerek devam eder. Üç'e ayrılırlar:

- ileri seçim yöntemi: ilk küçük modele terimler eklenir.
- geriye eleme yöntemi: ilk büyük modelden terimler çıkarılır.
- karma yöntemler: ilk modelden ya terimler çıkarılır ya da eklenir.

Bu tezde sadece ileri seçim ve geriye eleme yöntemleri uygulanacaktır.

Temel olarak herhangi bir 'x-faktör' terimi ele alındığında,
İleri seçim kuralı;

a) En anlamlı p değerine sahip olan, modelde henüz var olmayan ‘x-faktör’ terimi modele eklenir.

b) (α) anlamlılık seviyesinde modelin uyum iyiliği için kayda değer bir artış sağlanamadığı noktada terim ekleme işlemi durdurulur.

Geriye eleme kuralı;

a) Modele alınmamış x-faktör terimleri arasındaki en az anlamlı p değerine sahip olan s-faktör terimi silinir.

b) Modeldeki tüm terimler önceden belirlenmiş anlamlılığın(α) en düşük seviyesini koruyana kadar terim silmeye devam edilir.

Test istatistiğinin anlamlılığı P değerleri ile ölçülür. Bir test istatistiği, önceden belirlenmiş anlamlılığın en düşük seviyesine (α) ulaşmayı başaramazsa, yani $P < \alpha$ ise önem seviyesini korur. α seviyesi genelde 0.10, 0.05 ve 0.01 kabul edilir.

4.4. Uygun Modelin Seçiminde Kullanılan Akaike Bilgi Kriteri (AIC)

En uygun modelin seçiminde, birden fazla “en iyi model” olması durumunda kullanılan, Akaike (1973) tarafından ortaya atılan bir kriterdir. Genel olarak Akaike bilgi kriteri AIC ile gösterilip ,“L” log-olabilirlik değeri ve “d” modeldeki parametre sayısı olmak üzere, (-2 L+2d) şeklinde hesaplanır.

Log-lineer modeller için AIC değeri aşağıdaki gibi hesaplanır. AIC değerini maksimize etmek, log-lineer modeller için aşağıdaki eşitliği minimize etmektir.

$$AIC = A_M = G^2(M) - [q-2r] \qquad (4.19)$$

Bu eşitlikte;

$G^2(M)$, M modelini doymuş modele karşı test eden olabilirlik oran test istatistiğini,

r, M modeli için serbestlik derecesini,

q, doymuş model için serbestlik derecesini belirtmektedir.

Uygulamada AIC değerinden türetilen ve hesaplaması daha basit olan A_M-q kriteri tercih edilmektedir.

$$A_M\text{-}q = G^2(M) - 2\,[q-r] \qquad (4.20)$$
$$= G^2(M) - 2\,[sd]$$

Bu eşitlikte q değeri M modeline bağlı olmadığı için, A_M-q değerini minimize etmek A_M değerini minimize etmeye eştir [1]. Bu eşitlik, lojistik regresyonda model seçim aşamasında da kullanılmaktadır.

4.5. Üç Boyutlu Tablolarda Lojit Model ve Log-Lineer Modellerin Denkliği

Üç boyutlu tablolarda log-lineer modeller verildiğinde elde edilen lojit modeller aşağıdaki gibidir. Çizelge 4.1 dikkate alınırsa "1" satırları, "2" sütunları ve "3" tabakayı belirtmektedir.

Örnek 4.1: [1][2][3] ve [1][23] modellerinin ana etki β_0 ' a eşitliğinin gösterilmesi

[1][2][3] log-lineer modeli için:

$$\log\left(\frac{p_{1jk}}{p_{2jk}}\right) = \log\left(\frac{m_{1jk}}{m_{2jk}}\right)$$
$$= \log(m_{1jk}) - \log(m_{2jk})$$
$$= [\lambda_{1(1)} + \lambda_{2(j)} + \lambda_{3(k)}] - [\lambda_{1(2)} + \lambda_{2(j)} + \lambda_{3(k)}]$$
$$= \lambda_{1(1)} - \lambda_{1(2)} = \beta_0 \;:\; \text{ana etkidir.} \qquad (4.21)$$

[1] [23] modeli için:

$$\log\left(\frac{p_{1jk}}{p_{2jk}}\right) = \log\left(\frac{m_{1jk}}{m_{2jk}}\right)$$
$$= \log(m_{1jk}) - \log(m_{2jk})$$

$$= [\lambda_{1(1)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{23(jk)}] - [\lambda_{1(2)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{23(jk)}]$$

$$= \lambda_{1(1)} - \lambda_{1(2)} = \beta_0 \quad : \text{ana etkidir.} \tag{4.22}$$

Görüldüğü üzere [1][2][3] ve [1][23] log-lineer modelleri aynı ve tek bir lojit modele eşittir.

Örnek 4.2: [12][3] ve [12][23] modellerinin $\beta_0 + \beta_j$'ye eşitliğinin gösterilmesi

[12][3] modeli için:

$$\log(\frac{p_{1jk}}{p_{2jk}}) = \log(\frac{m_{1jk}}{m_{2jk}})$$

$$= \log(m_{1jk}) - \log(m_{2jk})$$

$$= [\lambda_{1(1)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{12(1j)}] - [\lambda_{1(2)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{12(2j)}]$$

$$= [\lambda_{1(1)} - \lambda_{1(2)}] + [\lambda_{12(1j)} - \lambda_{12(2j)}]$$

$$= \beta_0 + \beta_j \tag{4.23}$$

Burada β_j ikinci faktörün etkisidir.

[12][23] modeli için:

$$\log(\frac{p_{1jk}}{p_{2jk}}) = \log(\frac{m_{1jk}}{m_{2jk}})$$

$$= \log(m_{1jk}) - \log(m_{2jk})$$

$$= [\lambda_{1(1)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{12(1j)} + \lambda_{23(jk)}] - [\lambda_{1(2)} + \lambda_{2(j)} + \lambda_{3(k)} + \lambda_{12(2j)} + \lambda_{23(jk)}]$$

$$= [\lambda_{1(1)} - \lambda_{1(2)}] + [\lambda_{12(1j)} - \lambda_{12(2j)}]$$

$$= \beta_0 + \beta_j \tag{4.24}$$

Görüldüğü üzere [12][3] ve [12][23] modelleri eşittir. Diğer modeller için de aynı uygulama yapılabilir. Özet tablo aşağıdaki gibidir.

Çizelge 4.10. Log-lineer –Lojit ilişkisi [3]

Log-lineer model	Lojit
[1][2][3] ; [1] [23]	β_0
[12][3]; [12][23]	$\beta_0 + \beta_j$
[13][2] ; [13][23]	$\beta_0 + \beta_k$
[12][13]; [12][13]	$\beta_0 + \beta_j + \beta_k$
[123]	$\beta_0 + \beta_j + \beta_k + \beta_{jk}$

Uygulamada log-lineer modeller ile lojistik regresyonun farklı amaçları, yorumları, avantajları ve dezavantajları vardır. Burada dikkat edilmesi gereken, iki log-lineer modelin tek bir lojit modele eşit olmasıdır.

5. LOJİSTİK REGRESYON İLE LOG-LİNEER MODELLERİN ÖZELLİKLERİNİN KARŞILAŞTIRILMASI

Bu bölümde hem lojistik regresyonun hem de log-lineer modellerin kullanım alanları, birbirlerine göre üstünlükleri ele alınacaktır. Ancak daha önce istatistikte önemli yeri olan "sıfır hücreleri" ile ilgili ön bilgi verilecektir.

5.1. Yapısal ve Örnekleme Sıfırları

Olumsallık tablolarında bazı durumlarda $n_i = 0$ olan sıfırlı hücreler bulunmaktadır. Bunlar yapısal(structural) sıfırlar ve örnekleme (sampling) sıfırları olmak üzere iki çeşittir.

Çoğu durumda gözlenen frekans $n_i = 0$ olmasına rağmen, beklenen frekans $m_i > 0$ dır ve örneklem sayısı arttıkça $n_i > 0$ olur. Bu tip hücreler "örnekleme sıfırlı hücreler" olarak adlandırılır. Ancak eğer elde etmek istediğimiz gözlemler mümkün değilse bu tip hücreler "yapısal sıfırlı hücreler" olarak adlandırılır. Bu gibi hücreler için beklenen frekans $m_i = 0$, tahmin edilen frekans $\hat{m}_i = 0$ ve gözlenen frekans $n_i = 0$ dır. Yapısal sıfırlı olumsallık tabloları "tamamlanmamış tablolar" olarak adlandırılırlar [11]. Örneğin ameliyat yöntemi ve cinsiyet değişkenlerinin çapraz sınıflamasında bayanların prostat ameliyatı ya da erkeklerin jinekolojik ameliyat geçirmesi mümkün olmadığı için bu hücrelere düşen sıfırlar yapısal sıfırlardır [21].

Örnekleme sıfırları ise verinin bir parçasıdır. Sıfır sayıları, poisson ya da çokterimli rassal değişkenlerde ortaya çıkabilecek bir sonuçtur [11]. Özellikle ele alınan değişkenlerden bazıları çapraz sınıflandığında, bazı kategorilerin gözlenme olasılığı azalır. Sıfırlı hücrelerin varlığı, bu tür durumların yığında bulunmaması anlamına gelmez, sadece örneğe düşmediğini belirtir [21]. Yapısal sıfır hücreleri ise, verinin bir parçası değildir ve örnekleme sıfırlarından daha nadir ortaya çıkarlar [11].

Hosmer ve Lemeshow sıfırlı hücrelerden kaynaklanan sorunları ortadan kaldırmak için “ad hoc çözümleri” olarak adlandırdıkları üç yöntem önermişlerdir [7].

1) Sıfırlı hücre ile sıfırlı olmayan hücreler beraber değerlendirilerek bir nitel değişkenin kategorileri birleştirilir. Böylece iki ya da daha çok hücre birleştirilerek değişkenin kategori sayısı azaltılır ve sıfırlı hücreden kurtarılır.
2) Basitçe sıfırlı hücrenin bulunduğu kategori elenerek sıfırlı hücreden tablo kurtarılır.
3) Eğer sıfırlı hücreye sahip olan değişken sıralı ölçüm düzeyine sahipse, değişkene sanki eşit aralıklı ölçülmüş gibi davranılır.

Burada Ad hoc "amaca özel, niyete mahsus" anlamına gelen Latince ibaredir. Genelde bir soruna yönelik, geçici bir çözümü anlatmak için kullanılır. Bazen de bir yetersizliği ya da üstünkörü üretilen çözümleri vurgulamak için kullanılır [15].

5.2. Log-Lineer Modellerin Uygulamaları

Daha öncede belirtildiği gibi log-lineer modeller ile

1) Bağımlı - bağımsız değişken ayrımı yapılarak ya da yapılmayarak, değişken setleri arasındaki etkileşimin ölçüsünü belirlenir.
2) Her bir bağımlı-bağımsız değişken birlikteliğin derecesi, geriye kalan bağımlı ve bağımsız değişkenlerin etkisi sabitlenerek ölçülür [7].

Çoğu log-lineer uygulama, kısmi etkileşimleri ölçmek için ilk uygulamaya yoğunlaşmıştır. Değişkenler arasındaki kısmi etkileşimler, lambda (λ) ağırlıkları ve onların Wald Z testleri ile belirlenir. Wald z testi, λ parametresinin maksimum olabilirlik tahmininin kendi standart hatasına bölünmesiyle elde edilir.

İkinci uygulama ise aynen klasik regresyon analizi ya da ANOVA çalışmalarında olduğu gibidir, ancak burada araştırmacılar her bir değişkeni bağımlı ve bağımsız değişken olarak belirtmek zorundadırlar.

Hiyerarşik log-lineer modeller altında beklenen hücre frekanslarının maksimum olabilirlik tahminleri hangi uygulamanın seçileceğine bağlı olarak değişir. Kısmi etkileşimleri vurgulayan ilk uygulama için beklenen hücre frekansları, her bir değişkenin satır ve sütun etkilerinin ve bu değişkenler arasındaki etkileşimlerin bir fonksiyonudur [27-28].

Bağımlı, bağımsız değişkenlerin birlikte ele alındığı çalışmalarda ise, araştırmacılar sadece, bağımsız değişkenler arasındaki tüm olası etkileşimleri otomatik olarak oluşturan, olası hipotez modellerinin sınırlı bir alt setine yoğunlaşır. Özellikle, bağımsız değişkenler arasındaki en yüksek sıralı etkileşim terimi ele alınır ve bağımlı değişkenlerin her bir çifti arasındaki birlikteliğin testinin yanında, her bir bağımlı değişken ile her bir bağımsız değişken arasındaki birliktelik test edilir.

5.3. Lojistik Regresyonun Uygulamaları

1) Kategorik veri analizinde lojistik regresyon (lojit model) , log-lineer modellerin bir alt seti ya da özel bir durumudur ve iki değer alan bir bağımlı değişkenin log-oddsunu tahmin eder. Bununla birlikte, sınıflama ve sıralama ölçme düzeyine sahip değişken setinin analizi için, araştırmacılar log-lineer modeller ya da lojit modelden birini seçebilirler [11].

2) Lojistik regresyon, bir bağımsız değişkenin bağımlı değişken üzerindeki büyüklüğünü ve işaretini farklı yollarla tahmin edebilir [30]. Eğer modeldeki tüm değişkenler kategorikse, ağırlıklandırılmış en küçük kareler ya da maksimum olabilirlik yöntemleri kullanılabilir. Bununla birlikte, eğer tüm değişkenler eşit aralıklı ya da hem kategorik hem eşit aralıklı ölçüm düzeyli karma bir setse sadece maksimum olabilirlik yöntemleri kullanılmalıdır [7]. Uygulama kısmında maksimum olabilirlik tahminleri kullanılmıştır.

3) Lojistik regresyon için, maksimum olabilirlik tahminleri tutarlı ve asimptotik olarak etkilidir [31]. Ancak, olumsallık tablosunda çoklu sıfır hücreye sahip olan seyrek veri için kısmi eğim katsayısı (bir bağımsız değişkenin bağımlı değişken

üzerindeki etkisinin tahmini) sonsuz çıkabilir. Maksimum olabilirlik tahmin edicisinin optimal özellikleri sadece büyük çaplı örnekler için korunur. Küçük ya da orta çaplı örneklerde maksimum olabilirlik tahmin edicisinin özellikleri şüphelidir [7].

5.4. Lojistik Regresyonun Göreli Avantajları

1) Metodolojik erişebilirlik: lojistik regresyonun klasik en küçük kareler regresyonuna benzerliği

Lojistik regresyon, klasik regresyon analizinde olduğu gibi, her bir bağımsız değişkenin bağımlı değişken üzerindeki etkisinin yönünü, iki ya da daha fazla değişken arasındaki birlikteliğin büyüklüğünü, standart hataları tahmin eder. Bunun için Wald z ve olabilirlik oran test istatistiklerini kullanır [7].

Log-lineer modellerde ise iki ya da daha fazla değişken arasındaki birlikteliğin büyüklüğünün(λ) tahminleri kovaryans matrisinden elde edilir. Delta(Δ) metoduna dayalı (Bkz. Ek 4) standart hatalar sadece bu durumda mümkündür. Eğer test edilen modelin beklenen hücre frekansları matematiksel formüllerle elde edilemiyor, bilgisayar algoritması ile iteratif olarak tahmin edilebiliyorsa, o zaman log-lineer modeller λ ağırlıklarını, bunların standart hatalarının tahmini olmadan elde eder. Böylece λ ağırlıkları standart hatalarına bölünemez ve H_0:λ=0 hipotezi test edilemez [29].

Eğer modelde üçten fazla değişken varsa, çeşitli hipotez modelleri sadece bilgisayar algoritmaları ile tahmin edilebilir [11]. Bu durumda log-lineer modellerle kıyaslanırsa lojistik regresyon araştırmacılara her bir bağımsız değişkenin bağımlı değişken üzerindeki nicel etkisini tahmin etmek için fırsat sağlar [7].

2) Bağımlı değişken için koşullu oddsları tahmin etme

Lojistik regresyon, bağımsız değişkenlerin koşullu bir seti verildiğinde, bir olayın olması /olmaması oranının doğal logaritmasını (log-odds) tahmin etmek için nokta tahminleri ve güven aralıklarını elde eder. Bu gibi bilgiler ancak seçilen bir log-lineer model, eşit bir lojit modele dönüştürülerek elde edilebilir [32].

