

**KSÜ
FEN BİLİMLERİ ENSTİTÜSÜ**

YÜKSEK LİSANS TEZİ

**LOJİSTİK REGRESYON
VE
BİYOLOJİK ALANLARDA KULLANIMI**

85372

Mustafa ŞAHİN

**T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ**

ZOOTEKNİ ANABİLİM DALI

**KAHRAMANMARAŞ
EKİM 1999**

85372

**KSÜ
FEN BİLİMLERİ ENSTİTÜSÜ**

**LOJİSTİK REGRESYON
VE
BİYOLOJİK ALANLARDA KULLANIMI**

**Mustafa ŞAHİN
YÜKSEK LİSANS TEZİ
ZOOTEKNİ ANABİLİM DALI**

Bu tez ~~14.10.1999~~ ^{14.10.1998} Tarihinde Aşağıdaki Jüri Üyeleri Tarafından
Oy Birliği/~~Oy Çekliği~~ ile Kabul Edilmiştir.

İmza:.....

Doç. Dr. Ercan EFE
DANIŞMAN

İmza:.....

Prof. Dr. Yüksel BEK
ÜYE

İmza:.....

Prof. Dr. Zeynel CEBECİ
ÜYE

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Kod No:

.....

Prof. Dr. Nuri GÜZEL
Enstitü Müdürü
İmza ve Mühür

Bu Çalışma KSÜ Fen Bilimleri Enstitüsü Tarafından Desteklenmiştir.
Proje No: 97/7-11.

Not: Bu tezde kullanılan özgün ve başka kaynaktan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat eserleri Kanunundaki hükümlere tabidir.

İÇİNDEKİLER

SAYFA

ÖZET.....	I
ABSTRACT.....	II
TEŞEKKÜR	III
ÇİZELGELER DİZİNİ	IV
ŞEKİLLER DİZİNİ	VI
1. GİRİŞ	1
2. ÖNCEKİ ÇALIŞMALAR	3
2.1. Yöntem Çalışmaları	3
2.2. Tıp Alanındaki Uygulamaları	4
2.3. Bitkisel ve Hayvansal Üretim Alanlarındaki Uygulamaları	5
3. MATERYAL VE YÖNTEM	8
3.1. MATERYAL	8
3.2. YÖNTEM	9
3.2.1. Doğrusal Olasılık Modeli	9
3.2.2. Lojistik Model ve Lojistik fonksiyon	11
3.2.3. Basit Lojistik Regresyon Modeli	12
3.2.4. Lojistik Regresyon Katsayılarının önem Testi	17
3.2.5. Desen Değişkenleri Tanımlanması	21
3.2.6. Çoklu Lojistik Regresyon Modeli	22
3.2.7. Modelin ve Katsayıların Önem Testi	24
3.2.8. Basit Lojistik Regresyonda katsayıların Yorumlanması	25
3.2.9. Çoklu Lojistik Regresyonda katsayıların Yorumlanması	36
3.2.10. Model Yapılandırma	38
3.2.11. Uyum İyiliğinin Belirlenmesi	41
4. BULGULAR VE TARTIŞMA	45
5. SONUÇLAR VE ÖNERİLER	60
KAYNAKLAR	61
ÖZGEÇMİŞ	66
EKLER	67

ÖZET

YÜKSEK LİSANS TEZİ

LOJİSTİK REGRESYON VE BİYOLOJİK ALANLARDA KULLANIMI

Mustafa ŞAHİN

KSÜ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOOTEKNİ ANABİLİM DALI

DANIŞMAN: Doç. Dr. Ercan EFE

Yıl:1999, Sayfa:88

Jüri : Doç. Dr. Ercan EFE
: Prof. Dr. Yüksel BEK
: Prof. Dr. Zeynel CEBECİ

Biyolojik konularda çok yaygın olarak kullanılan lojistik regresyon; *log-linear model*, *ayrım* (discriminant), *kümeleme* (cluster) analizi gibi kesikli verilerle ilgili bir yöntemdir. Lojistik regresyonda, bir veya daha fazla bağımsız (açıklayıcı, explanatory, causal) değişken ile adlandırma ölçeğinde kesikli bir bağımlı (cevap, response) değişken arasındaki ilişki incelenmektedir.

Lojistik fonksiyon $P(\mathbf{x}_i) = e^{\sum \beta_k x_{ik}} / (1 + e^{\sum \beta_k x_{ik}}) = e^{\ell(\mathbf{x}_i)} / (1 + e^{\ell(\mathbf{x}_i)})$ şeklindedir. $\text{Log}[P(\mathbf{x}_i)/(1 - P(\mathbf{x}_i))]$ *lojit dönüşümü* ile *lojistik regresyon* modeli,

$$\ell(\mathbf{x}_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} = \beta_0 + \sum_{k=1}^p \beta_k x_{ik}$$

olarak elde edilir. Bu eşitlik *lojit model* ya da kısaca *lojit* olarak da adlandırılır.

Bu çalışmada, lojistik model katsayılarının en çok olabilirlik tahminleri, önem testleri, desen değişkenlerinin tanımlanması, olabilirlik oran testleri, model yeterliliği ve modelin uyum iyiliğinin belirlenmesi incelenmiştir. Analiz yöntemi sayısal bir örnek üzerinde gösterilmiş, ayrıca çocukların doğum ağırlığını etkileyen risk faktörleri ile ilgili gerçek verilere lojistik regresyon modeli uydurulmuştur.

Anahtar Kelimeler: Lojistik, lojit, odds, regresyon.

ABSTRACT

MSc THESIS

LOGISTIC REGRESSION AND ITS USING IN BIOLOGICAL FIELDS

Mustafa ŞAHİN

DEPARTMENT OF ANIMAL SCIENCE
INSTITUTE OF NATURAL AND APPLIED SCIENCE
UNIVERSITY OF KAHRAMANMARAS SUTCU IMAM

Supervisor: Assoc. Prof. Dr. Ercan EFE

Year:1999, Pages:88

Jury : Assoc. Prof. Dr. Ercan EFE
: Prof. Dr. Yüksel BEK
: Prof. Dr. Zeynel CEBECİ

Logistic regression widely used in biological fields is a method related to discrete type of data such as *log-linear model*, *discriminant* and *cluster analysis*. The associations between one or more explanatory variables and a nominal scaled discrete response variable has been investigated by using logistic regression.

Logistic function is $P(\mathbf{x}_i) = e^{\sum \beta_k x_{ik}} / (1 + e^{\sum \beta_k x_{ik}}) = e^{\ell(\mathbf{x}_i)} / (1 + e^{\ell(\mathbf{x}_i)})$. By using *logit transformation*, $\log[P(\mathbf{x}_i)/(1 - P(\mathbf{x}_i))]$, *logistic regression model* is obtained as below:

$$\ell(\mathbf{x}_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} = \beta_0 + \sum_{k=1}^p \beta_k x_{ik}$$

This equation is also called *logit model* or shortly *logit*.

In this study, maximum likelihood estimates of slope coefficients of logistic model, their significance tests, determining the desing variables, likelihood ratio test, adequacy of the model and assesing the goodness of fit were investigated. The method was shown on a numerical example and logistic model was fitted to the real birth weight data and its risk factors.

Keywords: Logistic, logit, odds, regression.

TEŞEKKÜR

Çalışmalarım sırasında beni yönlendiren, her türlü bilgi ve tecrübesini benden esirgemeyen sayın hocam, Doç. Dr. Ercan EFE'ye, tezin hazırlanması sırasında yardımlarını esirgemeyen bölüm araştırma görevlilerinden O. Soylu YILMAZ, Uğur SERBESTER, Ayhan CEYHAN ve Hayri YILMAZ'a en derin şükranlarımı sunarım.

Aynı zamanda bugünlere gelmemde büyük emeği olan aileme ve manevi desteğinden dolayı değerli eşim Belkıs ŞAHİN'e teşekkürü bir borç bilirim.