3) Logistik regresyonun karma veri setlerini analiz etmedeki kullanışlılığı

Lojistik regresyon sınıflama, sıralama ya da eşit-aralıklı ölçme düzeyine sahip değişkenlerin karma bir setini analiz etmede log-lineer modellerden daha kullanışlı bir araçtır [33]. Lojistik regresyonda; sınıflama düzeyine sahip değişkenler kolayca dummy(kukla) ya da çarpım değişkeni olarak uygun hale getirilirken, aralıklı değişkenler modelde olduğu gibi kullanılırlar. Fakat log-lineer modelin genel kullanımı, sınıflama ve sıralama ölçme düzeyine sahip değişkenlerle sınırlandırılmıştır [1].

5.5. Log-Lineer Modellerin Göreli Avantajları

Log-lineer modellerin logistik regresyona göre 3 avantajı vardır.

1) Logistik regresyonda araştırmacıların ilk amacı iki değer alan bağımlı değişkenin odds oranının logaritması üzerinde, her bir bağımsız değişkenin etkisini ölçmektir. Ancak lojistik regresyon araştırmacılara bağımsızlık yapısını ya da bağımsız değişkenler arasındaki birlikteliğe eş zamanlı değer biçme olanağını sağlamaz [33].

2) Log-lineer modellerin lojistik regresyona göre, çoklu bağımlı ve bağımsız değişken setleri arasındaki ilişkiyi eş zamanlı olarak test etme üstünlüğü vardır [6,11]. Bölüm 5.2'de belirtildiği gibi her bir bağımlı değişkenin her bir bağımsız değişken ile birlikteliğini test eder. Buna karşı lojistik regresyon, çoklu bağımlı değişkenlerin istatistiksel olarak anlamsız olduğu varsayımı altında tek bir bağımlılık modeli ile sıralı testlerin serisini oluşturmalıdır [33].

3) Lojistik regresyon örnekleme sıfırları ile yapısal sıfır hücrelerini birbirinden ayırt edemez [11]. Bununla birlikte sıfırlı hücre sayıları, lojistik regresyonda parametre tahminlerinde ve standart hatalarının hesaplanmasında ciddi sayısal hatalara neden olabilir çünkü odds, odds oranları sıfırlar ile tanımlanamazlar [21-31]. Geniş standart hatalar, bu ihtimalin uyarısı olabilir.

Daha önce lojistik regresyon model seçme yöntemlerinde de belirtildiği gibi değişkenlerin teker teker basit lojistik regresyon analizi ile incelenmesinde gözlenen frekansların sıfır olması durumuna dikkat edilmelidir. Bu durum odds oranlarının biri için tek değişkenli nokta tahminini sıfır ya da sonsuz yapar (0/1 veya 1/0). Böyle bir değişken lojistik regresyona dahil edilirse anlamsız sonuçlar çıkar. Çözüm olarak Bölüm 5.1'deki "ad-hoc çözümleri" uygulanabilir [19]. Log-lineer modeller ise sıfırlı hücrelere karşı duyarlıdır ve örnekleme sıfırları ile yapısal sıfırları birbirinden ayırt eder.

Lojistik regresyonla kıyaslanırsa log-lineer modeller sıfırlı hücreleri tanır ve bu hücrelerin ortaya çıkardığı problemlere çözümler önerir. Ayrıca parametrelerin maksimum olabilirlik tahminlerini elde etmek için lojistik regresyondan daha az sayısal hesaplamalara ve istatistiksel varsayımlara ihtiyaç duyar.

Sıfırlı hücrelerin örnekleme sıfırlarından mı yoksa yapısal sıfırlardan mı kaynaklandığı ayırt edilmelidir çünkü log-lineer modeller her iki durum için farklı çözümler sunar [7]. Bir ya da daha çok örnekleme sıfırına sahip olumsallık tabloları için pozitif beklenen hücre frekansları her bir hücreye küçük bir değer eklenerek ya da örnek çapı arttırılarak tahmin edilebilir. Ancak log-lineer modeller yapısal sıfırları mantıksal mümkünsüzlük olarak değerlendirir ve bu tür durumlar için beklenen hücre frekanslarını tahmin edemez [21]. Bunun sebebi beklenen hücre frekansı m_{ij}'nin sıfır olması ve dolayısıyla $\log(m_{ij})$'nin tanımsız olmasıdır [1]. Bunun yerine, log-lineer modeller de yapısal sıfırlara sabit değer olarak davranılır ve geriye kalan yapısal olmayan sıfır hücreler için beklenen hücre frekanslarının maksimum tahminleri üretilmeye çalışılır [7].

Çizelge 5.1. Log-lineer modeller ve lojistik regresyon'un istatistiksel özelliklerinin karşılaştırılması [7]

Özellik	Log-lineer Modeller	Lojistik Regresyon
Metodolojik varsayımlar	Birden fazla örnekleme planı vardır. • Poisson: Toplam birim sayısı n ve her bir hücredeki birim sayısı n_{ijk} poisson dağılımına sahiptir. • Çokterimli: Toplam birim sayısı n sabitlenmiştir ve her bir hücredeki birim sayısı n_{ijk} çokterimli dağılıma sahiptir. • Çarpımsal çokterimli: Bağımsız değişkenlerin her bir kategorisindeki gözlem sayıları her satırda sabitlenmiştir ($n_{ij.}$). Burada her satır bağımsız çokterimli dağılıma sahiptir. • Hipergeometrik: Hem satır toplamları hem sütun toplamları sabitlenmiştir. • Retrospective çarpımsal binom: Sadece bağımlı değişkene ait olan sütun toplamları sabitlenir.	• Hata terimleri binom dağılımına sahiptir. • Bağımlı değişken için log-odds'lar, bağımsız değişkenlerin doğrusal bir fonksiyonudur. • Kesişim katsayısı, her bir bağımsız değişkenin bağımlı değişkenle ilişkisini belirtir.
Bağımsız değişkenlerin birlikteliğini ölçme	Kısmi ve marjinal birliktelikler kullanılır.	İkili korelasyon matrisinden elde edilir.
Birden çok bağımlı değişken ile birden çok bağımsız değişkenin arasındaki ilişkiyi ölçme	Birden çok bağımlı değişken kullanılarak her bir bağımlı değişken ile bağımsız değişkenler setinin arasındaki birliktelik eş zamanlı olarak ölçülür.	Tek bir bağımlılık modeli ile sıralı testlerin serisi oluşturulur.

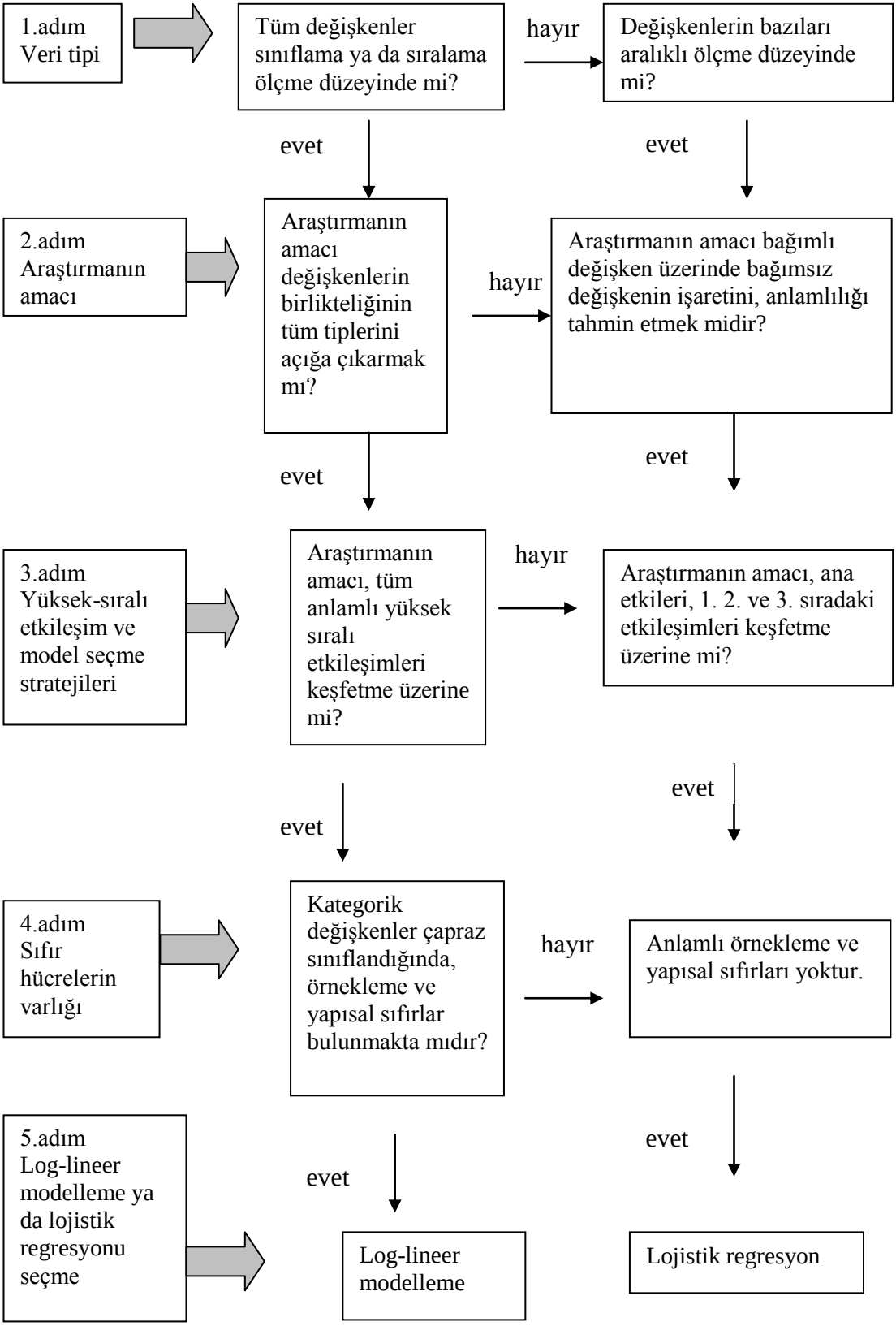
Çizelge 5.1. (Devam) Log-lineer modeller ve lojistik regresyon'un istatistiksel özelliklerinin karşılaştırılması [7]

Özellik	Log-lineer Modeller	Lojistik Regresyon
Sıfırlı hücreleri tanımlama ve problemleri çözme	• örnekleme ve yapısal sıfır hücreleri farklı değerlendirilir. • örnekleme sıfırları için maksimum olabilirlik yöntemi ile beklenen hücre frekansları tahmin edilir.	• Sıfırlı hücrelerin arasında hiçbir ayrım yapılmaz. • Sıfırlı hücre problemlerini ortadan kaldırmak için Bölüm 5.1'de belirtilen "ad hoc çözümleri" kullanılabilir.
Kategoriyi birleştirme	a- İki değişken arasındaki marjinal birliktelik, üçüncü değişkenin kategorileri birleştirilerek hesaplanır. b- Üçüncü değişkenin kategorilerini birleştirmek için G^2 koşullu bağımsızlık testi kullanılır. (Koşullu bağımsızlık, verilen z değişkeni için x ve y nin bağımsız olmadığını söyler, fakat sadece x ve y yi göz önünde tutarsak aslında bağımsızlardır.)	Belirli kuralları yoktur. Sınıflama /sıralama ölçme düzeyine sahip değişken için, sıfırlı hücreyi kapsayan bir ya da daha çok kategori "ad hoc çözümleri" ile ortadan kaldırılır.
Bir bağımsız değişkenin, bağımlı değişken üzerindeki etkisini ölçme	İki değişken arasındaki birlikteliği ölçmek için lambda (λ) ağırlıkları elde edilir. Ancak geriye kalan bağımsız değişkenlerin etkilerini kontrol ederek bir bağımsız değişkenin, bağımlı değişken üzerindeki etkisinin ne kadar olduğu ölçülemez.	Her bir bağımsız değişkenin bağımlı değişken üzerindeki etkisinin işaretini, büyüklüğünü ve standart hatalarını tahmin etmek için, Wald ve olabilirlik oran test istatistikleri kullanılır.
Bağımlı değişken için koşullu odds oranını tahmin etme	Log-lineer modellerle doğrudan elde edilemezler; log-lineer modelin eşit bir lojit modele dönüşmesi gereklidir.	Bağımlı değişken için koşullu oddslar ve odds'ların standart hataları tahmin edilebilir.
Nicel ve nitel değişken içeren karma veri seti için veri analizi	Log-lineer modeller, sınıflama ve sıralama ölçme düzeyine sahip veriler için sınırlandırılmıştır.	Bu tip karma veri setlerinin istatistiksel analizlerinde güçlü bir araçtır.

Kaynaklar:(Agresti,1984;Andersen,1980;Hosmer&Lemeshow,1989;Everett,1977; Fienberg,1980;Santner&Duffy,1989;Upton,1978)

Çizelge 5.2. Log-lineer modeller ile lojistik regresyonun bağımlı ve bağımsız değişkenler arasındaki birlikteliği test etmedeki farklılıkları [7]

	Log-lineer modeller	Lojistik regresyon
Anlamlı bir bağımlı – bağımsız değişken ilişkisine karar vermede kullanılan testler	2 aşamalı süreçtir: a- Bağımlı – bağımsız değişken birlikteliğinin anlamlılığı Pearson ya da olabilirlik oran test istatistikleri ile test edilir. b- Bağımlı – bağımsız değişkenin kategorileri arasındaki ilişkinin anlamlılığı wald z testi ile test edilir.	a ya da b'yi kullanarak bir aşamalıdır. a- Keşisim katsayısı b için H_0:b=0 hipotezi wald istatistiği ile test edilir. Bu bir z ya da ki-kare istatistiği gibi dağılır [33]. b- Olabilirlik oran istatistiği G^2'yi anlamlı olarak azaltarak tüm modelin uyum iyiliğini yükselten, bir bağımsız değişkeni ekleyen ya da çıkartan karşılaştırılabilir bir test önerir.
Bir bağımlı – bağımsız değişken arasındaki ilişkinin doğası hakkındaki varsayımlar	Doğrusallık ya da tekdüzelik hakkında hiçbir varsayım yoktur.	Bağımsız değişkenler ile log odds arasında doğrusal birliktelik vardır.
Parametre katsayılarını yorumlanabilmesi	Wald z testi sadece, bağımlı – bağımsız değişkenin kategorileri arasındaki ilişkinin anlamlılığını test ettiğinde sınırlı yorumlama yapılabilir. Bu iki değişken arasındaki birlikteliğin büyüklüğünü söyleyemez.	Wald istatistiği, a- Bağımlı değişkenin olasılığının ya da ilişkili oddslarının üzerinde bağımsız değişkenin bir birim değişmesinin etkisinin büyüklüğü ölçer. b- İlişkinin yönünü belirler.



Şekil 5.1. Lojistik regresyona karşı log-lineer modelleri seçme akış şeması [7]

6. HANEHALKI BÜTÇE ANKETİ ÜZERİNE BİR UYGULAMA

6.1. Hanehalkı Bütçe Anketi

Türkiye İstatistik Kurumu tarafından, hanelerin sosyo-ekonomik yapıları, yaşam düzeyleri, tüketim kalıpları hakkında bilgi elde etmek için Hanehalkı Bütçe Anketi uygulanmaktadır. Hanehalkı Bütçe Anketi, toplumun ihtiyaçlarının belirlenmesi, kullanılabilir gelirin haneler ya da fertler arasında ne şekilde dağıldığının belirlenmesi ve uygulanan sosyo-ekonomik politikaların geçerliliğinin test edilmesi amacıyla kullanılan en önemli kaynaklardan biridir [34].

2009 yılı Hanehalkı Bütçe Anketi, 1 Ocak–31 Aralık 2009 tarihleri arasında bir yıl süre ile her ay değişen aylık toplam 1050, yıllık toplam 12600 örnek hanehalkına uygulanarak bireylerin ve bunların oluşturduğu hanehalklarının tüketim yapılarını, gelir düzeylerini; sosyo-ekonomik gruplara, kır, kent ve bölgelere göre ortaya çıkarır. Ayrıca hanehalkının sosyo-ekonomik özellikleri, tüketim alışkanlıkları, tüketim harcaması türleri ile mal ve hizmet harcamalarının çeşitliliği, hanehalkı fertlerinin çalışma durumları, hanehalkının toplam geliri, gelirin elde edildiği kaynaklar vb. konular hakkında bilgiler derlenmektedir [34].