ÇİZELGELER DİZİNİ**SAYFA**

Çizelge 1.1.	Ayrım, Kümeleme ve Lojistik Regresyon Karşılaştırması.....	2
Çizelge 2.1.	Lojistik Modelin Kullanıldığı Çalışmaların Konulara Göre Dağılımı.....	7
Çizelge 3.1.	Düşük Doğum Ağırlığını ve Aday Risk Faktörlerinin Kodlanması.....	8
Çizelge 3.2.	Düşük Doğum ağırlığı ve Aday Risk Faktörleri İçin Gerçek Veriler.....	9
Çizelge 3.3.	X'in Sınıflandırılması, y=0 ve y=1 Frekansları ve y=1 İçin Nispi Frekansların Bulunmasına Bir Örnek.....	13
Çizelge 3.4.	r=2 İçin Desen Değişkenlerinin Kısmi ve Marjinal Yöntemle Tanımlanması.....	21
Çizelge 3.5.	r>2 İçin Desen Değişkenlerinin Kısmi ve Marjinal Yöntemle Tanımlanması.....	28
Çizelge 3.6.	İki Sınıflı Bağımsız Değişkene Sahip Basit Lojistik Regresyonda y=1 ve y=0 İçin Koşullu Olasılıklar.....	27
Çizelge 3.7.	İkişer Sınıflı X ve Y'nin Çapraz Sınıflamasına Bir Örnek.....	29
Çizelge 3.8.	Y ve İki'den Fazla Sınıflı X'in Çapraz Sınıflamasına Örnek.....	31
Çizelge 3.9.	X'in Kısmi Yöntemle Desen Değişkenine Dönüştürülmesi.....	32
Çizelge 3.10.	X'in Ortalamadan Sapma Yöntemi İle Desen Değişkenine Dönüştürülmesi.....	33
Çizelge 3.11.	Sabit Denek Sayılı Onlu Risk Grupları İçin Gözlenen ve Beklenen Frekanslar.....	43
Çizelge 3.12.	Sabit P(xi) Aralıklı Onlu Risk Grupları İçin Gözlenen ve Beklenen Frekanslar.....	44
Çizelge 4.1.	Sürekli Değişkenlerin Ortalama ve Standart Sapma Değerleri ($\bar{x} \pm S$).....	45
Çizelge 4.2.	Deneklerin Sigara İçme Durumuna Göre Dağılımları.....	45
Çizelge 4.3.	Deneklerin Gebelik Durumuna Göre Dağılımları.....	46
Çizelge 4.4.	Deneklerin Daha Önce Doğum Yapıp Yapmama Durumuna Göre Dağılımları.....	46
Çizelge 4.5.	Deneklerin Akrabalık Durumuna Göre Dağılımları.....	46
Çizelge 4.6.	Deneklerin Hipertansiyona Göre Dağılımları.....	46
Çizelge 4.7.	Deneklerin Bebek Cinsiyetine Göre Dağılımları.....	47
Çizelge 4.8.	Düşük Doğum Ağırlığı İle İlişkili Olabileceği Düşünülen Değişkenlerin Basit Lojistik Regresyon Analizi	47
Çizelge 4.9.	GS değişkeni için desen değişkenlerinin oluşturulması ve kodlanması	48
Çizelge 4.10.	Çoklu Lojistik Regresyon Analizi.....	48
Çizelge 4.11.	Doğum Durumu (D) Değişkeni İndirgendikten Sonra Geriye Kalan Değişkenler İçin Çoklu Lojistik Regresyon Analizi.....	49

Çizelge 4.12. YAŞ'ın Kısmi Yöntemle Desen Değişkenine Dönüştürülmesi.....	49
Çizelge 4.13. Kartil Bölünmüş ve Desen Değişkenine Dönüştürülmüş Yaş Değişkeni İçin Çoklu Lojistik Regresyon Analizi.....	50
Çizelge 4.14. YAS Değişkeni İçin , Katsayı Tahminleri, Standart Hata, Ki-Kare ve Odds Oranları	50
Çizelge 4.15. SADA Değişkeninin Desen Değişkenine Dönüştürülmesi.....	51
Çizelge 4.16. Kartil Bölünmüş ve Desen Değişkenine Dönüştürülmüş SADA Değişkeni İçin Çoklu Lojistik Regresyon Analizi.....	51
Çizelge 4.17. SADA Değişkeni İçin, Katsayı Tahminleri, Standart Hata, Ki-Kare ve Odds Oranları	52
Çizelge 4.18. SADA Değişkeninin Kesikli Değişken Olarak Kodlanması.....	52
Çizelge 4.19. DSADA Değişkeni Kullanılarak Elde Edilen Katsayı Tahminleri , Standart Hata, Ki-Kare ve Önemlilikleri.....	52
Çizelge 4.20. Ana Etkiler Modeline İnteraksiyon Terimlerinin Tek Tek Eklenmesiyle Elde Edilen Yeni Modellerin -2 Log Olabilirlik, Olabilirlik Oran Testi (G) ve P- Değerleri.....	53
Çizelge 4.21. Önemli İnteraksiyon Terimlerini İçeren Model İçin Çoklu Lojistik Regresyon Analizi.....	54
Çizelge 4.22. SIG ve DSADA'nın Dört Farklı Kombinasyonu İçin Lojit ve Lojit Farkları	56
Çizelge 4.23. SIG ve DSADA İçin Lojit Farklar ve Odds Oranı Tahminleri ($\hat{d}_1$, $\hat{d}_2$ ve $\hat{d}_3$; 35 Yaşlı Denekler İçin Hesaplanmıştır.).....	57
Çizelge 4.24. Sabit Denek Sayılı Onlu Risk Grupları İçin Gözlenen ve Beklenen Frekanslar.....	58
Çizelge 4.25. Sabit P(xi) Aralıklı Onlu Risk Grupları İçin Gözlenen ve Beklenen Frekanslar.....	59

ŞEKİLLER DİZİNİ

SAYFA

Şekil 3.1.	Lojistik Regresyonda x ve y Noktalarının genel durumu.....	13
Şekil 3.2.	Yaşlara Göre Yüksek Verimlilerin Oransal Miktarları.....	14
Şekil 3.3.	Tek Bağımsız Değişkenli Basit Lojistik Regresyon Modeli Analizi.....	17
Şekil 3.4.	Yalnızca Sabit Terimi İçeren Basit Lojistik Regresyon Modeli Analizi....	20



1. GİRİŞ

Kesikli veriler genellikle sayma yolu ile elde edilen veya gruplandırılmış olan dolayısı ile tamsayı olarak ifade edilen verilerdir. Kesikli veriler birkaç formda ortaya çıkar. Örneğin sürekli veriler, “yüksek”, “orta” , “düşük” gibi birkaç kesikli sınıfa ayrıldığında gruplanmış veriler oluşur. Fakat birçok değişken doğal olarak kesiklidir.

Sayısal bir sıraya sokulması mümkün sınıflardan oluşan kesikli bir değişken “sıralı (ordinal)” olarak, sıra numarasından ziyade isimlerle sınıflanan kesikli bir değişken “*adlandırma* (nominal)” ölçekli olarak tanımlanır. Kesikli bir değişken karşılıklı ayrık ve birbirini engelleyen gruplardan oluşabilir. Böyle bir değişken yalnızca iki değer alabiliyorsa “*ikil* (binary veya dichotomous)” olarak adlandırılır. Örneğin bir ürünün tüketicileri cinsiyet (erkek ve bayan), eğitim durumu (okumamış, ilk, orta, ön lisans, lisans, lisans üstü ya da kısaca A, B, C, D1, D2, E), medeni durumu (bekar, evli, dul, ayrı) bakımından incelenmiş olabilir. Bu çalışmada cinsiyet ikil değişken, eğitim durumu sıralı ölçekli bir değişken, ve medeni durum da adlandırma ölçekli değişkendir. Medeni durum değişkenine ait sınıfların belli bir diziliş sırası yoktur. Ancak eğitim durumunu en düşük eğitimden en yüksek eğitime doğru sıralamak mümkündür.

Araştırmacılar bazen değişik sınıfların her bir kombinasyonunda yer alan gözlem sayıları ile de ilgilenir. İkişer sınıflı iki değişken için 2x2’lik iki yönlü çapraz (contingency) çizelgeler en tanınmışlarıdır. İkidenden fazla değişken için de çok yönlü tablolar vardır. Bu tip çok değişkenli kesikli verilere *kategorik veriler* (categorical, frequency, count, attribute data) denir (Doğan, 1985; Chatfield ve Collins, 1992).

En basit olan iki yönlü çapraz tablo iki satır ve iki sütundan oluşur. İki yönlü tabloların genel şekli rxc tablolarıdır. Bu tablolarda r tane sıra ve c tane sütun vardır. Bu tabloların iyi bilinen analizi χ^2 ’dir. Sıra ve sütunlar arasındaki bağımsızlığa dair sıfır hipotezi test edilir. Analizi tamamlayıcı bilgiler olarak iki değişken arasında bazı ilişki katsayıları ile çeşitli oranlar hesaplanabilir (Yurtsever, 1984; Efe, 1995; Yoğurtçugil, 1978).

İkidenden fazla kategorik değişken olduğunda *çok boyutlu çapraz çizelgeler* (multidimensional contingency table) ve bunların analizleri söz konusudur (Doğan, 1985; Everitt, 1977). Bu gibi durumlarda araştırmacı, değişkenlerin karşılıklı bağımsızlığı şeklinde tek bir sıfır hipotezi yerine örneğin belli bir değişkenin diğer tüm değişkenler ile bağımsızlığı gibi birkaç hipotezle ilgilenebilir.

Burada önemli bir husus, genelleştirilmiş lineer modelde *cevap değişkeninin* (bağımlı değişken, response variable) beklenen değerinin ele alınan etkiler bakımından *eklemeli* (additive, toplanabilir) olmasıdır. Ancak, çapraz tablolarda farklardan ziyade frekansların oranları söz konusudur. Bu nedenle etkiler çarpımsaldır (multiplicative).

Çarpımsal etkileri eklemeli hale çevirmek için bir dizi logaritmik dönüştürme (transformasyon) gerekir. Kategorik verilerle ilgili olan bu tip modellere “*log-lineer modeller*” denir. Log-lineer model, çok boyutlu çapraz çizelgelerin tanımlanması ve ilgilenilen etkilerin tahmin edilmesi için araştırmacıya sistematik bir yaklaşım olanağı

sağlar. Bu modellerde, beklenen hücre frekansının logaritmasının sıra sütun ve interaksiyon terimlerinin lineer bir fonksiyonu olduğu varsayılır.

Kesikli verilerle ilgili modellerden birisi de “*lojistik model*” veya “*lojit model*” lerdir. Lojistik ya da lojit modeller, bir veya daha fazla kesikli ya da sürekli tipte *açıklayıcı* (explanatory, causal) *değişkenlere* bağlı olarak tek bir *ikil cevap* (binary response) değişkeninin varyasyonunu incelemek amacıyla kullanılan modellerdir. Örneğin, biyolojik assay (dose-response) çalışmalarında, bir ilacın farklı dozlarda uygulanması halinde ölen zararlıların oranı ölçülmüş olabilir. Burada esas olarak “*gerçekleşme*” *olasılığı*, P , ile ilgilenilmektedir. Burada P olasılığı $[0, 1]$ aralığında değer alır ve doğrudan doğruya lineer bir modelle tanımlanamaz. P olasılığının lojistik dönüşümü $\log[P/(1-P)]$ şeklindedir ve $[-\infty, +\infty]$ aralığında değer alır. Bu dönüşüm P ’nin “*lojit*”i olarak adlandırılır. Lojistik model, bir hücre olasılığının lojitinin, model parametrelerinin lineer bir fonksiyonu olduğunu varsayar. Böylece, model, sıradan bir regresyon modeline benzer. İki sütunlu ve sabit sıra toplamı bir çapraz tablo için lojit transformasyon log-lineer model ile eşdeğer olur.

Verilerin iki ya da daha çok gruba (küme) ayrılmasında kullanılan yöntemlerden üç tanesi ayrım (diskriminant), kümeleme (cluster) ve lojistik regresyon analizidir. Bu üç analiz yöntemin birbirine göre karşılaştırması Çizelge 1.1’deki gibidir.

Çizelge 1.1. Ayrım, Kümeleme ve Lojistik Regresyon Karşılaştırması

Ayrım	-Küme sayısı önceden bilinmekte ve analiz boyunca değişmemektedir. -Gelecekte kullanılabilir fonksiyonlar verir. Bu özelliği ile kümelemeden ayrılır; ana bileşenler, kanonik korelasyon ve çok değişkenli regresyon analizine benzer. -Değişkenlerin bazıları kesikli, bazıları sürekli olamaz. -Sık kullanılan, en çok bilinen yöntemdir. -Veriler normal dağılımlı olmalıdır.
Kümeleme	-Küme sayısı önceden tam bilinmemektedir. -Gelecekte kullanılabilir özelliği yoktur ve az kullanılan bir yöntemdir. -Veriler veya uzaklık değerleri normal dağılımlı olmalıdır. -Kovaryans matrisine ilişkin varsayım yoktur. -Çok değişkenli varyans analizi, lojistik regresyon analizi, çok boyutlu ölçekleme gibi çok değişkenli analizlerle yakından ilgilidir.
Lojistik Regresyon	-Değişkenlerden bazıları sürekli, bazıları kesikli olabilir. -Küme sayısı önceden bilinmektedir. -Az kullanılan ancak son yıllarda yaygınlaşan bir yöntemdir. -Varsayım (normallik, ortak varyans) bozulmalarında diskriminant ve çapraz tablolara alternatif oluşturur. Çünkü normallik varsayımı yoktur (y, ikil veya çok sınıflı, kesikli olduğundan). -Model esnek ve kolay yorumlanabilir.

2. ÖNCEKİ ÇALIŞMALAR

Bağımlı değişkenin binom dağılımına uyduğu durumlarda, lojistik regresyon modelinin kullanımı özellikle biyolojik alanlarda yaygın hale gelmiştir. Biyolojik çalışmalarda, çoğunlukla bağımlı değişkenin *ikil* (binary) olması, veri setlerinde çeşitli varsayım bozulmalarının görülmesi (normal dağılıma ve ortak kovaryansa sahip olmama) durumunda lojistik regresyon analizinin uygulanabilmesi ve lojistik regresyonla elde edilen katsayıların kolay yorumlanabilir olması araştırmacıların lojistik regresyon modellerini tercih etmesine neden olmuştur.

2.1. Yöntem Çalışmaları

Lojistik regresyon modellerinin yaygın bir şekilde kullanılır hale gelmesi, katsayı tahmin yöntemlerinin geliştirilmesi ve lojistik regresyon modellerinin daha ayrıntılı incelenmesine neden olmuştur. Cornfield (1962), lojistik regresyondaki katsayı tahmin işlemlerinde diskriminant fonksiyonu yaklaşımını ilk kez kullanarak popüler hale getirmiştir.

Le (1984) basit dönüşümlü (cross-over) deneme planları için lineer lojistik modeller üzerinde çalışmıştır.

Bonney (1987) lojistik regresyon modelinin kullanımı ve geliştirilmesi üzerine çalışmıştır. Aynı zamanda bağımlı değişkenleri tanımlayacak daha güvenilir modellerin geliştirilmesi işlemlerin; 2x2'lik tablolarda ve örnekler üzerinde göstermiştir.

Roberts ve ark. (1987) lojistik regresyonda standart χ^2 , olabilirlik oranları (G^2), "pseudo" en çok olabilirlik tahminleri, uyum mükemmelliği ve hipotez testleri üzerine çalışmalar yapmışlardır.

Houck (1988) simetrik olmayan verilere lojistik modelin uydurulmasında yetersizlik olduğunu ve bağımsız değişkenlere ait ölçeğin değiştirilmesi ile lojistik modelin simetrik olmayan verilere uyabileceğini göstermişlerdir.

Ruiz-Velasco (1989) açıklayıcı değişkenlerin normal dağılıma sahip olduğu varsayımı altında, parametreler hakkındaki hipotezleri test etmek amacıyla lineer diskriminant analizinde lojistik regresyonun asimtotik etkisini hesaplamıştır.