Anket çalışmasında kapsanan kitle Türkiye Cumhuriyeti sınırları içinde yaşayan tüm hanehalkı fertleridir. Ancak çalışmada kurumsal nüfus olarak tanımlanan yaşlılar evi, huzur evleri, hapishane, askeri kışla, özel nitelikli hastane, otel, çocuk yuvalarında bulunan nüfus ile uygulama zorluğundan dolayı göçer nüfus kapsam dışı tutulmuştur [34].

6.2. Değişkenler

Log-lineer modeller ve lojistik regresyonun karşılaştırılması için uygulamada sadece anket yapılan ayda çalışan ve düzenli ücrete sahip olan, örnekleme rasgele seçilen 6196 ferdin bilgisi kullanılacaktır. Ele alınan değişkenler ferdin aylık ücret geliri,

cinsiyeti, medeni durumu, yaşı, eğitimi ve haftalık çalışma süresidir. Değişken tanımları ve kategorileri aşağıdaki gibidir.

1) Cinsiyet (C): 1: Erkek, 2: Kadın

2) Yaş (Y): Hanehalkı fertlerinin anket ayında bitirmiş oldukları yaş grubu verilmiştir.

Çizelge 6.1. Hanehalkı fertlerinin anket ayında bitirmiş oldukları yaş grubu

Yaş aralığı	Atanan değer
15-24	1
25-34	2
35-44	3
45-54	4
55-64	5
65+	6

3) Eğitim durumu (E): Diploması alınan öğretim kurumu kaydedilmiştir.

Çizelge 6.2. Diploması alınan öğretim kurumu

Eğitim düzeyi	Atanan değer
Okur yazar değil	1
Okur yazar fakat okul bitirmemiş	2
İlkokul	3
İlköğretim-ortaokul	4
Lise	5
İki yıllık fakülte	6
Dört yıllık fakülte	7
Yüksek lisans ve doktora	8

4) Medeni durum (MD): 1: Evli, 2: Bekar.

5) Aylık Ücret Geliri (U): Bu gelir, maaş ücret veya yevmiye adı altında kişilere ödenen; emeklilik, sosyal sigortalar keseneği ve vergiler hariç, eline geçen net gelir olarak kapsanmıştır.

• Alınan iş riski, iş güçlüğü zammı, fazla mesai, prim ve ek görevlerden elde edilen gelirler maaş, ücret ve yevmiye gelirine dahil edilmiştir ve alınan vergi iadeleri transfer gelirlerinde kapsanmıştır. Ferdin ikramiye gelirleri varsa, alınan ikramiyeden aya düşen kısmı kaydedilmiştir [34].

6) Haftalık çalışma süresi(CS): Esas işteki haftalık normal çalışma süresi saat cinsinden kapsanmıştır.

• Fazla çalışma saatini kapsar ancak, öğle yemek aralarını ve kişinin işe gidip gelirken harcadığı zamanı içermez.

• Kişi kesin bir çalışma saati veremediği durumda, son dört haftadaki fiili çalışma süresinin ortalaması esas alınmıştır.

• Kendi işinde çalışanlar için, evde çalışılan planlama, rapor tutmada harcanan saatler dahil edilmiştir.

• Haftalık çalışma süresi 35 saatin altında olanlar Tam zamanlı (part-time), 35 ve daha fazla olanlar ise Tam zamanlı (full time) olarak değerlendirilmiştir. Tam zamanlı (full time) için "1", Yarı zamanlı (part time) için "2" kodu verilmiştir.

"Haftalık çalışma süresi" değişkeni uygulamamızda lojistik regresyon için nicel, lojit ve log-lineer modeller için nitel değişken olarak kullanacağız.

6.3. Log-lineer Modeller ile Lojistik Regresyonun Karşılaştırılması

Uygulamamızda log-lineer modeller ya da lojistik regresyondan hangisinin kullanılacağına Şekil 5.1'deki akış şeması ile karar vereceğiz.

6.3.1. Log-lineer modellerin uygulaması

1. adım: Tüm değişkenler sınıflama ya da sıralama ölçme düzeyinde mi?

Tüm değişkenler kategoriktir. Haftalık çalışma süresi part-time ve full-time olmak üzere iki kategoriye ayrılmıştır.

2. adım: Araştırmanın amacı değişkenlerin birlikteliğinin tüm tiplerini açığa çıkarmak mı?

Bağımlı değişken belirlemeden tüm değişkenlerin arasındaki ilişki açığa çıkarılmak istenmektedir.

3. adım: Araştırmanın amacı tüm anlamlı yüksek etkileşimleri ortaya koymak mı?

Tüm değişkenler arasındaki olası tüm etkileşimler ortaya çıkarılmak istenmektedir.

4.adım: Değişkenler çapraz sınıflandığında örnekleme ya da yapısal sıfırlı hücreler bulunuyor mu?

Asgari ücrete göre değerlendirilen ücret değişkeni için sonuç Çizelge 6.3'deki gibidir. Bu çizelgede cinsiyet, yaş, eğitim, medeni durum, haftalık çalışma saati ve aylık ücretin kategorileri belirtilmiştir. 2009 asgari ücretinden (486,855 TL) fazla geliri olanlara 1, az geliri olanlara 2 kodu verilmiştir. Bu değişkenlere ait olumsallık tablosu Ek 1'dedir. "Hücreler" kısmında belirtildiği gibi 418 örnekleme sıfırı bulunmaktadır. Bu durumda Şekil 5.1'deki akış şemasına göre log-lineer modeller kullanılmalıdır.

Çizelge 6.3. Asgari ücrete göre verinin genel durumu

		N
Gözlemler	Geçerli	6196
	Kayıp	0
Hücreler		
	Tanımlı hücreler	768
	Yapısal sıfırlar	0
	Örnekleme sıfırları	418
Kategoriler	cinsiyet	2
	yaş	6
	eğitim	8
	medeni durum	2
	haftalık çalışma süresi	2
	aylik ucret geliri	2

Log-lineer modellerde geriye eleme yöntemi ile 10 adımda seçilen model Çizelge 6.4'deki gibidir. Bu çizelgede etkiler bölümünde seçilen en yüksek etkileşim terimleri vardır. Bu etkileşim terimleri içinde barındırdıkları değişkenlerin daha alt etkileşimlerini de içerirler. Örneğin U*Y*E değişkeni U, Y, E, U*Y, U*E, Y*E değişkenlerini de içermektedir.

Çizelge 6.4. Log-lineer modellerde en son aşamada seçilen model

Adım		Etkiler	χ^2	sd	P değeri	A_x-q
10	Model	U*C*E*MD*CS, Y*E*MD*CS, C*Y*E*CS, U*C*Y*MD, U*C*Y*CS, U*Y*MD*CS, C*Y*MD*CS, U*Y*E	103,407	320	1,000	-536,593

Modelin anlamlılığının test edilmesi

H_0: model anlamlıdır.

H_1: model anlamlı değildir.

H_0 hipotezinin testi için Çizelge 6.5'deki uyum iyiliği test sonuçlarına bakılacak olursa hem olabilirlik oran hem de Pearson χ^2 testine göre H_0 hipotezi reddedilemez ve sonuç olarak model anlamlı bulunur.

Çizelge 6.5. Log-lineer modellerde en son aşamada seçilen modelin uyum iyiliği testi

	χ^2	sd	P değeri
Olabilirlik oran G^2	103,407	320	1,000
Pearson	93,204	320	1,000

Odds oranlarının yorumlanması

Çizelge 6.3'de belirtilen değişkenlerimize ait olumsallık tablosu ile verimiz hakkında çeşitli yorumlar yapabiliriz. Bu yorumlar için daha önce de belirtildiği gibi marjinal alt tablolar ve odds oranları kullanılır. Bu kısımda birkaç örnekle odds oranlarının yorumuna değinilecektir.

Örnek 6.1: Ele alınan alt grup 15-24 yaş aralığında lise mezunu bekar bayanlardır.

Çizelge 6.6. 15-24 yaş aralığında lise mezunu bekar bayanların haftalık çalışma süresi ve ücretlerinin olumsallık tablosu

Kadın, 15-24 yaş aralığında, lise mezunu, bekar	Asgari ücretten fazla ücret alıyor	Asgari ücretten az ücret alıyor	Toplam
Tam zamanlı	78	49	127
Yarı zamanlı	2	7	9
Toplam	80	56	136

Eş. 2.5'de belirtilen odds oranı formülü için önce olasılıkların hesaplanması gerekmektedir.

Olasılıklar:

p_{11}= 78/ 127= 0,614

p_{12}= 49/ 127= 0,386

1. satır için odds : $\Omega_1 = p_{11}/ p_{12}$=0,614/0,386= 1,59067

p_{21}= 2/ 9= 0,222

p_{22}= 7/ 9= 0,778

2. satır için odds : $\Omega_2 = p_{21}/ p_{22}$= 0,222/0,778=0,28535

Odds oranı: $\Omega = \Omega_1/ \Omega_2$= 1,591/0,285=5,57

Bu değer ayrıca Eş. 6.1 yardımıyla kolayca bulunabilir.

Odds oranı: Ω=(78x7)/(2x 49)=546/98=5,57 (6.1)

Bu durumda "15-24 yaş aralığında lise mezunu bekar bayanlardan full-time çalışanların asgari ücretten fazla ücret alma oranı, part-time çalışanların asgari ücretten fazla ücret alma oranından 5.57 katıdır" yorumu yapılabilir. Bu zaten beklenilen bir sonuçtur.

Örnek 6.2: Ele alınan alt grup 15-24 yaş aralığında lise mezunu bayanlardır.

Çizelge 6.7. 15-24 yaş aralığında lise mezunu bayanların medeni durumu ve ücretlerinin olumsallık tablosu

Kadın, 15-24 yaş aralığında, lise mezunu	Asgari ücretten fazla ücret alıyor	Asgari ücretten az ücret alıyor	Toplam
Evli	17	7	24
Bekar	80	56	136
Toplam	97	63	136

Örnek 6.1'de olduğu gibi önce olasılıkların hesaplanması gerekmektedir.

Olasılıklar:

p_{11}= 17/24=0,708

p_{12}= 7/24=0,292

1. satır için odds : $\Omega_1 = p_{11}/p_{12}$ = 0,708/0,292=2,43

p_{21}= 80/136=0,588

p_{22}= 56/136=0,412

2. satır için odds : $\Omega_2 = p_{21}/p_{22}$= 0,588/0,412=1,427

Odds oranı: $\Omega = \Omega_1/\Omega_2$= 2,43/1,427=1,703

Bu durumda "15-24 yaş aralığında lise mezunu bayanlardan evli olanların asgari ücretten fazla ücret alma oranı, bekarların asgari ücretten fazla ücret alma oranının 1,703 katıdır" yorumu yapılabilir.

Örnek 6.3: Ele alınan alt grup 25-34 yaş aralığındaki erkek nüfustur.

Çizelge 6.8. 25-34 yaş aralığındaki erkeklerin eğitim ve ücretlerinin olumsallık tablosu

25-34 yaş aralığında, erkek	Asgari ücretten fazla ücret alıyor	Asgari ücretten az ücret alıyor	Toplam
Dört yıllık üniversite mezunu	251	3	254
Yüksek lisans veya doktora mezunu	31	1	32
Toplam	282	4	286

Olasılıklar:

p_{11}= 251/254= 0,988

p_{12}= 3/254= 0,012

1. satır için odds : $\Omega_1 = p_{11}/p_{12}$ = 0,988/0,012=82,33

p_{21}= 31/32=0,969

p_{22}= 1/32=0,031

2. satır için odds : Ω_2= p_{21}/ p_{22}= 0,969/0,031=31,258

Odds oranı: Ω= Ω_1/ Ω_2= 82,33/31,258=2,63

Bu durumda "25-34 yaş aralığındaki erkeklerden dört yıllık üniversite mezunlarının asgari ücretten fazla ücret alma oranı, yüksek lisans veya doktora mezunlarının asgari ücretten fazla ücret alma oranının 1,703 katıdır" yorumu yapılabilir. Bu gerçekten çarpıcı bir sonuçtur.

6.3.2. Lojistik regresyon uygulaması

1.adım: Tüm değişkenler sınıflama ya da sıralama ölçme düzeyinde mi?

Haftalık çalışma süresi, sürekli değişken olmak üzere işleme alınmıştır. Yani tüm değişkenler kategorik değildir.

2. adım: Araştırmanın amacı değişkenlerin birlikteliğinin tüm tiplerini açığa çıkarmak mı?

Araştırmanın amacı, iki değer alan aylık ücret değişkeni üzerinde bağımsız değişkenlerin etkisinin tahmin edilmesidir.

3. adım: Ana etkileri, 1.,2. ve 3. Sıralı etkileşimleri mi ortaya çıkarmak istiyoruz?

Evet, tüm sıralı etkileşimler ile ilgilenilmemektedir.

4. adım: Kategorik değişkenler çapraz sınıflandığında örnekleme ve yapısal sıfır bulunuyor mu?

Çizelge 6.3'de belirtildiği gibi 418 örnekleme sıfırı bulunmaktadır. Bu durumda Bölüm 5.1 de belirtilen "ad-hoc çözümleri" kullanılabilir. Burada herhangi bir çözümleme yapılmadan lojistik regresyon uygulanmıştır.

Lojistik regresyon ile en son elde edilen model Eş. 6.2'deki gibidir. Aylık ücret gelir değişkeni bağımlı değişken olacak şekilde p_i, bireyin asgari ücretten fazla ücret alma olasılığını belirtsin. Dolayısıyla $1-p_i$, bireyin asgari ücretten az ücret alma olasılığını belirtir.

Eş. 6.2'de parantez içindeki ifadeler değişken kategorilerini belirtir. Ek 2'de ayrıca parametre tahminleri ile Wald testi sonuçları mevcuttur. α=0,05 olmak üzere istatistiksel olarak anlamsız bulunan değişkenler modele dahil edilmemiştir.

$$\log(p_i/(1-p_i)) = -4{,}50\ C(1) - 2{,}420\ Y(2) - 2{,}444\ Y(3) - 2{,}268\ Y(4) + 8{,}732\ E(1) + 7{,}748\ E(2) + 6{,}161\ E(3) + 5{,}461\ E(4) + 3{,}512\ E(5) + 1{,}770\ E(6) - 1{,}980\ MD(1) - 1{,}175\ (MD*C) - 0{,}031\ (C*CS) + 0{,}341\ (Y*E) + 0{,}557\ (E*MD) - 0{,}197\ (E*MD*C) + 0{,}007\ (C*E*CS) - 0{,}259\ (Y*E*MD) + 0{,}115\ (Y*E*MD*C) - 0{,}002\ (C*E*Y*CS) \quad (6.2)$$

Geriye eleme yöntemi ile hem log-lineer modeller hem de lojistik regresyon için belirlenen en son modelde yer alan değişkenler Çizelge 6.9'da özetlenmiştir.

Çizelge 6.9. Asgari ücret sınırına göre log-lineer modeller ve lojistik regresyonun model seçimleri

Log-lineer modeller	Lojistik regresyon
U*C*E*MD*CS, **Y*E*MD*CS,** *C*Y*E*CS,* U***C*Y*MD**, U***C*Y*CS**, U***Y*MD*CS,** **C*Y*MD*CS**, U**Y*E*	C, Y, E, MD, MD*C C*CS *Y*E* E*MD E*MD*C C*E*CS Y*E*MD **Y*E*MD*C** *C*E*Y*CS*
Ax-q: -536,593	AIC: 2279,122
G^2: 103,407	G^2: 4495,110
Sd: 320	Sd:13

Bağımlı değişken aylık ücret geliri U'yu içeren değişkenlerin dışındaki değişkenleri incelediğimizde C*Y*MD*CS etkileşimi log-lineer modelde anlamlı iken bu etkileşim lojistik regresyonda modele alınmamıştır. Bunun gibi, modeller tarafından ortak seçilmeyen değişkenler koyu renk ile belirtilmiştir. İtalik altı çizgili şekilde yazılanlar aynı değişkenleri, normal karakterler daha yüksek etkileşimlerde zaten içerilen değişkenleri belirtir. Örneğin lojistik regresyonda bağımlı değişkeni açıklamada anlamlı bulunan Y*E*MD değişkeni, log-lineer modelde Y*E*MD*CS değişkeni içinde kapsanmaktadır. Bu durumda tüm değişkenleri modele alan log-lineer modeller ile daha yüksek etkileşimli terimlerin içerildiğini söyleyebiliriz.