Ali ve Khan (1989) tek dereceli istatistiklerin fonksiyonel momentleri ve log lojistik dağılımdan ikinci derece istatistiklerin bölünmüş momentlerini elde etmişlerdir.

Hosmer ve ark. (1989) lineer olmayan modellerde en iyi alt setin seçimi yöntemleri üzerine çalışmışlar ve oldukça zaman alan bir modelleme işlemi olduğu için paket programların kullanımını önermişlerdir.

Duffy (1990) lojistik regresyonda hata terimlerinin dağılışı ve parametre değerlerinin gerçek değerlere yaklaşımını incelemiştir.

Corroll ve Wand (1990) lojistik regresyon parametrelerinin yarı parametrik tahminleri üzerine çalışmışlar ve parametre tahminlerinin bulunmasında Kernel regresyon tekniğini kullanmışlardır.

Vach ve Schumacher (1991) sınıflanabilir bağımsız değişken içeren gözlemlerde lojistik regresyon kullanımını incelemişler ve üç farklı yöntemi karşılaştırmışlardır. Bunlar sırası ile en çok olabilirlik yöntemi, “pseudo” en çok olabilirlik yöntemi ve “probability imputation” yöntemidir.

Hsu ve Leonard (1995) lojistik regresyon fonksiyonlarında Bayes tahminlerinin elde edilmesi işlemleri üzerinde çalışmışlar ve lojistik regresyonda Monte Carlo dönüşümünün kullanılabileceğini göstermişlerdir.

Heise ve Myers (1996) tek değişkenli lojistik regresyon için optimal deneme planları üzerine çalışmışlar ve iki tip deneme planı sunmuşlardır. Bunlar, Q-optimal ve D-optimal deneme planlarıdır. Q-optimal planının D-verimliliği ve D-optimal planının Q-verimliliği üzerinde durulmuştur.

Akkaya ve Pazarlıoğlu (1998) lojistik regresyon modellerinin ekonomi alanında kullanımını örneklerle incelemişlerdir.

2.2. Tıp Alanındaki Uygulamaları

Tıp alanındaki çalışmalar özellikle çeşitli kanser türleri, şeker hastalığı (anemi), kalp hastalıkları ve kontrol altına alınabilir yetersiz beslenmeye dayalı hastalıklar üzerinde yoğunlaşmaktadır.

Fears ve Brown (1986) *durum-kontrol* (case-control) çalışmalarında kullanılabilecek deneme desenlerini sayısal örneklerle göstermişlerdir. Risk faktörlerinin kontrol altında tutulabilmesi için tabakalı deneme desenlerini kullanmışlar ve her bir tabaka için en çok olabilirlik yöntemini kullanarak ayrı ayrı lojistik regresyon tahminlerini elde etmişlerdir.

Breslow ve Cain (1988) durum-kontrol çalışmaları için lojistik regresyon modelleri üzerine çalışmışlardır. Örnekler hasta olan veya hastalık riski taşıyan popülasyonlardan seçilmiş ve hasta olmayan gruplar ise kontrol grubu olarak ele alınmıştır.

Margaret ve Liang (1988) kümelenmiş ikil veriler için şartlı lojistik regresyon modelleri üzerine çalışmışlar ve çalışmalarında parametre tahminleri için klinik veriler üzerinde en çok olabilirlik yöntemini kullanmışlardır.

Scoot ve Wild (1990) durum-kontrol çalışmalarında lojistik regresyonun uydurulması ve en çok olabilirlik yöntemi ile parametre tahminleri üzerine çalışmışlardır.

Başarır (1990) klinik verilerde çok değişkenli lojistik regresyon analizi ve ayırimsama sorunu üzerine çalışmıştır.

Drescher ve Schill (1991) eşlenik olmayan durum-kontrol verilerine şartsız lojistik regresyon modeli uydurarak, incelenen faktör için risk tahminlerinin yapılması üzerinde çalışmışlardır.

Ganguly ve Nimbalkar (1992) *eşlenik durum-kontrol* (match case-control) çalışmalarında ikiden fazla düzeyli lojistik regresyon modellerinin kullanımı üzerine çalışmışlar ve özellikle 2x2’lik tabloların düzenlenmesi ve odds oranlarının tahmini üzerinde durmuşlardır.

Gartside ve Glueck (1995) insanlarda beslenme şekli, sigara ve alkol kullanımı, fiziksel aktivite gibi risk faktörlerinin kalp hastalığı üzerindeki etkilerini incelemişlerdir.

O'neil ve Barry (1995) trafik kazalarında ölen ve yaşayanlara ait bir veri setinde, şartlı lojistik regresyon modeli ve kesikli lojistik regresyon modelinin karşılaştırmasını yapmışlardır. İki yöntem arasındaki en önemli fark, kesikli lojistik regresyonda grup seviye kovaryanslarının bilinmesi ve grup seviye etkilerinin tahmin edilmesi gerektiğidir.

Muscati ve ark. (1996) sigaranın yaşlanma üzerine etkisini incelemişlerdir.

Kloiber ve ark.(1996) Peoples ve ark. (1991), Buescher ve ark. (1993) kadınlarda düşük doğum ağırlığını etkileyen risk faktörlerini; Santos ve ark. (1998) kafein tüketimi ve düşük doğum ağırlığı arasındaki ilişkiyi; Sable ve Herman (1997) erken doğum ve düşük doğum ağırlığı arasındaki ilişkiyi; Stevens ve Mcanarney (1992) genç yaşta gebe kalan bayanlarda ağırlık artışına etki eden faktörleri incelemişlerdir.

Gibbons ve Hedeker (1996) klinik çalışmalarda kullanılan ikil değişkenlerin lojistik regresyon analizindeki şansa bağlı etkilerinin güvenilirliği ve üç düzeyli verilerde lojistik regresyon modelleri üzerine çalışmalar yapmışlardır.

La-Vecchia ve ark. (1997) göğüs kanserinde risk faktörlerini (durum-kontrol); Meyer ve ark. (1997) prostat kanserinde risk faktörlerini (durum-kontrol); Stefani ve ark. (1998) mide kanserinde risk faktörlerini (sigara ve alkol değişkenlerine ait odds oranlarını kullanarak); Lee ve ark. (1998) diyet ve akciğer kanseri arasındaki ilişkiyi incelemişlerdir.

Elhan (1997) Lojistik regresyonu incelemiş ve kronik arter hastalığına etki eden risk faktörleri üzerinde örneklendirmiştir.

Singh ve ark. (1998) kırsal bölge ve şehirlerdeki insanlarda şeker hastalığı ve damar hastalıkları arasındaki ilişkiyi lojistik regresyonu kullanarak incelemişlerdir.

Cox ve ark. (1998) kardiovasküler hastalıklar ve hipertansiyon arasındaki ilişkiyi; Perzanowski ve ark.(1998) çocuklarda astım ve alerji arasındaki ilişkiyi incelemişlerdir.

2.3. Bitkisel ve Hayvansal Üretim Alanlarındaki Uygulamaları

Lojistik regresyon klinik çalışmalarda olduğu kadar bitki ve hayvanlar üzerindeki çalışmalarda da yaygın bir şekilde kullanılmıştır.

Bazzaz ve Harper (1977) keten bitkisi; Constable ve Glaeson (1977) , Marani (1979) pamuk bitkisi; Dale ve ark.(1980) ve Major (1980) mısır bitkisi üzerinde lojistik regresyon uygulamalarını incelemişlerdir.

Grossman ve ark. (1985) tavuklarda lojistik büyüme eğrilerini incelemişler ve parametre tahmin tekniklerini kırmızı Rhode Island ve beyaz Leghorn kültür ırkları üzerinde karşılaştırmışlardır.

Grossman ve Anderson (1985) tavuklarda lojistik büyüme eğrileri ve parametre (kalıtım parametreleri) tahminlerini kırmızı Rhode Island ve beyaz Leghorn kültür ırkları üzerinde incelemişlerdir.