6.4. Log-Lineer Modeller ile Lojit Modelin Karşılaştırması

Bu bölümde amacımız lojistik regresyonda tüm değişkenlerin kategorik olduğu durumda log-lineer modeller ile ilişkisini göstermektir. Lojistik regresyonda haftalık

çalışma süresi tam zamanlı (Full time) ve yarı-zamanlı (part time) olmak üzere iki kategoriye ayrılmıştır ve lojistik regresyon "Lojit model" olarak adlandırılmıştır.

Lojit ve log-lineer modellerin geriye eleme yöntemiyle en son aşamada seçtikleri modeller Çizelge 6.10'da belirtilmiştir. Burada da ortak olarak belirlenen değişkenler altı çizgili ve italik karakterle, farklı seçilenler koyu renkte ve daha yüksek etkileşimlerde zaten var olan değişkenler normal karakterle gösterilmiştir.

Dikkat edilecek olursa Çizelge 6.9'da da olduğu gibi bağımlı değişken aylık ücret geliri U'yu içeren değişkenlerin dışındaki değişkenleri incelediğimizde C*Y*MD*CS etkileşimi log-lineer modelde anlamlı iken bu etkileşim lojit modele alınmamıştır.

Çizelge 6.10. Asgari ücret sınırına göre log-lineer modeller ve lojit model seçimleri

Log-lineer modeller	Lojit
U*C*E*MD*CS, **Y*E*MD*CS,** *C*Y*E*CS,* U**C*Y*MD,* U***C*Y*CS**, U***Y*MD*CS,** **C*Y*MD*CS**, U***Y*E**	SABİT C, Y, MD, Y*MD, E*MD, MD*CS, C*Y, C*MD, C*CS, *C*Y*MD,* C*E*MD, C*E*CS, **C*Y*E*MD,** *C*Y*E*CS,*
Ax-q: -536,593	AIC: **699,792**
G^2: 103,407	G^2: **1202,714**
Sd: 320	Sd:14

Lojit modele ait parametre tahminleri ve Wald test istatistikleri Ek 3'de verilmiştir.

Lojit modele göre seçilen en son model Eş. 6.3'deki gibidir ancak α=0,05 olmak üzere istatistiksel olarak anlamsız bulunan cinsiyet, medeni durum ve haftalık çalışma süresi etkileşim terimi C*MD*CS modele dahil edilmemiştir.

$\log(p_i/(1-p_i))$ = -14,819 + 6,204 C + 3,707 Y + 7,636 MD - 2,567 Y*MD – 0,267 E*MD + 1,155 MD*CS – 2,069 C*Y – 3,978 C*MD + 2,039 C*CS + 1,394 C*Y*MD + 0,266 C*E*MD – 0,377 C*E*CS – 0,079 C*Y*E*MD + 0,041 C*Y*E*CS (6.3)

6.4.1. İki boyutlu tablolarda lojit model ve log-lineer modellerin denkliğinin gösterilmesi

Bu kısımda, Bölüm 3.3'de bahsedilen iki boyutlu olumsallık tablolarında lojit model ve log-lineer modellerin eşitliği, eğitim ve aylık ücret değişkenleri üzerine bir örnekle gösterilmiştir.

Eş. 3.44'ü tekrar yazalım. Burada ücret değişkeni bağımlı değişken olacak şekilde p_i, bireyin asgari ücretten fazla ücret alma olasılığını; $1-p_i$, bireyin asgari ücretten az ücret alma olasılığını belirtir.

$$\begin{aligned}\log\left(\frac{p_i}{1-p_i}\right) &= \log\left(\frac{m_{i1}}{m_{i2}}\right)\\ &= \log(m_{i1}) - \log(m_{i2})\\ &= (\lambda_{(i)}^{E} + \lambda_{(1)}^{Ü} + \lambda_{(i1)}^{Ü*E}) - (\lambda_{(i)}^{E} + \lambda_{(2)}^{Ü} + \lambda_{(i2)}^{Ü*E})\\ &= (\lambda_{(1)}^{Ü} - \lambda_{(2)}^{Ü}) + (\lambda_{(i1)}^{Ü*E} - \lambda_{(i2)}^{Ü*E})\\ &= \beta_0 + \beta_1 x_E \end{aligned} \quad (6.4)$$

Görüldüğü üzere lojit model, log-lineer modellerle de elde edilebilir. SPSS paket programı yardımıyla bu eşitliği göstermeye çalışacağız. İlk önce $\log(m_{ij}) = \lambda + \lambda^{U} + \lambda^{E} + \lambda^{U*E}$ doymuş log-lineer model parametreleri tahmin edilecek ve daha sonra lojit modele ilişkin parametrelerle karşılaştırılacaktır.

Çizelge 6.11. Eğitim* Aylık Ücret olumsallık tablosu

		Aylık ücret geliri	
		asgari ucretten fazla	asgari ucretten az
		Kişi Sayısı	Kişi Sayısı
eğitim	[1]:okur_yazar_değil	32	24
	[2]:okur_yazar_okul bitirmemiş	80	63
	[3]:ilkokul	1494	268
	[4]:ilköğretim-ortaokul	719	299
	[5]lise	1546	195
	[6]: iki yıllık fakülte	465	28
	[7]dört yıllık fakülte	855	19
	[8]yükseklisans_doktora	106	3

$\log(m_{ij}) = \lambda + \lambda^{U} + \lambda^{E} + \lambda^{U*E}$ doymuş log-lineer modeli için parametre tahminleri Çizelge 6.12'deki gibidir. Burada U[1] yani asgari ücretten fazla alma durumu için λ katsayısı 3.415 bulunmuştur ve referans kategori olan asgari ücretten az alma durumu U[2] sıfırdır. Aynı şekilde her bir kategori için etkileşimi gösteren λ katsayıları, her bir katsayının anlamlılığın testi (Wald Z) hesaplanmıştır. Örneğin [U = 1] * [E = 1] yani asgari ücretten fazla alan okur yazar olmama durumu için λ katsayısı -3,133 bulunmuştur.

Çizelge 6.12. $\log(m_{ij}) = \lambda + \lambda^{U} + \lambda^{E} + \lambda^{U*E}$ doymuş log-lineer modeli için parametre tahminleri

Parametre	Tahmin	Std. Hata	Wald Z	Sig.	95% Güven aralığı	
					Alt sınır	Üst sınır
Sabit	1,253	,535	2,344	,019	,205	2,300
[U = 1]	**3,415**	,543	6,287	,000	2,351	4,480
[U = 2]	0	.	.	.	.	.
[E = 1]	1,946	,571	3,405	,001	,826	3,066
[E = 2]	2,898	,549	5,279	,000	1,822	3,974
[E = 3]	4,340	,538	8,067	,000	3,286	5,395
[E = 4]	4,449	,538	8,276	,000	3,396	5,503
[E = 5]	4,023	,539	7,460	,000	2,966	5,080
[E = 6]	2,097	,566	3,703	,000	,987	3,207
[E = 7]	1,718	,580	2,959	,003	,580	2,855
[E = 8]	0	.	.	.	.	.
[U = 1] * [E = 1]	**-3,133**	,606	-5,174	,000	-4,320	-1,946
[U = 1] * [E = 2]	**-3,178**	,569	-5,590	,000	-4,293	-2,064
[U = 1] * [E = 3]	**-1,699**	,547	-3,104	,002	-2,771	-,626
[U = 1] * [E = 4]	**-2,539**	,548	-4,637	,000	-3,612	-1,466
[U = 1] * [E = 5]	**-1,347**	,548	-2,456	,014	-2,422	-,272
[U = 1] * [E = 6]	**-,622**	,576	-1,079	,280	-1,752	,508
[U = 1] * [E = 7]	**,366**	,590	,621	,535	-,790	1,521
[U = 1] * [E = 8]	0(a)	.	.	.	.	.
[U = 2] * [E = 1]	0(a)	.	.	.	.	.
[U = 2] * [E = 2]	0(a)	.	.	.	.	.
[U = 2] * [E = 3]	0(a)	.	.	.	.	.
[U = 2] * [E = 4]	0(a)	.	.	.	.	.
[U = 2] * [E = 5]	0(a)	.	.	.	.	.
[U = 2] * [E = 6]	0(a)	.	.	.	.	.
[U = 2] * [E = 7]	0(a)	.	.	.	.	.
[U = 2] * [E = 8]	0(a)	.	.	.	.	.

Bağımlı değişken "aylık ücret geliri" ve bağımsız değişken "eğitim" için lojit model parametre tahminleri Çizelge 6.13'de belirtilmiştir. Burada da referans kategori, değişkenlerin en son kategorisidir.

Çizelge 6.13. Bağımlı değişken "aylık ücret geliri" ve bağımsız değişken "eğitim" için lojit model parametre tahminleri

Aylık Ücret		B	Std. Hata	Wald	sd	P değeri	Exp(B)= odds oranı	Exp (B) için %95 Güven Aralığı	
								Alt sınır	Üst sınır
asgari ucretten fazla	Sabit	**3,565**	,585	37,075	1	,000			
	[E=1]	**-3,277**	,645	25,836	1	,000	**,038**	,011	,134
	[E=2]	**-3,326**	,609	29,805	1	,000	**,036**	,011	,119
	[E=3]	**-1,847**	,589	9,822	1	,002	**,158**	,050	,501
	[E=4]	**-2,687**	,589	20,783	1	,000	**,068**	,021	,216
	[E=5]	**-1,494**	,590	6,407	1	,011	**,224**	,071	,714
	[E=6]	**-,755**	,617	1,498	1	,221	**,470**	,140	1,575
	[E=7]	**,242**	,630	,147	1	,701	**1,274**	,371	4,376
	[E=8]	**0(b)**	**.**	**.**	**0**	**.**	**.**	**.**	**.**

Tahmin edilen parametreler B sütununda, odds oranları Exp(B) sütununda belirtilmiştir. Ayrıca parametre tahminlerine ilişkin standart hatalar, tahminin anlamlılığının sınanması için wald istatistiği ve P değeri, odds oranlarına ilişkin güven aralıkları da bu tabloda yer almaktadır.

Elde edilen lojit model Eş. 6.5 şeklinde yazılabilir ancak α=0,05 olmak üzere istatistiksel olarak anlamsız bulunan iki yıllık ve dört yıllık fakülteden mezun olma durumu E[6] ve E[7] değişkenleri modele dahil edilmemiştir.

$$\log(p_i/(1-p_i)) = 3{,}565 - 3{,}277\ E(1) - 3{,}326\ E(2) - 1{,}847\ E(3) - 2{,}687\ E(4) - 1{,}494\ E(5) \qquad (6.5)$$

Dikkat edilecek olursa, Eş. 6.5'de $\hat{\beta}_0$ değeri 3,565 olarak tahmin edilmiştir ve bu değer log-lineer model tarafından tahmin edilen [U=1]= 3,415 değerine yaklaşık olarak eşittir (Bkz. Çizelge 6.12). Ayrıca $\hat{\beta}_1$ tahmini -3,277 değeri [U=1]*[E=1] etkileşimi -3,133 değerine yaklaşık eşit çıkmıştır.

Burada dikkat edilmesi gereken diğer bir eşitlik iki boyutlu tablolarda log-lineer modellere ait parametre tahminlerinin odds oranı (Exp(B)) ye eşitliğidir. Bu durum hem marjinal alt tablo yardımıyla hem de Çizelge 6.13'de gösterilmiştir.
U(1)*E(1) yani okur yazar olmayanların asgari ücretten fazla ücret alma durumu için marjinal alt tablo Çizelge 6.14'de belirtilmiştir. Burada referans kategori son kategori olarak yükseklisans ve doktoralılar kategorisidir.

Odds oranı: θ= 32*3/106*24= **0,0377** olarak hesaplanır.

Çizelge 6.14. Eğitim* Aylık Ücret Geliri olumsallık tablosu

	Aylık Ücret Geliri	
Eğitim	asgari ücretten fazla[1]	asgari ücretten az[2]
[1]okur_yazar_değil	32	24
[8]yükseklisans_doktora	106	3

Çizelge 6.12'de [U = 1] * [E = 1] için hesaplanan λ katsayısı -3,133'dür.

$$e^{\lambda U[1]*E[1]} = e^{-3,133} = 0,04358 \sim \mathbf{0,0377} = \theta = \text{odds oranı} \qquad (6.6)$$

Aynı eşitliği doğrudan Çizelge 6.13'de E[1] için Exp(B) yani odds oranı kısmında da görebiliriz. Bu değer **0,038** olarak bulunmuştur. Bu oran 1/θ=1/0,038=26,53 olarak bulunabilir ve "yükseklisans- doktora mezunlarının asgari ücretten fazla ücret alma oranı, okur yazar olmayan bireylerin asgari ücretten fazla ücret alma oranının yaklaşık 27 katıdır." diye yorumlanabilir.

6.4.2. Üç boyutlu tablolarda lojit model ve log-lineer modellerin denkliğinin gösterilmesi

Bölüm 4.3'de belirtilmiş olan üç boyutlu tablolarda lojit model ile log-lineer modellerin eşitliği gösterilmiştir. Bu bölümde de her iki yöntem tarafından tahmin edilen parametrelerin eşitliği, Çizelge 6.15'deki eşitlikler yardımıyla gösterilecektir. Log-lineer modellerde tüm değişkenler, bağımlı – bağımsız değişken ayrımı

yapılmadan modele alınırken, lojit modelde aylık ücret geliri [U] bağımlı değişken, yaş ve eğitim değişkenleri bağımsız değişken olarak modele alınmıştır.

Çizelge 6.15. Lojit- log lineer modeller eşitliği

Log-lineer model	Lojit
[UY][E]; [UY][YE]	$= \beta_0 + \beta_Y$
[UE][Y] ; [UE][YE]	$= \beta_0 + \beta_E$

[UY] [E] = [UY][YE] $= \beta_0 + \beta_Y$ *eşitliğinin gösterilmesi*

[UY] [E] = [UY][YE] $= \beta_0 + \beta_Y$ eşitliğinin gösterilmesi için ilk önce tüm modeller için parametre tahminleri elde edilecektir.

- [UY] [E] log-lineer modeli $\log(m_{ijk}) = \lambda + \lambda^U + \lambda^Y + \lambda^E + \lambda^{U*Y}$ şeklinde yazılabilir. Bu modele ait parametre tahminleri Çizelge 6.16'da verilmiştir.

Çizelge 6.16. $\log(m_{ijk}) = \lambda + \lambda^U + \lambda^Y + \lambda^E + \lambda^{U*Y}$ modelinin parametre tahmini

Parametre	Tahmin	Std. Hata	Wald Z	P değeri	95% Güven aralığı	
					Alt sınır	Üst sınır
Sabit	-1,555	,304	-5,119	,000	-2,151	-,960
[E = 1]	-,666	,164	-4,051	,000	-,988	-,344
[E = 2]	,271	,127	2,135	,033	,022	,521
[E = 3]	2,783	,099	28,195	,000	2,589	2,976
[E = 4]	2,234	,101	22,170	,000	2,037	2,432
[E = 5]	2,771	,099	28,064	,000	2,577	2,964
[E = 6]	1,509	,106	14,259	,000	1,302	1,717
[E = 7]	2,082	,102	20,494	,000	1,883	2,281
[E = 8]	0(a)	.	.	.	.	.
[U = 1]	**-,182**	,428	-,426	,670	-1,021	,657
[U = 2]	0(a)	.	.	.	.	.
[Y = 1]	3,574	,293	12,213	,000	3,001	4,148
[Y = 2]	2,725	,298	9,144	,000	2,141	3,309
[Y = 3]	2,512	,300	8,371	,000	1,924	3,101

Çizelge 6.16. (Devam) $\log(m_{ijk}) = \lambda + \lambda^{U} + \lambda^{Y} + \lambda^{E} + \lambda^{U*Y}$ modelinin parametre tahmini

Parametre	Tahmin	Std. Hata	Wald Z	P değeri	95% Güven aralığı	
					Alt sınır	Üst sınır
[Y = 4]	1,910	,309	6,174	,000	1,303	2,516
[Y = 5]	1,365	,323	4,222	,000	,731	1,999
[Y = 6]	0(a)	.	.	.	.	.
[U = 1] * [Y = 1]	**,698**	,432	1,615	,106	-,149	1,546
[U = 1] * [Y = 2]	**2,561**	,435	5,888	,000	1,709	3,414
[U = 1] * [Y = 3]	**2,590**	,437	5,933	,000	1,734	3,446
[U = 1] * [Y = 4]	**2,485**	,444	5,601	,000	1,615	3,354
[U = 1] * [Y = 5]	**1,281**	,460	2,785	,005	,379	2,183
[U = 1] * [Y = 6]	0(a)	.	.	.	.	.
[U = 2] * [Y = 1]	0(a)	.	.	.	.	.
[U = 2] * [Y = 2]	0(a)	.	.	.	.	.
[U = 2] * [Y = 3]	0(a)	.	.	.	.	.
[U = 2] * [Y = 4]	0(a)	.	.	.	.	.
[U = 2] * [Y = 5]	0(a)	.	.	.	.	.
[U = 2] * [Y = 6]	0(a)	.	.	.	.	.

- [UY][YE] log-lineer modeli $\log(m_{ijk}) = \lambda + \lambda^{U} + \lambda^{Y} + \lambda^{E} + \lambda^{U*Y} + \lambda^{Y*E}$ şeklinde yazılabilir. Bu modele ait parametre tahminleri Çizelge 6.17'de verilmiştir.

Çizelge 6.17. $\log(m_{ijk}) = \lambda + \lambda^{U} + \lambda^{Y} + \lambda^{E} + \lambda^{U*Y} + \lambda^{Y*E}$ modelinin parametre tahmini

Parametre	Tahmin	Std. Hata	Wald Z	P değ.	95% Güven aralığı	
					Alt sınır	Üst sınır
Sabit	,087	,733	,119	,906	-1,350	1,524
[E = 1]	3,30E-014	1,000	,000	1,000	-1,960	1,960
[E = 2]	3,04E-014	1,000	,000	1,000	-1,960	1,960
[E = 3]	1,792	,764	2,346	,019	,295	3,289
[E = 4]	-,693	1,225	-,566	,571	-3,094	1,707
[E = 5]	,405	,913	,444	,657	-1,384	2,195
[E = 6]	-15,332	915,344	-,017	,987	-1809,373	1778,709
[E = 7]	-15,332	915,344	-,017	,987	-1809,373	1778,709
[E = 8]	0(a)	.	.	.	.	.
[U = 1]	**-,182**	,428	-,426	,670	-1,022	,657
[U = 2]	0(a)	.	.	.	.	.
[Y = 1]	-,378	1,019	-,371	,711	-2,376	1,620
[Y = 2]	1,506	,748	2,012	,044	,039	2,973
[Y = 3]	,677	,763	,887	,375	-,819	2,173
[Y = 4]	,348	,780	,447	,655	-1,180	1,876

Çizelge 6.17. (Devam) $\log(m_{ijk}) = \lambda + \lambda^{U} + \lambda^{Y} + \lambda^{E} + \lambda^{U*Y} + \lambda^{Y*E}$ modelinin parametre tahmini

Parametre	Tahmin	Std. Hata	Wald Z	P değ.	95% Güven aralığı	
					Alt sınır	Üst sınır
[Y = 5]	-,087	,897	-,097	,923	-1,844	1,670
[Y = 6]	0(a)	.	.	.	.	.
[Y = 1] * [E = 1]	-2,52E-015	1,414	,000	1,000	-2,772	2,772
[Y = 1] * [E = 2]	3,481	1,231	2,828	,005	1,069	5,894
[Y = 1] * [E = 3]	1,846	1,047	1,763	,078	-,207	3,898
[Y = 1] * [E = 4]	6,073	1,415	4,292	,000	3,300	8,846
[Y = 1] * [E = 5]	4,930	1,156	4,265	,000	2,664	7,195
[Y = 1] * [E = 6]	19,192	915,344	,021	,983	-1774,849	1813,234
[Y = 1] * [E = 7]	18,664	915,344	,020	,984	-1775,378	1812,706
[Y = 1] * [E = 8]	0(a)	.	.	.	.	.
[Y = 2] * [E = 1]	-1,863	1,062	-1,754	,079	-3,945	,219
[Y = 2] * [E = 2]	-,925	1,030	-,898	,369	-2,944	1,094
[Y = 2] * [E = 3]	,490	,776	,631	,528	-1,031	2,011
[Y = 2] * [E = 4]	2,071	1,234	1,679	,093	-,347	4,488
[Y = 2] * [E = 5]	2,047	,923	2,218	,027	,238	3,857
[Y = 2] * [E = 6]	16,447	915,344	,018	,986	-1777,594	1810,489
[Y = 2] * [E = 7]	17,309	915,344	,019	,985	-1776,732	1811,350
[Y = 2] * [E = 8]	0(a)	.	.	.	.	.
[Y = 3] * [E = 1]	-,314	1,045	-,300	,764	-2,361	1,734
[Y = 3] * [E = 2]	,074	1,036	,072	,943	-1,957	2,105
[Y = 3] * [E = 3]	1,515	,789	1,920	,055	-,032	3,063
[Y = 3] * [E = 4]	2,782	1,242	2,240	,025	,347	5,217
[Y = 3] * [E = 5]	2,370	,935	2,534	,011	,537	4,202
[Y = 3] * [E = 6]	17,050	915,344	,019	,985	-1776,991	1811,092
[Y = 3] * [E = 7]	17,542	915,344	,019	,985	-1776,499	1811,583
[Y = 3] * [E = 8]	0(a)	.	.	.	.	.
[Y = 4] * [E = 1]	,163	1,053	,154	,877	-1,901	2,226
[Y = 4] * [E = 2]	,111	1,054	,106	,916	-1,955	2,178
[Y = 4] * [E = 3]	1,140	,803	1,419	,156	-,434	2,715
[Y = 4] * [E = 4]	2,622	1,252	2,094	,036	,168	5,076
[Y = 4] * [E = 5]	2,065	,947	2,180	,029	,208	3,921
[Y = 4] * [E = 6]	16,657	915,344	,018	,985	-1777,384	1810,699
[Y = 4] * [E = 7]	17,396	915,344	,019	,985	-1776,645	1811,438
[Y = 4] * [E = 8]	0(a)	.	.	.	.	.
[Y = 5] * [E = 1]	-3,23E-014	1,225	,000	1,000	-2,400	2,400
[Y = 5] * [E = 2]	,405	1,190	,341	,733	-1,927	2,738
[Y = 5] * [E = 3]	1,166	,920	1,267	,205	-,637	2,969
[Y = 5] * [E = 4]	2,565	1,337	1,918	,055	-,056	5,186
[Y = 5] * [E = 5]	1,642	1,056	1,555	,120	-,428	3,712
[Y = 5] * [E = 6]	16,430	915,344	,018	,986	-1777,611	1810,472
[Y = 5] * [E = 7]	17,278	915,344	,019	,985	-1776,764	1811,319
[Y = 5] * [E = 8]	0(a)	.	.	.	.	.

Çizelge 6.17. (Devam) $\log(m_{ijk}) = \lambda + \lambda^{U} + \lambda^{Y} + \lambda^{E} + \lambda^{U*Y} + \lambda^{Y*E}$ modelinin parametre tahmini

Parametre	Tahmin	Std. Hata	Wald Z	P değ.	95% Güven aralığı	
					Alt sınır	Üst sınır
[Y = 6] * [E = 1]	0(a)	.	.	.	.	.
[Y = 6] * [E = 2]	0(a)	.	.	.	.	.
[Y = 6] * [E = 3]	0(a)	.	.	.	.	.
[Y = 6] * [E = 4]	0(a)	.	.	.	.	.
[Y = 6] * [E = 5]	0(a)	.	.	.	.	.
[Y = 6] * [E = 6]	0(a)	.	.	.	.	.
[Y = 6] * [E = 7]	0(a)	.	.	.	.	.
[Y = 6] * [E = 8]	0(a)	.	.	.	.	.
[U = 1] * [Y = 1]	**,698**	,433	1,614	,106	-,149	1,546
[U = 1] * [Y = 2]	**2,561**	,435	5,886	,000	1,708	3,414
[U = 1] * [Y = 3]	**2,590**	,437	5,931	,000	1,734	3,446
[U = 1] * [Y = 4]	**2,485**	,444	5,600	,000	1,615	3,355
[U = 1] * [Y = 5]	**1,281**	,460	2,784	,005	,379	2,183
[U = 1] * [Y = 6]	**0(a)**	.	.	.	.	.
[U = 2] * [Y = 1]	0(a)	.	.	.	.	.
[U = 2] * [Y = 2]	0(a)	.	.	.	.	.
[U = 2] * [Y = 3]	0(a)	.	.	.	.	.
[U = 2] * [Y = 4]	0(a)	.	.	.	.	.
[U = 2] * [Y = 5]	0(a)	.	.	.	.	.
[U = 2] * [Y = 6]	0(a)	.	.	.	.	.

- $\log(p_i/1-p_i) = \hat{\beta}_0 + \hat{\beta}_Y$ lojit modeli için tahmin edilen parametreler Çizelge 6.18'de belirtilmiştir.

Çizelge 6.18. Bağımlı değişken "aylık ücret geliri" ve bağımsız değişken "yaş" için lojit model parametre tahminleri

Aylık Ücret Geliri		B	Std. Hata	Wald	sd	P değeri	Exp(B)= odds oranı	Exp (B) için %95 Güven Aralığı	
								Alt sınır	Üst sınır
asgari ucretten fazla	Sabit	**-,182**	,428	,181	1	,670			
	[Y=1]	**,698**	,433	2,607	1	,106	2,010	,861	4,693
	[Y=2]	**2,561**	,435	34,651	1	,000	12,951	5,520	30,384
	[Y=3]	**2,590**	,437	35,176	1	,000	13,330	5,664	31,371
	[Y=4]	**2,485**	,444	31,358	1	,000	12,000	5,029	28,635
	[Y=5]	**1,281**	,460	7,750	1	,005	3,600	1,461	8,871
	[Y=6]	**0(b)**	.	.	0	.	.	.	.

Lojit model Eş. 6.7'deki gibi tahmin edilmiştir. α=0,05 olmak üzere istatistiksel olarak anlamsız bulunan değişkenler modele dahil edilmemiştir.

$$\log(p_i/(1-p_i)) = -0{,}182 + 0{,}698\ Y(1) + 2{,}561\ Y(2) + 2{,}590\ Y(3) + 2{,}485\ Y(4) + 1{,}281\ Y(5) \quad (6.7)$$

Çizelge 6.16 ve 6.17'de hem [UY][E] hem de [UY][YE] log lineer modelleri için tahmin edilen U[1] yani asgari ücretten fazla alma durumu parametre tahmini -0,182 bulunmuştur ve bu değer lojit modelde sabit etki değeri -0,182' ye eşittir. Aynı zamanda hem [UY][E] hem de [UY][YE] log lineer modellerinde [U][Y] yani ücret, yaş etkileşimi için belirlenen λ katsayıları lojit modeldeki yaş değişkenine ait parametre tahminlerine eşit bulunmuştur. Örneğin; her iki loglineer modelde de asgari ücretten fazla alan 15-24 yaş aralığında olma [U=1][Y=1] için λ tahmini (0,698) ,Çizelge 6.18'deki [Y=1] tahmini (0,698)'e eşittir. Bu eşitlik diğer kategoriler için de görülebilir. Ayrıca Bölüm 6.4.1'de odds oranlarının eşitliğine ilişkin yorumlar burada da geçerlidir. $e^{\lambda_{U[1]*Y[1]}} = e^{0,698} = 2,0097$'nin Çizelge 6.18'deki Exp(B) sütununa bakılırsa [Y=1] tahmininin odds oranının, "2,010" değerine eşit olduğu görülebilir.

[UE] [Y] = [UE][YE] $= \beta_0 + \beta_E$ *eşitliğinin gösterilmesi*

[UE] [Y] = [UE][YE] $= \beta_0 + \beta_E$ eşitliğinin gösterilmesi için ilk önce tüm modeller için parametre tahminleri yapılacaktır.

- [UE] [Y] log-lineer modeli $\log(m_{ijk}) = \lambda + \lambda^U + \lambda^Y + \lambda^E + \lambda^{U*E}$ şeklinde yazılabilir. Bu modele ait parametre tahminleri Çizelge 6.19'da verilmiştir.

Çizelge 6.19. $\log(m_{ijk}) = \lambda + \lambda^{U} + \lambda^{Y} + \lambda^{E} + \lambda^{U*E}$ modelinin parametre tahmini

Parametre	Tahmin	Std. Hata	Wald Z	P değ.	95% Güven Aralığı	
					Alt sınır	Üst sınır
Sabit	-4,542	,615	-7,383	,000	-5,748	-3,336
[E = 1]	2,079	,612	3,396	,001	,879	3,279
[E = 2]	3,045	,591	5,153	,000	1,887	4,202
[E = 3]	4,492	,580	7,740	,000	3,355	5,630
[E = 4]	4,602	,580	7,933	,000	3,465	5,739
[E = 5]	4,174	,582	7,177	,000	3,034	5,314
[E = 6]	2,234	,607	3,678	,000	1,043	3,424
[E = 7]	1,846	,621	2,972	,003	,628	3,063
[E = 8]	0(a)	.	.	.	.	.
[U = 1]	**3,565**	,585	6,090	,000	2,418	4,712
[U = 2]	0(a)	.	.	.	.	.
[Y = 1]	3,952	,215	18,362	,000	3,530	4,374
[Y = 2]	4,586	,214	21,401	,000	4,166	5,006
[Y = 3]	4,400	,215	20,513	,000	3,980	4,820
[Y = 4]	3,701	,216	17,150	,000	3,278	4,124
[Y = 5]	2,145	,225	9,521	,000	1,704	2,587
[Y = 6]	0(a)	.	.	.	.	.
[U = 1] * [E = 1]	**-3,277**	,645	-5,084	,000	-4,541	-2,014
[U = 1] * [E = 2]	**-3,326**	,609	-5,461	,000	-4,520	-2,132
[U = 1] * [E = 3]	**-1,847**	,589	-3,135	,002	-3,001	-,692
[U = 1] * [E = 4]	**-2,687**	,589	-4,560	,000	-3,843	-1,532
[U = 1] * [E = 5]	**-1,494**	,590	-2,532	,011	-2,651	-,338
[U = 1] * [E = 6]	**-,755**	,617	-1,224	,221	-1,964	,454
[U = 1] * [E = 7]	**,242**	,630	,384	,701	-,992	1,476
[U = 1] * [E = 8]	**0(a)**	.	.	.	.	.
[U = 2] * [E = 1]	0(a)	.	.	.	.	.
[U = 2] * [E = 2]	0(a)	.	.	.	.	.
[U = 2] * [E = 3]	0(a)	.	.	.	.	.
[U = 2] * [E = 4]	0(a)	.	.	.	.	.
[U = 2] * [E = 5]	0(a)	.	.	.	.	.
[U = 2] * [E = 6]	0(a)	.	.	.	.	.
[U = 2] * [E = 7]	0(a)	.	.	.	.	.
[U = 2] * [E = 8]	0(a)	.	.	.	.	.

- [UE] [YE] log-lineer modeli $\log(m_{ijk}) = \lambda + \lambda^{U} + \lambda^{Y} + \lambda^{E} + \lambda^{U*E} + \lambda^{Y*E}$ şeklinde yazılabilir. Bu modele ait parametre tahminleri Çizelge 6.20'de verilmiştir.