Koops (1986) insanlarda kullanılan boy büyüme eğrisinin n tane lojistik büyüme fonksiyonunun toplamı olduğunu ve pika, fare ve tavşanların ağırlık artışlarına uydurulan çok devreli büyüme eğrisinin, otokorelasyon olmaması durumunda tek devreli modelden daha iyi olduğunu göstermişlerdir. Aynı zamanda

matematiksel büyüme fonksiyonunun her bir büyüme devresinde lojistik büyümeye bağlı olduğu ve her bir devredeki günlük canlı ağırlık kazancı eğrisinin simetrik olması nedeni ile lojistik regresyonun kullanılabileceğini göstermiştir.

Boer ve ark.(1989) İsrail Holstein'larında çok devreli fonksiyonların uydurulması üzerine çalışmışlar ve laktasyon eğrilerinin şeklini tanımlamak için kullanılan matematiksel fonksiyonların yetersiz kaldığını savunmuşlardır.

Alho (1990) balıkçılıkta lojistik regresyonun uygulanabilirliği üzerine çalışmıştır. Bu amaçla tutulan balıklara ait ölçümler alınmış ve numara verildikten sonra tekrar bırakılmıştır. Daha sonra balıklar tekrar yakalanmış ve yeni ölçümler alınmıştır. Olasılıklar, şartlı en çok olabilirlik kullanılarak gözlenen verilerden tahmin edilmiştir. Denek sayısı ne kadar çok olursa o kadar tutarlı sonuçlar elde edileceği bildirilmiştir.

Lees ve ark. (1990) genç yaşlardaki koçlarda döl verimi ve risk faktörleri; Schumann ve ark. (1990) etçi sığır ırklarının buzağılarında görülen ishal ve risk faktörleri; Emanuelson ve Bendixen (1991) sığırlarda ovaryum kisti oluşumu ve risk faktörleri; Berger ve ark. (1992) sığırlarda genç yaşlarda görülen ölümler ve risk faktörleri; Pinard ve ark. (1993) tavuklarda marek hastalığı ve risk faktörleri; Lust ve ark.(1993) Labrador cinsi av köpeklerinde kalça çıkıklığı ve risk faktörleri üzerinde çalışmışlardır.

Oetzel (1991) süt sığırlarında süt verim düzeyini etkileyen risk faktörleri üzerinde çalışmıştır.

Mohammed ve ark.(1991) süt sığırlarında kan selenyum düzeyi ile yumurtalık kisti oluşumu arasındaki ilişkiyi incelemişlerdir.

Morris ve Silk (1992) mısır bitkisinde kök büyümesinin hücresel gelişimini inceleyerek zamana bağlı lojistik regresyon modelini geliştirmişlerdir. Burada zaman (t), önceden tespit edilen düzeyde hücrenin gelişmesi için geçen süredir.

Huang ve Tu (1998) domates bitkisinin kök kısmında gelişen zararlı kurtçuklarla ilgili çalışmalar yapmışlardır.

Thompson ve ark. (1994) sığırlarda gebelikte yıpranma ve rektal palpasyon üzerinde çalışmışlardır.

Humblot ve ark.(1996) süt sığırlarında gebelik ve ovulasyon döngüsü arasındaki ilişkiyi incelemişlerdir.

Shrestha ve Ghimire (1996) göçebe koyun ve keçi yetiştirme sisteminde sodyum etkisini incelemişlerdir.

Judge ve ark. (1997), Enevoldsen ve Sorensen (1992), Houben ve ark. (1993), Berning ve Shook (1992), Nielen ve ark. (1995), Morin ve ark. (1998), süt sığırlarında mastitise neden olan risk faktörlerinin tespiti üzerinde çalışmışlardır.

Dietz ve ark. (1997) süt sığırlarında süt üretimini etkileyen genler üzerinde çalışmışlardır.

Domecq ve ark. (1997) süt sığırcılığında, süt üretimi ve fiziksel görünüme dayalı puanlama (body condition score) arasındaki ilişkiyi incelemişlerdir.

Bu çalışmada ulusal ve uluslar arası olmak üzere 50'den fazla dergide son 10 yıllık çalışmalar taranmış, lojistik modelin çeşitli uygulamalarının kullanıldığı 3000 civarında çalışma tespit edilmiştir. Bunlardan incelenebilen 801 tanesinin konulara göre dağılımı Çizelge 2.1'de verilmiştir.

Çizelge 2.1. Lojistik Modelin Kullanıldığı Çalışmaların Konulara Göre Dağılımı

ÇALIŞMA ALANLARI		ÇALIŞMA SAYISI		TOPLAM		
TIP	Genel	214	%26.7	262 %32.7		
	Halk Sağlığı	17	%2.1			
	Besleme-Diyet (İnsanlar)	31	%3.9			
TARIM (Hayvansal ve Bitkisel) ve VETERİNERLİK	Entomoloji	37	%4.6	73	%9.1	381 %47.6
	Fitopatoloji	36	%4.5	225	%28.1	
	Sütçü Sığırlar	112	%14			
	Etçi Sığırlar	8	%1			
	Mastitis	17	%2.1			
	Domuzlar	15	%1.9			
	Atlar	10	%1.2			
	Koyunlar	10	%1.2			
	Tavuklarda	8	%1			
	Maymunlar	1	%0.1			
	Hindiler	1	%0.1			
	Arıcılık	3	%0.4			
	Geyikler	1	%0.1			
	Keçiler	3	%0.4			
	Kediler	5	%0.6			
	Köpekler	18	%2.2			
	Fareler	3	%0.4			
	Sindirim-Besleme	5	%0.6			
	Su Ürünleri	5	%0.6			
	Tarla Bitkileri	61	%7.6	76	%9.5	
	Bahçe Bitkileri	4	%0.4			
	Bitki Besleme	11	%1.4			
		Genetik	7	%0.9	7	
ÇEVRE-DOĞA-EKOLOJİ		51	%6.4	109 %13.5		
ORMANCILIK		42	%5.2			
GIDA		8	%0.1			
MİKROBİYOLOJİ		8	%0.1			
DİĞERLERİ	Eğitim-Sosyal-Ekonomi	39	%4.9	49 %6.2		
	Turizm	3	%0.4			
	Jeoloji-Uzaktan Algılama	7	%0.9			
TOPLAM		801 %100				

Çizelgeden de görüldüğü gibi, biyolojik alanlarda lojistik modelin kullanımı oldukça yaygındır. İncelenen çalışmaların %32.7'si Tıp, %47.6'sı Tarım ve Veterinerlik, %13.5'i Çevre, Doğa, Ekoloji, Ormancılık, Gıda ve Mikrobiyoloji alanındadır. Diğer alanlardaki lojistik model kullanımı ise %6.2 olmuştur.

3. MATERİYAL VE YÖNTEM

3.1. MATERİYAL

Lojistik regresyonun teorik olarak anlatılmasından sonra, çocuklarda doğum ağırlığı ve buna etki eden risk faktörleri ile ilgili gerçek verilere lojistik regresyon uydurulmuştur. Bu veriler Kahramanmaraş'ta özel bir klinikten alınmıştır. Bu veri setinde, bağımlı değişken olarak düşük doğum ağırlığını (DA), bağımsız değişken olarak da yaş (YAS), son adet dönemi vücut ağırlığı (SADA), annenin sigara içime durumu (SIG), annenin daha önce doğum yapıp yapmadığı (D), annenin toplam gebelik sayısı (GS), annenin eşi ile akrabalık durumu (AK), bebeğin ultrasonda tespit edilen cinsiyeti (C), hemogloblin düzeyi (HB) ve hipertansiyon (HT) değişkenleri kullanılmıştır.

Doğum ağırlığı ile aday risk faktörlerinin kısaltmaları ve kodlama şekilleri Çizelge 3.1'de, azaltılmış veriler ise Çizelge 3.2'de verilmiştir.