Çizelge 6.20. $\log(m_{ijk}) = \lambda + \lambda^{U} + \lambda^{Y} + \lambda^{E} + \lambda^{U*E} + \lambda^{Y*E}$ modelinin parametre tahmini

Parametre	Tahmin	Std. Hata	Wald Z	P değ.	95% Güven Aralığı	
					Alt sınır	Üst sınır
Sabit	-2,900	,908	-3,194	,001	-4,679	-1,120
[E = 1]	2,745	1,161	2,365	,018	,470	5,021
[E = 2]	2,773	1,155	2,402	,016	,510	5,036
[E = 3]	3,501	,954	3,669	,000	1,631	5,372
[E = 4]	1,674	1,351	1,239	,215	-,974	4,323
[E = 5]	1,809	1,078	1,678	,093	-,304	3,922
[E = 6]	-14,439	841,506	-,017	,986	-1663,761	1634,883
[E = 7]	-15,322	809,528	-,019	,985	-1601,967	1571,323
[E = 8]	0(a)	.	.	.	.	.
[U = 1]	**3,565**	,585	6,089	,000	2,417	4,712
[U = 2]	0(a)	.	.	.	.	.
[Y = 1]	2,68E-014	1,000	,000	1,000	-1,960	1,960
[Y = 2]	3,367	,719	4,682	,000	1,958	4,777
[Y = 3]	2,565	,734	3,495	,000	1,127	4,003
[Y = 4]	2,140	,748	2,863	,004	,675	3,605
[Y = 5]	,693	,866	,800	,423	-1,004	2,391
[Y = 6]	0(a)	.	.	.	.	.
[U = 1] * [E = 1]	**-3,277**	,645	-5,083	,000	-4,541	-2,013
[U = 1] * [E = 2]	**-3,326**	,609	-5,459	,000	-4,520	-2,132
[U = 1] * [E = 3]	**-1,847**	,589	-3,134	,002	-3,001	-,692
[U = 1] * [E = 4]	**-2,687**	,589	-4,559	,000	-3,843	-1,532
[U = 1] * [E = 5]	**-1,494**	,590	-2,531	,011	-2,652	-,337
[U = 1] * [E = 6]	**-,755**	,617	-1,224	,221	-1,964	,454
[U = 1] * [E = 7]	**,242**	,630	,384	,701	-,992	1,476
[U = 1] * [E = 8]	**0(a)**	.	.	.	.	.
[U = 2] * [E = 1]	0(a)	.	.	.	.	.
[U = 2] * [E = 2]	0(a)	.	.	.	.	.
[U = 2] * [E = 3]	0(a)	.	.	.	.	.
[U = 2] * [E = 4]	0(a)	.	.	.	.	.
[U = 2] * [E = 5]	0(a)	.	.	.	.	.
[U = 2] * [E = 6]	0(a)	.	.	.	.	.
[U = 2] * [E = 7]	0(a)	.	.	.	.	.
[U = 2] * [E = 8]	0(a)	.	.	.	.	.
[Y = 1] * [E = 1]	-2,68E-014	1,414	,000	1,000	-2,772	2,772
[Y = 1] * [E = 2]	3,481	1,231	2,828	,005	1,069	5,894
[Y = 1] * [E = 3]	1,846	1,047	1,763	,078	-,207	3,898
[Y = 1] * [E = 4]	6,073	1,415	4,292	,000	3,300	8,846
[Y = 1] * [E = 5]	4,930	1,156	4,265	,000	2,664	7,195
[Y = 1] * [E = 6]	19,024	841,506	,023	,982	-1630,298	1668,346
[Y = 1] * [E = 7]	18,418	809,528	,023	,982	-1568,227	1605,064
[Y = 1] * [E = 8]	0(a)	.	.	.	.	.
[Y = 2] * [E = 1]	-1,863	1,062	-1,754	,079	-3,945	,219

Çizelge 6.20. (Devam) $\log(m_{ijk}) = \lambda + \lambda^{U} + \lambda^{Y} + \lambda^{E} + \lambda^{U*E} + \lambda^{Y*E}$ modelinin parametre tahmini

Parametre	Tahmin	Std. Hata	Wald Z	Sig.	95% Güven Aralığı	
					Alt sınır	Üst sınır
[Y = 2] * [E = 2]	-,925	1,030	-,898	,369	-2,944	1,094
[Y = 2] * [E = 3]	,490	,776	,631	,528	-1,031	2,011
[Y = 2] * [E = 4]	2,071	1,234	1,679	,093	-,347	4,488
[Y = 2] * [E = 5]	2,047	,923	2,218	,027	,238	3,857
[Y = 2] * [E = 6]	16,279	841,506	,019	,985	-1633,042	1665,601
[Y = 2] * [E = 7]	17,063	809,528	,021	,983	-1569,581	1603,708
[Y = 2] * [E = 8]	0(a)	.	.	.	.	.
[Y = 3] * [E = 1]	-,314	1,045	-,300	,764	-2,361	1,734
[Y = 3] * [E = 2]	,074	1,036	,072	,943	-1,957	2,105
[Y = 3] * [E = 3]	1,515	,789	1,920	,055	-,032	3,063
[Y = 3] * [E = 4]	2,782	1,242	2,240	,025	,347	5,217
[Y = 3] * [E = 5]	2,370	,935	2,534	,011	,537	4,202
[Y = 3] * [E = 6]	16,882	841,506	,020	,984	-1632,439	1666,204
[Y = 3] * [E = 7]	17,296	809,528	,021	,983	-1569,349	1603,941
[Y = 3] * [E = 8]	0(a)	.	.	.	.	.
[Y = 4] * [E = 1]	,163	1,053	,154	,877	-1,901	2,226
[Y = 4] * [E = 2]	,111	1,054	,106	,916	-1,955	2,178
[Y = 4] * [E = 3]	1,140	,803	1,419	,156	-,434	2,715
[Y = 4] * [E = 4]	2,622	1,252	2,094	,036	,168	5,076
[Y = 4] * [E = 5]	2,065	,947	2,180	,029	,208	3,921
[Y = 4] * [E = 6]	16,489	841,506	,020	,984	-1632,832	1665,811
[Y = 4] * [E = 7]	17,151	809,528	,021	,983	-1569,494	1603,796
[Y = 4] * [E = 8]	0(a)	.	.	.	.	.
[Y = 5] * [E = 1]	-3,10E-015	1,225	,000	1,000	-2,400	2,400
[Y = 5] * [E = 2]	,405	1,190	,341	,733	-1,927	2,738
[Y = 5] * [E = 3]	1,166	,920	1,267	,205	-,637	2,969
[Y = 5] * [E = 4]	2,565	1,337	1,918	,055	-,056	5,186
[Y = 5] * [E = 5]	1,642	1,056	1,555	,120	-,428	3,712
[Y = 5] * [E = 6]	16,262	841,506	,019	,985	-1633,060	1665,584
[Y = 5] * [E = 7]	17,032	809,528	,021	,983	-1569,613	1603,677
[Y = 5] * [E = 8]	0(a)	.	.	.	.	.
[Y = 6] * [E = 1]	0(a)	.	.	.	.	.
[Y = 6] * [E = 2]	0(a)	.	.	.	.	.
[Y = 6] * [E = 3]	0(a)	.	.	.	.	.
[Y = 6] * [E = 4]	0(a)	.	.	.	.	.
[Y = 6] * [E = 5]	0(a)	.	.	.	.	.
[Y = 6] * [E = 6]	0(a)	.	.	.	.	.
[Y = 6] * [E = 7]	0(a)	.	.	.	.	.
[Y = 6] * [E = 8]	0(a)	.	.	.	.	.

- $\log(p_i/1-p_i) = \hat{\beta}_0 + \hat{\beta}_E$ lojit modeli için tahmin edilen parametreler Çizelge 6.21'de belirtilmiştir.

Çizelge 6.21. Bağımlı değişken "aylık ücret geliri" ve bağımsız değişken "eğitim" için lojit model parametre tahminleri

Aylık Ücret		B	Std. Hata	Wald	sd	P değ.	Exp(B)= odds oranı	Exp (B) için %95 Güven Aralığı	
								Alt sınır	Üst sınır
asgari ucretten fazla	Intercept	**3,565**	,585	37,075	1	,000			
	[E=1]	**-3,277**	,645	25,836	1	,000	,038	,011	,134
	[E=2]	**-3,326**	,609	29,805	1	,000	,036	,011	,119
	[E=3]	**-1,847**	,589	9,822	1	,002	,158	,050	,501
	[E=4]	**-2,687**	,589	20,783	1	,000	,068	,021	,216
	[E=5]	**-1,494**	,590	6,407	1	,011	,224	,071	,714
	[E=6]	**-,755**	,617	1,498	1	,221	,470	,140	1,575
	[E=7]	**,242**	,630	,147	1	,701	1,274	,371	4,376
	[E=8]	**0(b)**	.	.	0	.	.	.	.

Lojit model Eş. 6.8'deki gibi tahmin edilmiştir. α=0,05 olmak üzere istatistiksel olarak anlamsız bulunan değişkenler modele dahil edilmemiştir.

$$\log(p_i/(1-p_i)) = 3{,}565 - 3{,}277\ E(1) - 3{,}326\ E(2)\ -1{,}847\ E(3) - 2{,}687\ E(4\ )\ -1{,}494\ E(5) \quad (6.8)$$

Çizelge 6.19 ve 6.20'de hem [UE][Y] hem de [UE][YE] log lineer modelleri için tahmin edilen U[1] yani asgari ücretten fazla alma durumu parametre tahmini 3,565 bulunmuştur ve bu değer lojit modelde sabit etki değeri 3,565'e eşittir (Bkz. Çizelge 6.21). Aynı zamanda hem [UE][Y] hem de [UE][YE] log lineer modellerinde [U][E] yani ücret, eğitim etkileşimi için belirlenen λ katsayıları lojit modeldeki yaş değişkenine ait parametre tahminlerine eşit bulunmuştur. Örneğin; her iki loglineer modelde de "asgari ücretten fazla alan okuryazar olmayan"ı belirten [U=1][E=1] için λ tahmini (-3,277) ,Çizelge 6.21'deki [E=1] tahmini -3,277'e eşittir. Bu eşitlik diğer kategoriler için de görülebilir.

Ayrıca Bölüm 6.4.1'de odds oranlarının eşitliğine ilişkin yorumlar burada da geçerlidir. $e^{\lambda_{U[1]*E[1]}} = e^{-3,277} = 0,03774$ 'ün Çizelge 6.21'deki [E=1] tahmininin odds oranı "0,038" değerine eşit olduğu görülebilir. Sonuç olarak log-lineer modellerde koşullu odds oranlarının hesaplanması için eşit bir lojit modele dönüştülmesi gerektiğini söyleyebiliriz.

7. SONUÇ VE ÖNERİLER

Bu çalışmanın amacı, kategorik veri analizinde sıklıkla kullanılan lojistik regresyon, lojit model ve log-lineer modellerin istatistiksel olarak karşılaştırılmasıdır. İlk önce lojistik regresyon ile log-lineer modellerin tanımı, kullanım amaçları, istatistiksel varsayımları, birbirlerine göre avantaj ve dezavantajları ele alınmış daha sonra Türkiye İstatistik Kurumu tarafından düzenlenen 2009 Hanehalkı Bütçe anket verileri ile iki uygulama yapılmıştır.

İlk uygulamada lojistik regresyon ile log-lineer modeller, ikinci uygulamada lojit model ile log-lineer modeller ele alınmış ve en son aşamada modele seçilen değişkenler karşılaştırılmıştır. Elde edilen bulgulara göre log-lineer modeller, bağımlı bağımsız değişken ayrımı yapmadan tüm değişkenler arasındaki ilişkiyi ölçmede, lojistik regresyon ve lojit modele karşı üstünlük sağlamaktadır. Ancak amacımız bağımlı değişken üzerinde bağımsız değişkenlerin etkisini ölçmek ise lojistik regresyon tercih edilmelidir. Ayrıca bağımsız değişkenlerin ölçüm düzeylerine göre modellerin farklılık gösterdiği vurgulanmıştır. Lojistik regresyonda, bağımsız değişkenler sınıflama, sıralama ve eşit aralıklı ölçüm düzeyine sahip olabilirken, log-lineer modellerde değişkenler sadece sınıflama ya da sıralama ölçüm düzeyine sahiptir. Diğer bir ayrım da sıfırlı hücrelerin varlığı durumda ortaya çıkmaktadır. Log-lineer modeller yapısal ve örnekleme sıfırlarını birbirinden ayırır ve maksimum olabilirlik tahminlerini elde edebilir. Ancak lojistik regresyon sıfırlı hücreler arasında böyle bir ayrım yapamaz. Çözüm olarak ad-hoc çözümleri'nin kullanılabileceği belirtilmiştir.

Uygulama kısmı ele alınan tüm bağımsız değişkenlerin kategorik değişken olması ve olmaması durumuna göre ikiye ayrılmıştır. İlk uygulamada aylık ücret geliri üzerinde etkili olduğu düşünülen bağımsız değişkenlerden "haftalık çalışma süresi" nicel diğerleri nitel değişken olarak modele alınmıştır. Lojistik regresyon ya da log-lineer modellerden hangisinin uygulanması gerektiğine bir akış şeması ile karar verilmiştir. İkinci uygulamada ise tüm değişkenlerin kategorik olması durumunda lojit modellerin, log-lineer modellerin özel bir durumu olduğu gösterilmiştir. İki ve üç

boyutlu tablolarda lojit model için tahmin edilen parametre tahminlerinin log-lineer modeller ile de elde edilebileceği belirtilmiş ve iki farklı log-lineer modelin aynı lojit modele eşit olabileceği tespit edilmiştir. Log-lineer modeller için koşullu oddsların hesaplanması da log-lineer modelin eşit bir lojit modele dönüştürülmesi ile gerçekleşmektedir.

Sonuç olarak log-lineer modeller ya da lojistik regresyondan hangisinin uygulama için seçilmesi gerektiğine çalışmanın genel amacına ve değişkenlerin ölçüm düzeylerine göre karar verilmelidir.

KAYNAKLAR

1. Christensen, R ., "Log-Linear Models and Logistic Regression 2nd ed.", ***Springer- Verlag***, New York, 1-246, 279-290(1997).

2. Powers, D. A., Xie, Y., "Statistical Metods For Categorical Data Analysis 2nd ed.", ***Academic Press***, England, 41-119 (2008).

3. Çilan, Ç. A., "Sosyal Bilimlerde Kategorik Verilerle İlişki Analizi Kontenjans Tabloları Analizi 1. Baskı ",***Pegem Akademi***, Ankara, 5-163 (2009).

4. Simonoff, S. J., "Analyzing Categorical Data 1nd ed.", ***Springer- Verlag*** ,New York, 55-102 (2003).

5. Erbaş,O. S., "Olasılık ve İstatistik 2.baskı", ***Gazi Kitabevi***,Ankara , 170-202 (2008).

6. Fienberg, E. S., "Analysis Of Cross-Classified Categorical Data 2nd ed.", ***Springer***, New York , 71-116 (2007).

7. Tansey, R., White, M., Long, R.G., Smith, M., "A Comparison of Loglinear Modeling and Logistic Regression in Management Research", ***Journal Of Management*** , Vol. 22. No. 2: 339- 358 (1996).

8. Kleinbaum, D. G., Klein, M., "Logistic Regression A Self-Learning Set 3rd ed.", ***Springer***, New York , 4-28, 107-122 (2010).

9. Gujarati, D. N., "Temel Ekonometri 1. baskı" , Şeneşen, Ü., Şeneşen, G., ***Literatür Yayıncılık***, İstanbul , 282-311 (1999).

10. Ürük, E., "İstatistiksel Uygulamalarda Lojistik Regresyon Analizi", Yüksek lisans, ***Marmara Üniversitesi Fen Bilimleri Enstitüsü***, İstanbul, 1-42 (2007).

11. Agresti, A., "Categorical Data Analysis 2nd ed.", ***Wiley-Interscience,*** New Jersey, 1-357 (2002).

12. Agresti, A., "Analysis of Ordinal categorical ", ***John Wiley and Sons***, New York, 286 (1984).

13. Everitt, S.B., Dunn, G., "Applied multivariate data analysis", ***John Wiley and Sons***, New York, 304 (1991).

14. Becanım, C., "Log-Lineer Modeller ve Doktor-Tıbbi Satış Mümessilleri İlişkileri Üzerine Bir Uygulama", Yüksek lisans, ***Gazi Üniversitesi Fen Bilimleri Enstitüsü***, Ankara, 4-25 (2006).

15. İnternet: Vikipedi özgür ansiklopedi
http: // en.wikipedia.org/wiki

16. Myung, I. J., "Tutorial on maximum likelihood estimation", ***Journal of Mathematical Pscyhology***, 47: 90-100 (2003).

17. Hanson, T. E., Johnson, W.O., Gardner, I. A., "Log-linear and lojistic modeling of dependence among diagnostic tests", ***Preventive Veterinary Medicine***, 45: 123- 137 (2000).

18. Büyükbebeci, E., "İşgücü durumunu belirleyen değişkenlerin log-lineer modellerle van bölge örneğinde analizi", ***TÜİK Uzmanlık Tezi***, Van, 19-41 (2009).

19. Şahin, M., "Lojistik Regresyon ve Biyolojik Alanlarda Kullanımı", Yüksek lisans, ***Kahramanmaraş Sütçü İmam Üniversitesi Fen Bilimleri Enstitüsü***, Kahramanmaraş, 9-41(1999).