Çizelge 3.1. Düşük Doğum Ağırlığını ve Aday Risk Faktörlerinin Kodlanması

Özellik	Değişken Adı	Sembolik Adı
Doğum Ağırlığı (Doğum Ağırlığı ≥ 2500 gr=0, Doğum Ağırlığı < 2500 gr=1)	DA	y_i
Annenin Yaşı (yıl)	YAS	x_{ik}
Annenin Son Adet Dönemi Vücut Ağırlığı (kg)	SADA	
Annenin Sigara İçme Durumu (İçiyor=1, İçmiyor=0)	SIG	
Annenin Daha Önce Doğum Yapıp Yapmadığı (Yaptı=1, Yapmadı=0)	D	
Annenin Gebelik Sayısı (1 Gebelik=1, 2 Gebelik=2, 3 Gebelik=3, 4 Gebelik=4)	GS	
Annenin Eşi İle Akrabalık Durumu (Var=1, Yok=0)	AK	
Bebeğin Ultrasonda Tespit Edilen Cinsiyeti (Erkek=0, Kız=1)	C	
Annenin Hipertansiyon Durumu (Var=1, Yok=0)	HT	
Annenin Hemogloblin Düzeyi (g/dl)	HB	

Çizelge 3.2. Düşük Doğum ağırlığı ve Aday Risk Faktörleri İçin Gerçek Veriler

No	DA	SİG	YAŞ	GS	D	AK	SADA	C	HT	HB
1	0	1	14	1	0	0	50.0	0	0	13.1
2	0	0	15	1	0	1	58.0	0	0	10.2
3	0	0	21	1	0	1	61.0	0	0	11.0
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
282	1	0	20	1	0	0	66.0	1	0	10.8
283	1	0	39	1	0	0	54.0	0	0	10.9
284	1	0	40	1	0	0	53.0	0	0	13.5

3.2. YÖNTEM

Lojistik regresyon analizinin kullanım amacı diğer model oluşturma teknikleri ile aynıdır. Yani mümkün olan en az sayıda değişkeni kullanarak, sonuç değişkeni ile bağımsız değişken arasındaki ilişkiyi doğru bir şekilde tanımlayabilen, iyi bir uyuma sahip ve biyolojik olarak anlamlı bir model oluşturmaktır.

Lojistik regresyon analizi normal dağılıma sahip olmama, ortak kovaryansa sahip olmama gibi çeşitli varsayım bozulmaları durumunda *ayrimsama analizi* ve *çapraz çizelgelere* (contingency table) bir alternatif olmaktadır. Diğer yandan, bağımlı değişkenin ikil (0, 1) veya daha fazla (0, 1, 2, ..., c) düzey içeren kesikli bir değişken olması durumunda normallik varsayımının bozulması nedeni ile *lineer regresyon* analizine de alternatif olmaktadır.

Bu çalışmada yalnızca *ikil bağımlı değişken* durumu incelenecektir.

Lojistik regresyon modelinin kullanım amacı ve tercih sebepleri, veri setlerinin lojistik regresyona uygun hale getirilmesi (ikil değişken olarak kodlanması) ve ikil (binary) değişkenlerin gerekliliği izah edilecektir. Lojistik regresyon analizi teorik olarak anlatılırken konunun anlaşılabilirliği açısından eşzamanlı olarak sayısal örnek sonuçları ile desteklenecektir.

3.2.1. Doğrusal Olasılık Modeli

Öncelikle ikil bağımlı y için “*doğrusal olasılık modeli*” kavramını inceleyerek başlamak “*lojistik model*” kavramının anlaşılmasına yardımcı olacaktır.

Genelleştirilmiş lineer bir modelde p tane bağımsız değişken (x_{ik}) için sürekli bir bağımlı değişkenin (y_i) beklenen değeri,

$$E(y_i | x_{i1}, x_{i2}, \dots, x_{ip}) = \sum_{k=0}^p \beta_k x_{ik} \quad (3.1)$$

şeklindedir ($i=1, 2, \dots, n$; n = denek sayısı; $x_{i0}=1$). Bu eşitlik “*genel doğrusal regresyon modeli*” olarak da bilinir. Herhangi bir y_i gözlemi, doğal olarak, belli bir ε_i hatası ile gerçekleşecektir. Buna göre y_i gözlemi için,

$$y_i = \sum_{k=0}^p \beta_k x_{ik} + \varepsilon_i \quad (3.2)$$

yazılabilir. Bu modelde bağımsız değişkenler sürekli veya kesikli olabilirler. Bunun sonucunda da y_i , $[-\infty, +\infty]$ aralığında bir değer alır. Diğer bir ifade ile y_i sürekli dir.

Ancak, gerçekte bir çok biyolojik ve sosyal olaylarda bağımlı değişken kesiklidir. Yani, y 'nin aldığı değerler birkaç sınıfa ayrılmaktadır. Eğer bağımlı değişken y 'nin aldığı değerler yalnızca iki sınıfa düşüyorsa *ikil* (iki sınıflı, iki gruplu, dichotomous, binary) olarak adlandırılır. İki den fazla sınıf söz konusu ise *çok sınıflı* (polychotomous) denir. Bu çalışmada yalnızca ikil bağımlı değişkenler ele alınacaktır.

İkil bağımlı değişken y 'nin aldığı değerler 0 ve 1 ile kodlandığı için beklenen değeri,

$$E(y_i) = 1 \cdot P(y_i = 1) + 0 \cdot P(y_i = 0) = P(y_i = 1)$$

olarak bulunur. Bu ikil bağımlı değişkenin 1 değerini alması olasılığı demektir ve $P(y_i = 1)$ 'in değişim aralığı $0 \leq P(y_i = 1) \leq 1$ şeklindedir. Buna göre ikil y 'nin 1 değerini alması olasılığı regresyon denklemi olarak,

$$E(y_i) = P(y_i = 1) = \sum_{k=0}^p \beta_k x_{ik} \quad (3.3)$$

şeklinde yazılabilir ($i=1, 2, \dots, n$; $x_{i0}=1$). Bu eşitlik “*doğrusal olasılık modeli*” (linear probability model) olarak da adlandırılmaktadır. İkil y_i 'nin gözlenen değeri ise yine belli bir ε_i hatası ile,

$$y_i = \sum_{k=0}^p \beta_k x_{ik} + \varepsilon_i$$

şeklinde gerçekleşecektir. İkil y_i yalnızca 0 ve 1 değerlerini aldığına göre her iki durum için hata, ε_i ,

$$y_i=0 \text{ ise } \varepsilon_i = - \sum_{k=0}^p \beta_k x_{ik}$$

ve

$$y_i=1 \text{ ise } \varepsilon_i = 1 - \sum_{k=0}^p \beta_k x_{ik}$$

olarak bulunur. Hata teriminin beklenen değeri ve varyansı sırasıyla,

$$E(\varepsilon_i) = 0$$

ve

$$Var(\varepsilon_i) = \sum_{k=0}^p \beta_k x_{ik} [1 - \sum_{k=0}^p \beta_k x_{ik}]$$

şeklindedir (kanıtlama Ek-1).

β_k katsayısının en küçük kareler tahmini yansız ancak en iyi değildir. Çünkü ε_i hatası için değişken varyanslılık söz konusudur. Bunu ortadan kaldırmak için y_i 'nin x üzerindeki regresyonu için $\hat{\beta}_k$ katsayıları bulunur ve ,

$$w_i = 1/Var(\varepsilon_i) = 1 / \left[\sum \hat{\beta}_k x_{ik} (1 - \sum \hat{\beta}_k x_{ik}) \right]$$

ağırlık değerleri ile ağırlıklandırılır, yani hata terimi, $w_i \varepsilon_i$, sabit varyanslı olacaktır (Tatlıdil, 1996).

3.2.2. Lojistik Model ve Lojistik Fonksiyon

İkil y_i 'nin doğrusal olasılık modelinin, $E(y_i) = P(y_i = 1 | \mathbf{x}_i) = \sum_{k=0}^p \beta_k x_{ik}$

olduğu ve 0 ile 1 aralığında olasılık değerleri aldığına yukarıda değinilmişti. Burada $\mathbf{x}_i$, $(x_{i1}, x_{i2}, \dots, x_{ip})$ şeklinde p tane x_{ik} bağımsız (açıklayıcı) değişkeni içeren satır vektörüdür. Lojistik regresyonda kesikli veya sürekli olabilen ve $[-\infty, \infty]$ aralığında değer alabilen bağımsız x_{ik} değişkenleri ile yalnızca 0 ve 1 değeri alabilen ikil y_i değişkeni ilişkilendirildiğinden bu eşitlik her zaman gerçekleşmemektedir. Çözüm olarak $P(y_i=1|\mathbf{x}_i)$ olasılık değerlerini $[-\infty, \infty]$ aralığına dönüştürmek gerekir. Bu amaçla,

- 1) lojit dönüşüm,
- 2) probit dönüşümü (probability unit)

yaygın kullanılır. İkil y için doğrusal olasılık modelinde öncelikle,

$$P(\mathbf{x}_i)/(1-P(\mathbf{x}_i))$$

dönüşümü ile değişim aralığı $[0, +\infty]$ aralığına, daha sonra da **doğal logaritması** alınarak $[-\infty, +\infty]$ aralığına dönüştürülür. Lojit dönüşümden sonra (3.3) nolu eşitlikle verilen doğrusal olasılık modeli,

$$\ell(\mathbf{x}_i) = E(y_i | \mathbf{x}_i) = \log[P(\mathbf{x}_i)/(1 - P(\mathbf{x}_i))] = \sum_{k=0}^p \beta_k x_{ik} \quad (3.4)$$

şekline dönüşür ($i=1, 2, \dots, n; k=1, 2, \dots, p; x_{i0}=1$). Bu eşitlik “*lojistik model*” veya “*lojit model*” ya da kısaca “*lojit*” olarak adlandırılır. Burada $P(\mathbf{x}_i)$ olasılığı,

$$P(\mathbf{x}_i) = \frac{e^{\sum \beta_k x_{ik}}}{1 + e^{\sum \beta_k x_{ik}}} = \frac{e^{\ell(\mathbf{x}_i)}}{1 + e^{\ell(\mathbf{x}_i)}} \quad (3.5)$$

şeklindedir. Bu eşitliğe “*lojistik fonksiyon*” denir. Bazen $P(\mathbf{x}_i)$ kısaca P_i olarak da gösterilebilmektedir. İkili y değişkeni yalnızca 0, 1 değerini aldığından hata terimi için,

$$\varepsilon_i \sim \text{binom}, E(\varepsilon_i) = 0$$

ve

$$\text{Var}(\varepsilon_i) = P(\mathbf{x}_i)[1 - P(\mathbf{x}_i)]$$

yazılabilir.

3.2.3. Basit Lojistik Regresyon Modeli

Basit lojistik regresyon modeli, yalnızca bir tane ($P=1$) bağımsız değişkenin bulunduğu durumdur. Buna göre (3.4) ve (3.5) nolu eşitliklerden basit lojistik model,

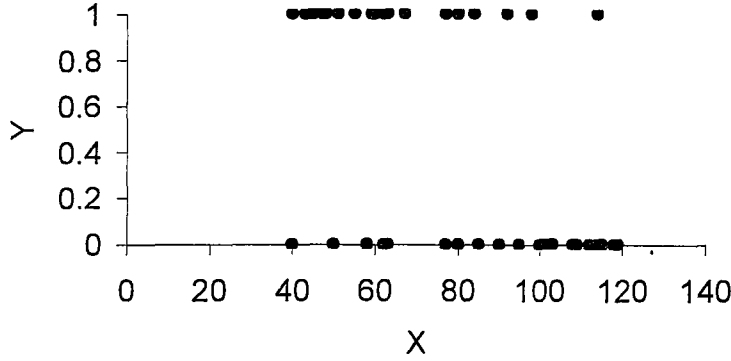
$$\ell(x_i) = \log[P(x_i)/(1 - P(x_i))] = \beta_0 + \beta_1 x_{i1} \quad (3.6)$$

ve basit lojistik fonksiyon ,

$$P(x_i) = \frac{e^{\ell(x_i)}}{1 + e^{\ell(x_i)}} = \frac{e^{\sum \beta_k x_{ik}}}{1 + e^{\sum \beta_k x_{ik}}} = \frac{e^{\beta_0 + \beta_1 x_{i1}}}{1 + e^{\beta_0 + \beta_1 x_{i1}}} \quad (3.7)$$

şeklinde yazılabilir.

Şekil 3.1’de herhangi bir x ’e göre y düzeylerinin dağılımı verilmiştir. Şekilden, noktaların 0 ve 1 düzeyine ait iki paralel çizgi üzerinde dağıldığı görülmektedir. $y=0$ düzeyindeki noktaların x arttıkça yoğunlaştığı, $y=1$ düzeyindeki noktaların ise x arttıkça seyreltiği görülmektedir.



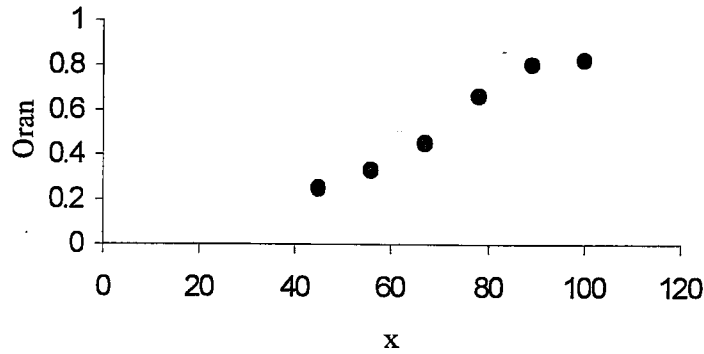
Şekil 3.1. Lojistik Regresyonda x ve y Noktalarının Genel Durumu.

Çizelge 3.3’deki gibi gruplandırılmış bir frekans tablosunun oluşturulduğu varsayalım. Çizelgenin en sonundaki sütunda yer alan oranlar, bir x grubundaki $y=1$ ’in oransal miktarıdır. Örneğin 1-inci gruptaki $y=1$ ’in oransal miktarı $3/12=0.25$ ’tir.

Çizelge 3.3. X ’in Sınıflandırılması, $y=0$ ve $y=1$ Frekansları ve $y=1$ İçin Nispi Frekansların Bulunmasına Bir Örnek

X Grupları	X_i	n	$y=1$	$y=0$	$y=1$ için oranlar
40 – 50	45	12	9	3	0.25
51 – 61	56	9	6	3	0.33
62 – 72	67	11	6	5	0.45
73 – 83	78	12	4	8	0.66
84 – 94	89	10	2	8	0.80
95 – 105	100	11	2	9	0.82
Σ		$n=65$	$n_1=29$	$n_0=36$	0.55

Bu çizelgeden, x ile y arasındaki ilişkinin yapısı daha açık olarak görülmektedir. Dikkat edileceği gibi x arttıkça $y=0$ ’a ait oranlarda artmaktadır. Bu durum Şekil 3.2’de görsel hale getirilmiştir.



Şekil 3.2. Yaşlara Göre Yüksek Verimlilerin Oransal Miktarları.

Şekil 3.2’de x ile y arasındaki ilişki daha iyi anlaşılabilir. Buna göre x ile y arasında doğrusal olmayan, “S” şekilli bir ilişki görülmektedir.

İkil değişkenlerle çalışıldığında şartlı ortalama $0 \leq E(y|x) \leq 1$ aralığında değer alır. Bu durum Şekil 3.2’de de görülmektedir. x deki bir birim değişme $E(y|x)$ deki değişimi gittikçe azalan miktarda 0’a veya 1’e yaklaştırır. “S” şekilli dağılım, bir rastgele değişkenin birikimli dağılışının grafiğine benzer. Çok iyi bilinen birikimli dağılışlar ile y ’nin ikil olduğu durumlar için $E(y|x)$ ’in dağılışının birbirine benzer olduğu dikkat çekicidir.