20. Harrell, F.E., "Bölüm 9.3.1",Regression Modeling Strategies, ***Springer-Verlag***, New York, 186-187 (2001).

21. Knoke, D., Burke, P., "Log-lineer Models", ***Sage University***, United States of America, 63-65 (1980).

22. Xu, J., Long, J. S., "Using the Delta Method to Construct Confidence Intervals for Predicted Probabilities, Rates, and Discrete Changes", ***Indiana University***, United States of America, 1-8 (2005).

23. Junker, B., "Log-Linear Models: Three-Way Tables", 36-720 (2007). http://www.stat.cmu.edu/~brian/720/week03/04-loglinear.pdf

24. Conover , W.J., "Practical Nonparametric Statistics 3. Baskı", ***John Wiley & Sons***, United States of America ,179-269 (1999).

25. Hosmer, D. W., Lemeshow, S., "Applied logistic regression 2. Baskı ", ***John Wiley&Sons*** , United States of America, 1-31, 47-260(2000).

26- Greasley, P., "Quantitative Data Analysis Using SPSS ", ***Open University Press***, England, 61-75 (2008)

27. White, M., Tansey, R.. Smith, M., Barnett, T., "Log-linear modeling in personnel research" ***Personnel Psychology***, 46 : 667-686 (1993).

28. White, M., Tansey, R. & Smith. M., "A causal modeling test of the relationship between Chief Executive Officer experience and corporate strategy", ***Jounal of Occupational and Organizational Psychology***, 67: 259-278 (1994).

29. Lee, S., "On the asymptotic variance of $\hat{u}$ terms in log-linear models of multidimensional contingency tables", ***Journal of American Statistical Association***, 72: 412-419 (1977).

30. Morgan, S., Teachman, J., "Logistic regression: Description, examples and comparisons", ***Journal of Marriage and the Family***, 50: 929-936 (1989).

31. Hirji, K., Tsiatis, A., Mehta. C., "Median unbiased estimation for binary data", ***The American Statistician***, 43: 7-l1 (1989).

32. Aldrich, J., Nelson, F., "Linear probability, logit and probit models", ***Sage University Paper Series on Quantitative Applications in the Social Sciences***, Beverly Hills, 7-45 (1984).

33. Miller, R.L, Acton, C., Fullerton, D.A., Maltby, J., "SPSS for Social Scientists ", ***Polgrave Mactillan***, New York, 186-213, (2002).

34. TÜİK 2009 Hanehalkı Bütçe Anketi Mikro Veri CD tanımları

35. Anderson, E.B., "The statistical analysis of categorical data", ***Springer- Verlag***, Berlin, 520 (1990).

36. Tatlıdil, H., "Uygulamalı Çok Değişkenli İstatistiksel Analiz" , ***Akademi 1***, Ankara, 289-324 (1996)

37. Özdamar, K., "Paket Programlar ile İstatistiksel Veri Analizi 1", ***Kaan Kitabevi***, Eskişehir, 589-618 (2004).

38. Blasius, J.,Greenacre, M., "Vizualization of Categorical Data", ***Academic Press***, USA, 32-49 (1998).

39. Wooldridge, J. M., "Introductory econometrics: a modern approach", ***Cengage Learning***, Canada, 761(2009).

40. Harrell, F.E., "Bölüm 9.2", Regression Modeling Strategies, ***Springer-Verlag***, New York, 183-186 (2001).

41. Harrell, F.E., "Bölüm 9.3.3", Regression Modeling Strategies, ***Springer-Verlag***, New York, 189-190 (2001).

42. Green, P.E., Carmone, F. J., Wachspress, D. P., "On The Analysis of Qualitative Data in Marketing Research ", ***Journal of Marketing Research***, Vol. XIV: 52-9 (1977).

EKLER

EK- 1. Asgari ücrete göre olumsallık tablosu

Çizelge 1.1. Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			aylık ücret		Toplam
						asgari ucretten fazla	asgari ucretten az	
erkek	15-24	okur yazar değil	bekar	haftalık çalışma süresi	full time		1	1
					part time		1	1
				toplam			2	2
		Okur yazar ancak okul bitirmemiş	evli	haftalık çalışma süresi	full time	3	3	6
				Toplam		3	3	6
			bekar	haftalık çalışma süresi	full time	10	19	29
					part time	1	2	3
				Toplam		11	21	32
		ilkokul	evli	haftalık çalışma süresi	full time	12	3	15
				Toplam		12	3	15
			bekar	haftalık çalışma süresi	full time	35	7	42
					part time	1	0	1
				Toplam		36	7	43

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
		ilkogretim-ortaokul	evli	haftalık çalışma süresi	full time	31	3	34
					part time	1	0	1
				Toplam		32	3	35
			bekar	haftalık çalışma süresi	full time	122	142	264
					part time	2	22	24
				Toplam		124	164	288
		lise	evli	haftalık çalışma süresi	full time	39	3	42
				Toplam		39	3	42
			bekar	haftalık çalışma süresi	full time	159	39	198
					part time	1	14	15
				Toplam		160	53	213
		iki yıllık fakülte	evli	haftalık çalışma süresi	full time	2		2
				Toplam		2		2
			bekar	haftalık çalışma süresi	full time	36	6	42
					part time	1	2	3
				Toplam		37	8	45
		dört yıllık fakulte	evli	haftalık çalışma süresi	full time	3		3
				Toplam		3		3

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
			bekar	haftalık çalışma süresi	full time	15	1	16
					part time	3	0	3
				Toplam		18	1	19
	25-34	okur yazar değil	evli	haftalık çalışma süresi	full time	2	1	3
				Toplam		2	1	3
			bekar	haftalık çalışma süresi	full time	1	2	3
				Toplam		1	2	3
		Okuryazar ancak okul bitirmemiş	evli	haftalık çalışma süresi	full time	7	1	8
					part time	0	2	2
				Toplam		7	3	10
			bekar	haftalık çalışma süresi	full time	8	1	9
				Toplam		8	1	9
		ilkokul	evli	haftalık çalışma süresi	full time	350	25	375
					part time	5	2	7
				Toplam		355	27	382
			bekar	haftalık çalışma süresi	full time	68	7	75
					part time	2	1	3
				Toplam		70	8	78

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
		ilkogretim-ortaokul	evli	haftalık çalışma süresi	full time	140	14	154
					part time	1	0	1
				Toplam		141	14	155
			bekar	haftalık çalışma süresi	full time	27	7	34
					part time	1	1	2
				Toplam		28	8	36
		lise	evli	haftalık çalışma süresi	full time	356	13	369
					part time	10	1	11
				Toplam		366	14	380
			bekar	haftalık çalışma süresi	full time	134	8	142
					part time	3	1	4
				Toplam		137	9	146
		iki yıllık fakülte	evli	haftalık çalışma süresi	full time	59	2	61
					part time	1	0	1
				Toplam		60	2	62
			bekar	haftalık çalışma süresi	full time	38	3	41
				Toplam		38	3	41

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
		dört yıllık fakulte	evli	haftalık çalışma süresi	full time	120		120
					part time	33		33
				Toplam		153		153
			bekar	haftalık çalışma süresi	full time	81	2	83
					part time	17	1	18
				Toplam		98	3	101
		Yüksek lisans -doktora	evli	haftalık çalışma süresi	full time	16	1	17
					part time	4	0	4
				Toplam		20	1	21
			bekar	haftalık çalışma süresi	full time	11		11
				Toplam		11		11
	35-44	okur yazar değil	evli	haftalık çalışma süresi	full time	5	1	6
					part time	0	2	2
				Toplam		5	3	8
			bekar	haftalık çalışma süresi	full time		2	2
				Toplam			2	2
		Okuryazar ancak okul bitirmemis	evli	haftalık çalışma süresi	full time	13	3	16
					part time	0	2	2
				Toplam		13	5	18

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
		ilkokul	evli	haftalık çalışma süresi	full time	524	37	561
					part time	13	10	23
				Toplam		537	47	584
			bekar	haftalık çalışma süresi	full time	11	3	14
					part time	1	1	2
				Toplam		12	4	16
		ilkogretim-ortaokul	evli	haftalık çalışma süresi	full time	170	11	181
					part time	1	1	2
				Toplam		171	12	183
			bekar	haftalık çalışma süresi	full time	4		4
				Toplam		4		4
		lise	evli	haftalık çalışma süresi	full time	311	8	319
					part time	5	0	5
				Toplam		316	8	324
			bekar	haftalık çalışma süresi	full time	10	1	11
				Toplam		10	1	11
		iki yıllık fakülte	evli	haftalık çalışma süresi	full time	88	1	89
					part time	7	0	7

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
				Toplam		95	1	96
			bekar	haftalık çalışma süresi	full time	3		3
				Toplam		3		3
		dört yıllık fakulte	evli	haftalık çalışma süresi	full time	120	1	121
					part time	27	0	27
				Toplam		147	1	148
			bekar	haftalık çalışma süresi	full time	10		10
				Toplam		10		10
		Yüksek lisans-doktora	evli	haftalık çalışma süresi	full time	12		12
					part time	1		1
				Toplam		13		13
			bekar	haftalık çalışma süresi	full time	1		1
					part time	2		2
				Toplam		3		3
	45-54	okur yazar değil	evli	haftalık çalışma süresi	full time	12	1	13
					part time	0	1	1
				Toplam		12	2	14
		Okuryazar ancak okul bitirmemis	evli	haftalık çalışma süresi	full time	9	2	11
				Toplam		9	2	11
		ilkokul	evli	haftalık çalışma süresi	full time	233	23	256

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
					part time	13	8	21
				Toplam		246	31	277
			bekar	haftalık çalışma süresi	full time	5		5
				Toplam		5		5
		ilkogretim-ortaokul	evli	haftalık çalışma süresi	full time	83	6	89
					part time	3	2	5
				Toplam		86	8	94
			bekar	haftalık çalışma süresi	full time	4		4
				Toplam		4		4
		lise	evli	haftalık çalışma süresi	full time	164	7	171
					part time	2	2	4
				Toplam		166	9	175
			bekar	haftalık çalışma süresi	full time	1	1	2
				Toplam		1	1	2
		iki yıllık fakülte	evli	haftalık çalışma süresi	full time	41		41
					part time	7		7
				Toplam		48		48
			bekar	haftalık çalışma süresi	full time	2		2

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
				Toplam		2		2
		dört yıllık fakulte	evli	haftalık çalışma süresi	full time	81	2	83
					part time	18	1	19
				Toplam		99	3	102
			bekar	haftalık çalışma süresi	full time	3		3
					part time	1		1
				Toplam		4		4
		Yüksek lisans-doktora	evli	haftalık çalışma süresi	full time	11	1	12
					part time	2	0	2
				Toplam		13	1	14
	55-64	Okuryazar ancak okul bitirmemis	evli	haftalık çalışma süresi	full time	2	2	4
					part time	0	1	1
				Toplam		2	3	5
		ilkokul	evli	haftalık çalışma süresi	full time	46	23	69
					part time	1	4	5
				Toplam		47	27	74
			bekar	haftalık çalışma süresi	full time	1		1
				Toplam		1		1
		ilkogretim-ortaokul	evli	haftalık çalışma süresi	full time	16	3	19

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
					part time	1	2	3
				Toplam		17	5	22
			bekar	haftalık çalışma süresi	full time	2	1	3
				Toplam		2	1	3
		lise	evli	haftalık çalışma süresi	full time	23	3	26
					part time	0	1	1
				Toplam		23	4	27
			bekar	haftalık çalışma süresi	full time	1		1
				Toplam		1		1
		iki yıllık fakülte	evli	haftalık çalışma süresi	full time	7	0	7
					part time	1	1	2
				Toplam		8	1	9
		dört yıllık fakulte	evli	haftalık çalışma süresi	full time	18	0	18
					part time	7	2	9
				Toplam		25	2	27
			bekar	haftalık çalışma süresi	full time	1		1
				Toplam		1		1
		Yüksek lisans-doktora	evli	haftalık çalışma süresi	full time	4		4

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
				Toplam		4		4
	65	okur yazar değil	evli	haftalık çalışma süresi	full time	1	0	1
					part time	0	1	1
				Toplam		1	1	2
		Okuryazar ancak okul bitirmemis	evli	haftalık çalışma süresi	full time	1	1	2
				Toplam		1	1	2
		ilkokul	evli	haftalık çalışma süresi	full time	3	5	8
					part time	1	2	3
				Toplam		4	7	11
			bekar	haftalık çalışma süresi	full time		1	1
				Toplam			1	1
		ilkogretim-ortaokul	bekar	haftalık çalışma süresi	part time		1	1
				Toplam			1	1
		lise	evli	haftalık çalışma süresi	full time	2	1	3
				Toplam		2	1	3
		Yüksek lisans-doktora	evli	haftalık çalışma süresi	full time	1		1
					part time	1		1
				Toplam		2		2
kadin	15-24	Okuryazar ancak okul bitirmemis	evli	haftalık çalışma süresi	full time	3	2	5
				Toplam		3	2	5

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
			bekar	haftalık çalışma süresi	full time	9	11	20
					part time	0	2	2
				Toplam		9	13	22
		ilkokul	evli	haftalık çalışma süresi	full time	4	2	6
				Toplam		4	2	6
			bekar	haftalık çalışma süresi	full time	8	1	9
					part time	1	2	3
				Toplam		9	3	12
		ilkogretim-ortaokul	evli	haftalık çalışma süresi	full time	5	6	11
				Toplam		5	6	11
			bekar	haftalık çalışma süresi	full time	39	52	91
					part time	0	9	9
				Toplam		39	61	100
		lise	evli	haftalık çalışma süresi	full time	16	6	22
					part time	1	1	2
				Toplam		17	7	24
			bekar	haftalık çalışma süresi	full time	78	49	127
					part time	2	7	9
				Toplam		80	56	136

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
		iki yıllık fakülte	evli	haftalık çalışma süresi	full time	5		5
				Toplam		5		5
			bekar	haftalık çalışma süresi	full time	33	5	38
					part time	3	2	5
				Toplam		36	7	43
		dört yıllık fakulte	evli	haftalık çalışma süresi	part time	2		2
				Toplam		2		2
			bekar	haftalık çalışma süresi	full time	23	3	26
					part time	6	0	6
				Toplam		29	3	32
		Yüksek lisans-doktora	evli	haftalık çalışma süresi	full time	1		1
				Toplam		1		1
			bekar	haftalık çalışma süresi	full time	1		1
				Toplam		1		1
	25-34	okur yazar değil	evli	haftalık çalışma süresi	full time		2	2
					part time		1	1
				Toplam			3	3
		Okuryazar ancak okul bitirmemis	evli	haftalık çalışma süresi	full time	1		1
				Toplam		1		1

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
			bekar	haftalık çalışma süresi	full time	1	1	2
					part time	0	1	1
				Toplam		1	2	3
		ilkokul	evli	haftalık çalışma süresi	full time	43	19	62
					part time	1	12	13
				Toplam		44	31	75
			bekar	haftalık çalışma süresi	full time	21	9	30
					part time	0	3	3
				Toplam		21	12	33
		ilkogretim-ortaokul	evli	haftalık çalışma süresi	full time	16	5	21
				Toplam		16	5	21
			bekar	haftalık çalışma süresi	full time	14	3	17
					part time	1	0	1
				Toplam		15	3	18
		lise	evli	haftalık çalışma süresi	full time	76	8	84
					part time	0	3	3
				Toplam		76	11	87
			bekar	haftalık çalışma süresi	full time	53	7	60
					part time	1	0	1

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
				Toplam		54	7	61
		iki yıllık fakülte	evli	haftalık çalışma süresi	full time	37	2	39
					part time	3	0	3
				Toplam		40	2	42
			bekar	haftalık çalışma süresi	full time	27	3	30
					part time	1	1	2
				Toplam		28	4	32
		dört yıllık fakulte	evli	haftalık çalışma süresi	full time	52	1	53
					part time	40	2	42
				Toplam		92	3	95
			bekar	haftalık çalışma süresi	full time	54	2	56
					part time	13	1	14
				Toplam		67	3	70
		Yüksek lisans-doktora	evli	haftalık çalışma süresi	full time	14	0	14
					part time	4	1	5
				Toplam		18	1	19
			bekar	haftalık çalışma süresi	full time	7		7
				Toplam		7		7
	35-44	okur yazar değil	evli	haftalık çalışma süresi	full time	6	1	7

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
					part time	0	2	2
				Toplam		6	3	9
		Okuryazar ancak okul bitirmemis	evli	haftalık çalışma süresi	full time	4	0	4
					part time	2	3	5
				Toplam		6	3	9
			bekar	haftalık çalışma süresi	full time	1		1
				Toplam		1		1
		ilkokul	evli	haftalık çalışma süresi	full time	47	23	70
					part time	3	13	16
				Toplam		50	36	86
			bekar	haftalık çalışma süresi	full time	15	6	21
					part time	0	3	3
				Toplam		15	9	24
		ilkogretim-ortaokul	evli	haftalık çalışma süresi	full time	11	2	13
					part time	1	2	3
				Toplam		12	4	16
			bekar	haftalık çalışma süresi	full time	7		7
				Toplam		7		7
		lise	evli	haftalık çalışma süresi	full time	42	4	46
					part time	6	3	9
				Toplam		48	7	55

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
			bekar	haftalık çalışma süresi	full time	24	2	26
					part time	1	0	1
				Toplam		25	2	27
		iki yıllık fakülte	evli	haftalık çalışma süresi	full time	26		26
					part time	4		4
				Toplam		30		30
			bekar	haftalık çalışma süresi	full time	16		16
				Toplam		16		16
		dört yıllık fakulte	evli	haftalık çalışma süresi	full time	35		35
					part time	22		22
				Toplam		57		57
			bekar	haftalık çalışma süresi	full time	19		19
					part time	3		3
				Toplam		22		22
		Yüksek lisans-doktora	evli	haftalık çalışma süresi	full time	6		6
					part time	1		1
				Toplam		7		7
			bekar	haftalık çalışma süresi	full time	3		3
				Toplam		3		3

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
	45-54	okur yazar değil	evli	haftalık çalışma süresi	full time	3	1	4
					part time	0	2	2
				Toplam		3	3	6
		Okuryazar ancak okul bitirmemis	evli	haftalık çalışma süresi	full time	2	3	5
					part time	0	1	1
				Toplam		2	4	6
			bekar	haftalık çalışma süresi	full time	2		2
				Toplam		2		2
		ilkokul	evli	haftalık çalışma süresi	full time	19	6	25
					part time	0	2	2
				Toplam		19	8	27
			bekar	haftalık çalışma süresi	full time	6	4	10
				Toplam		6	4	10
		ilkogretim-ortaokul	evli	haftalık çalışma süresi	full time	11	2	13
					part time	0	1	1
				Toplam		11	3	14
			bekar	haftalık çalışma süresi	full time	3	1	4

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum		asgari ücretten fazla	asgari ücretten az	Toplam	
					part time	1	0	1
				Toplam		4	1	5
		lise	evli	haftalık çalışma süresi	full time	17	1	18
					part time	1	0	1
				Toplam		18	1	19
			bekar	haftalık çalışma süresi	full time	5		5
				Toplam		5		5
		iki yıllık fakülte	evli	haftalık çalışma süresi	full time	9		9
					part time	4		4
				Toplam		13		13
			bekar	haftalık çalışma süresi	full time	1		1
				Toplam		1		1
		dört yıllık fakulte	evli	haftalık çalışma süresi	full time	12		12
					part time	10		10
				Toplam		22		22
			bekar	haftalık çalışma süresi	full time	3		3
					part time	3		3
				Toplam		6		6
		Yüksek lisans-doktora	evli	haftalık çalışma süresi	full time	1		1

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum		asgari ücretten fazla	asgari ücretten az	Toplam	
					part time	2		2
				Toplam		3		3
	55-64	okur yazar değil	evli	haftalık çalışma süresi	full time	1		1
				Toplam		1		1
			bekar	haftalık çalışma süresi	full time	1	1	2
					part time	0	1	1
				Toplam		1	2	3
		Okuryazar ancak okul bitirmemis	evli	haftalık çalışma süresi	full time	1		1
				Toplam		1		1
		ilkokul	evli	haftalık çalışma süresi	full time	1		1
				Toplam		1		1
			bekar	haftalık çalışma süresi	part time		1	1
				Toplam			1	1
		ilkogretim-ortaokul	evli	haftalık çalışma süresi	full time	1		1
				Toplam		1		1
		lise	evli	haftalık çalışma süresi	full time	2	0	2
					part time	0	1	1
				Toplam		2	1	3
		iki yıllık fakülte	evli	haftalık çalışma süresi	full time	1		1

Çizelge 1.1. (Devam) Haftalık çalışma süresi * aylık ücret * medeni durum * eğitim * yaş * cinsiyet olumsallık tablosu

cinsiyet	yaş	eğitim	medeni durum			asgari ücretten fazla	asgari ücretten az	Toplam
				Toplam		1		1
			bekar	haftalık çalışma süresi	full time	1		1
					part time	1		1
				Toplam		2		2

EK-2. Asgari ücrete göre seçilen en son lojistik regresyon modelinin parametre tahminleri

Çizelge 2.1. Hosmer - Lemeshow testi

Aşama	χ^2	s.d	Sig.
1	18,997	8	,015
2	17,508	8	,025
3	16,984	8	,030
4	18,406	8	,018
5	17,857	8	,022
6	16,276	8	,039
7	17,122	8	,029
8	16,990	8	,030
9	17,242	8	,028
10	14,549	8	,069
11	18,342	8	,019

Çizelge 2.2. Asgari ücrete göre seçilen en son lojistik regresyon modelinin parametre tahminleri

		B	Std. Hata	Wald	sd	Sig.
10.adım	Sabit	,898	3,151	,081	1	,776
	C(1)	-4,500	,682	43,483	1	,000
	Y			71,017	5	,000
	Y(1)	-1,723	,870	3,926	1	,048
	Y(2)	-2,420	,724	11,176	1	,001
	Y(3)	-2,444	,622	15,439	1	,000
	Y(4)	-2,268	,552	16,852	1	,000
	Y(5)	-,923	,517	3,189	1	,074
	E			80,381	7	,000
	E(1)	8,732	1,878	21,629	1	,000
	E(2)	7,748	1,606	23,287	1	,000
	E(3)	6,161	1,363	20,443	1	,000
	E(4)	5,461	1,156	22,315	1	,000
	E(5)	3,512	,960	13,390	1	,000
	E(6)	1,770	,809	4,784	1	,029
	E(7)	,354	,704	,252	1	,615
	MD(1)	-1,980	,801	6,112	1	,013
	MD*C	-1,175	,499	5,541	1	,019
	C*CS	-,031	,007	20,867	1	,000
	Y*E	,341	,106	10,458	1	,001
	E*MD	,557	,144	15,071	1	,000

Çizelge 2.2. (Devam) Asgari ücrete göre seçilen en son lojistik regresyon modelinin parametre tahminleri

		B	Std. Hata	Wald	sd	Sig.
	E*MD*C	-,197	,070	7,808	1	,005
	C*E*CS	,007	,002	18,660	1	,000
	Y*E*MD	-,259	,080	10,537	1	,001
	E*Y*CS	,001	,001	,951	1	,330
	Y*E*MD*C	,115	,050	5,366	1	,021
	C*E*Y*CS	-,002	,001	6,035	1	,014

EK-3. Asgari ücrete göre seçilen en son lojit modelin parametre tahminleri

Çizelge 3.1. Asgari ücrete göre seçilen en son lojit modelinin parametre tahminleri

Aylık ücret geliri		B	Std. Hata	Wald	sd	P değ.	Exp(B)	Exp(B) için 95% güven aralığı	
								Üst sınır	Alt sınır
asgari ucrette n az	Sabit	-14,82	1,237	143,447	1	,000			
	C	6,204	1,013	37,478	1	,000	494,94	67,902	3607,63
	Y	3,707	,497	55,581	1	,000	40,727	15,369	107,925
	MD	7,636	,856	79,526	1	,000	2070,4	386,58	11088,5
	Y*MD	-2,567	,378	46,209	1	,000	,077	,037	,161
	E*MD	-,267	,070	14,714	1	,000	,766	,668	,878
	MD*CS	1,155	,299	14,958	1	,000	3,174	1,768	5,700
	C.Y	-2,069	,360	32,963	1	,000	,126	,062	,256
	C*MD	-3,978	,715	30,951	1	,000	,019	,005	,076
	C*CS	2,039	,414	24,279	1	,000	7,679	3,413	17,277
	C*Y*MD	1,394	,258	29,323	1	,000	4,033	2,434	6,680
	C*E*MD	,266	,067	15,910	1	,000	1,304	1,145	1,486
	C*E*CS	-,377	,064	34,510	1	,000	,686	,605	,778
	C*MD*CS	-,426	,312	1,868	1	,172	,653	,354	1,203
	C*Y*E*MD	-,079	,024	10,409	1	,001	,924	,881	,970
	C*Y*E*CS	,041	,021	3,899	1	,048	1,041	1,000	1,084

EK-4. Delta metodu teorisi

Örneğin ele alınan parametreyi β ve maksimum olabilirlik tahmin edicisini $\hat{\beta}$ ile gösterelim. Burada olabilirlik fonksiyonu $\ell(\beta)$ ve log-olabilirlik fonksiyonu $L(\beta) = \log \ell(\beta)$ olsun. Çoğu modelde $L(\beta)$ konkav bir şekle sahiptir ve $\hat{\beta}$ türevin sıfıra eşit olduğu noktada oluşur. Burada SH , $\hat{\beta}$ nın standart hatasını ; $\text{cov}(\hat{\beta})$, $\hat{\beta}$'nın asimptotik kovaryans matrisini göstersin. Normal koşullar altında, $\text{cov}(\hat{\beta})$ bilgi matrisinin tersidir. Standart hatalar da bilgi matrisinin tersinin köşegen elemanlarının karekökleridir.

Fisher bilgi matrisinin (jk) faktörü aşağıdaki gibidir [11].

$$-E\left(\frac{\partial^2 L(\beta)}{\partial\beta_j \partial\beta_k}\right) \qquad (5.1)$$

Log-lineer modellerde ise iki ya da daha fazla değişken arasındaki birlikteliğin büyüklüğünün(λ) tahminleri bilgi matrisinin tersinden elde edilir. Delta(Δ) metoduna dayalı standart hatalar sadece bu durumda mümkündür.

İstatistikte delta metodu, sınırlı varyans bilgisine sahip asimptotik normal dağılan istatistiksel tahmin edicinin bir fonksiyonu için yaklaşık bir olabilirlik dağılımı türetme yöntemidir. Genel olarak delta metodu merkezi limit teoremi olarak dikkate alınabilir [15].

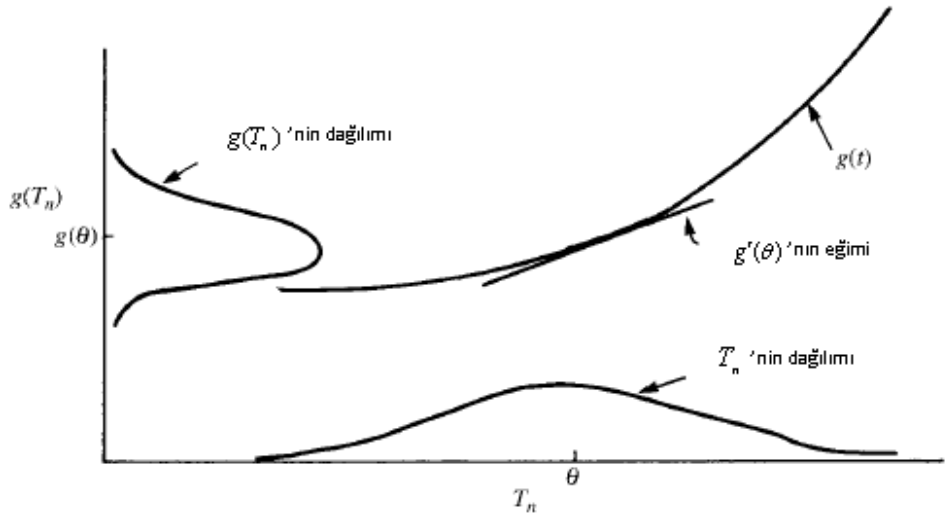
Delta metodu, büyük çaplı örnek çıkarımları için standart hataları elde etmede basit ve kullanışlı bir metodtur. T_n, θ parametreli asimptotik normal dağılımı olan bir istatistik olsun. n örnek çapını belirtsin. Varsayalım ki bir tahmin edici, T_n'nin $g(T_n)$ fonksiyonudur. O zaman $g(T_n)$ geniş örnek çaplı normal dağılıma sahip olur. Standart hatalar, θ'nın yanında t için $g(t)$nin ne kadar hızlı değiştiğine bağlıdır. Özellikle, büyük n için, T_n'nin, ortalaması θ ve standart hatası $\sigma/\sqrt{n}$ olan normal

dağılıma sahip olduğu varsayılsın. $n \to \infty$'a gittikçe, $\sqrt{n}(T_n - \theta)$'nın kümülatif dağılım fonksiyonu, 0 ortalamalı σ^2 varyanslı normal dağılan rasgele değişkenin kümülatif dağılım fonksiyonuna yakınsar. Bu sınırlandırılmış tutum dağılımdaki yakınsamanın bir örneğidir ve $\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \sigma^2)$ olarak gösterilir.

g, θ da en az iki kere türevi alınabilir bir fonksiyon olsun. $t = \theta$' nın bir komşuluğunda $g(t)$ için Taylor serileri kullanılarak $\sqrt{n}[(g(T_n) - g(\theta)] \approx \sqrt{n}(T_n - \theta)g'(\theta)$ 'yı büyük örnek çapları için gösterebiliriz. Burada $t = \theta$'da $g'(\theta) = \partial g / \partial t$ dir. Eğer değişken $Y \sim N(0, \sigma^2)$ ise o zaman $cY \sim N(0, \sigma^2)$ olur. Böylece

$$\sqrt{n}[(g(T_n) - g(\theta)] \xrightarrow{d} N(0, [g'(\theta)]^2 \sigma^2) \text{ olur.} \qquad (5.2)$$

Diğer bir ifadeyle $g(T_n)$, $g(\theta)$ ve $[g'(\theta)]^2 \sigma^2 / n$ varyans ile normal dağılır. Şekil 5.1 de bu sonuç özetlenmiştir [11].



Şekil 5.1 Delta metodunun resmedilmesi [11]

$g(t)$, $g'(\theta)$ eğimiyle θ etrafında yaklaşık doğrusaldır. O zaman $g(T_n)$ yaklaşık normal dağılır diyebiliriz çünkü normal dağılıma sahip rasgele değişkenlerin dönüşümleri de normal dağılır. Sonuç olarak Eş. 5.1 delta metodu olarak adlandırılır.

$g'(\theta)$ ve $\sigma^2 = \sigma^2(\theta)$, bilinmeyen θ parametresine bağlı olduğundan asimptotik varyans bilinmez. θ için T_n'nin yerine güven aralıkları kullanılır ve $\sqrt{n}[(g(T_n) - g(\theta)]/\left|g'(T_n)\right|\sigma(T_n)$ asimptotik standart normal dağılır. Örneğin, büyük örnekler için $g(\theta)$'ın Wald %95 güven aralığı aşağıdaki gibidir [11].

$$(g(T_n) \pm 1.96\left|g'(T_n)\right|\sigma(T_n)/\sqrt{n} \tag{5.3}$$

Delta metodu maksimum olabilirlik tahminin fonksiyonları için de güven aralıklarını hesaplamada kullanılan genel bir yaklaşımdır. Normal koşullar altında, eğer $\hat{\beta}$ maksimum olabilirlik tahmin vektörü ise o zaman, $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N[0, \text{var}(\hat{\beta})]$ dir [2].

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : OKTAY, G.Merve

Uyruğu : T.C.

Doğum tarihi ve yeri : 27.06.1986 Trabzon

Medeni hali : Bekar

Telefon : 0 (312) 481 94 00

e-mail : merve.oktay@tuik.gov.tr

Eğitim

Derece	Eğitim Birimi	Mezuniyet tarihi
Lisans	Gazi Üniversitesi /İstatistik Bölümü	2008
Lisans	Gazi Üniversitesi/ Ekonometri Bölümü	2009
Lise	Cumhuriyet Anadolu Lisesi	2004

İş Deneyimi

Yıl	Yer	Görev
2009-halen	Türkiye İstatistik Kurumu	Uzman Yrd.

Yabancı Dil

İngilizce

Hobiler

Sinema, tiyatro, plates