Şekil 3.2’deki dağılım için kullanılacak uygun modellerden biri lojistik dağılım modelidir. Lojistik dağılışın seçiminde başlıca iki neden vardır:

1. Matematiksel olarak esnek ve kolay kullanılan bir fonksiyondur,
2. Biyolojik olarak anlamlı yorumlara izin verir.

Parametrelerin Tahmini

Basit lojistik regresyon modelinde yalnızca bir tane bağımsız değişken (x_{i1}) olduğundan yalnızca β_0 ve β_1 katsayılarının tahmini gerekmektedir.

(x_i, y_i) gibi n tane bağımsız gözlem çiftini içeren bir örnek var olsun. Burada y_i ikil sonuç değişkeninin ve x_i de i -inci denek için incelenen bağımsız değişkenin değeridir. Bağımsız değişken için 0 ve 1 kodlarının sırasıyla belli bir özelliğin varlığı ve yokluğunu temsil ettiği varsayalım. (3.6) eşitliğindeki basit lojistik regresyon modelini bir veri grubuna uydurmak için, β_0 ve β_1 bilinmeyen parametrelerinin tahmin edilmesi gerekir.

Lineer regresyonda bilinmeyen parametreleri tahmin etmek için en çok kullanılan yöntem en küçük kareler yöntemidir. Bu yöntemde modelden tahmin edilen ile gözlenen değerler arasındaki sapmaların (hataların) kareleri toplamını minimize eden β_0 ve β_1 değerleri seçilir. Lineer regresyon için bilinen varsayımlar altında en küçük kareler yöntemi istenen istatistiki özelliklere sahip tahmin edicileri

verir. Ancak, lojistik regresyon analizinde en küçük kareler yöntemi uygulandığı zaman tahmin ediciler aynı özelliklere sahip olmazlar. Lojistik modelin parametrelerinin tahmini için kullanılan yöntemler,

- 1) En çok olabilirlik yöntemi,
- 2) Yeniden ağırlıklandırılmış iteratif en küçük kareler yöntemi ve
- 3) Minimum lojit ki-kare yöntemidir.

Bu çalışmada yalnızca en çok olabilirlik yöntemi üzerinde durulacaktır. Genel anlamda en çok olabilirlik yöntemi, gözlenmiş bir veri grubunun elde edilmesi olasılığını maksimum yapan parametre değerlerini verir. Bu yöntemin uygulanması için öncelikle en çok olabilirlik fonksiyonunun oluşturulması gerekmektedir. Bu fonksiyon, bilinmeyen parametrelerin bir fonksiyonu olarak gözlenen verinin olasılığını verir. Bu parametrelerin en çok olabilirlik tahmin edicileri, fonksiyonu maksimum yapan değerleri bulacak şekilde seçilir. Bu nedenle sonuçta elde edilen tahmin ediciler, gözlenen verilerle çok yakın değerlere sahiptir. Bu değerlerin lojistik regresyon modelinden nasıl bulunacağı aşağıda tanımlanmıştır.

Eğer y , 0 ve 1 olarak kodlanırsa, $P(x)$ ifadesi x verildiğinde y 'nin 1'e eşit olma koşullu olasılığını $[P(x) = P(y=1|x)]$ ve verilen herhangi bir x için y 'nin 0'a eşit olma koşullu olasılığını $[1 - P(x) = P(y=0|x)]$ göstermektedir. (x_i, y_i) çiftinin $y_i = 1$ olduğu zaman olabilirlik fonksiyonuna katkısı $P(x_i)$ iken, $y_i = 0$ olduğu zaman olabilirlik fonksiyonuna katkısı $1 - P(x_i)$ kadardır. (x_i, y_i) çiftinin olabilirlik fonksiyonuna katkısını ifade etmenin diğer bir yolu,

$$P(y_i | x_i) = P(x_i)^{y_i} [1 - P(x_i)]^{1-y_i} \quad (3.8)$$

eşitliğini kullanmaktır.

Gözlemlerin birbirinden bağımsız oldukları varsayıldığı için, olabilirlik fonksiyonu denklem (3.8)'deki terimlerin n gözlem için çarpılmasıyla elde edilir. Buna göre olabilirlik fonksiyonu,

$$L(y | x, \beta) = \prod_{i=1}^n P(x_i)^{y_i} [1 - P(x_i)]^{1-y_i} \quad (3.9)$$

olarak yazılır.

En çok olabilirliğin temel ilkesi, β tahmininin (3.9)'daki eşitliği maksimum yaptığıdır. Matematiksel olarak denklem (3.9)'daki eşitliğin **doğal logaritmasıyla** çalışmak daha kolay olacağından log-olabilirlik fonksiyonu,

$$\log L(y | x, \beta) = \sum_{i=1}^n \{y_i \log[P(x_i)] + (1 - y_i) \log[1 - P(x_i)]\} \quad (3.10)$$

şekline dönüştürülebilir.

$\log L(y|x, \beta)$ 'yı maksimum yapan β değerini bulmak için, β_0 ve β_1 'e göre türevi alındıktan sonra, elde edilen ifadeler sıfıra eşitlenir. Sonuçta elde edilen eşitlikler,

$$\sum_{i=1}^n [y_i - P(x_i)] = 0 \quad (3.11)$$

ve

$$\sum_{i=1}^n x_i [y_i - P(x_i)] = 0 \quad (3.12)$$

şeklinde. Bu eşitliklere lojistik fonksiyonun “*olabilirlik eşitlikleri*” denir.

Lineer regresyonda fonksiyondan sapmalar kareler toplamının türevi ile elde edilen normal eşitlikler, bilinmeyen parametreler bakımından lineerdir ve kolayca çözülür. Lojistik regresyonda ise (3.11) ve (3.12) eşitliklerindeki ifadeler, β_0 ve β_1 açısından lineer olmayan denklemlerdir ve bu denklemlerin çözümü için özel yöntemler gerekmektedir. Bu yöntemler iteratiftir. Bu çalışmada parametreler SAS istatistik paket programı ile tahmin edilecektir (Anonim.,1988). Ayrıca BMDP, SPSSX gibi istatistik programlar ile de iteratif yolla katsayıların tahminleri elde edilebilir. Olabilirlik eşitliklerinin çözümü ile elde edilen β değerleri *en çok olabilirlik tahmini* olarak bilinir ve $\hat{\beta}$ olarak gösterilecektir. Ayrıca $\hat{P}(x_i)$, $P(x_i)$ 'nin en çok olabilirlik tahminidir. Bu ifade belli bir x_i için y_i 'nin 1'e eşit olması şartlı olasılığının bir tahminini verir. Yani lojistik regresyon modeli için uydurulmuş veya modelden tahmin edilmiş değeri temsil eder. (3.11) eşitliği ile ilgili olarak,

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

şeklinde bir ilişki de söz konusudur. Bu, gözlenen y değerleri toplamının beklenen (modelden tahmin edilen) değerlerin toplamına eşit olduğunu göstermektedir. Bu özellik, uyum yeteneğinin belirlenmesi aşamasında önemlidir.

Sürekli bağımsız değişken olarak YAŞ ve bağımlı değişken olarak doğum ağırlığı için analiz (Ek-2) sonuçları Şekil 3.3'te verilmiştir.

Görüldüğü gibi β_0 ve β_1 'in maksimum olabilirlik tahminleri sırasıyla $\hat{\beta}_0 = 3.3212$ ve $\hat{\beta}_1 = -0.1817$ olarak bulunmuştur. Buna göre, (3.7) ve (3.6) eşitlikleri kullanılarak en çok olabilirlik tahmini, $\hat{P}(x_i)$, ve lojit tahmini, $\hat{\ell}(x_i)$, sırasıyla;

$$\hat{P}(x_i) = \frac{e^{3.3212 - 0.1817 (YAS_i)}}{1 + e^{3.3212 - 0.1817 (YAS_i)}}$$

ve

$$\hat{\ell}(x_i) = 3.3212 - 0.1817(YAS_i)$$

şeklinde oluşturulabilir.

MAXIMUM-LIKELIHOOD ANALYSIS					
(1)					
Iteration	Sub Iteration	-2 log Likelihood	Convergence Criterion	Parameter Estimates	
				1	2
0	0	393.7676	1.0000	0	0
1	0	302.01788	0.2329	1.9971	-0.1176
2	0	296.92922	0.0168	3.0883	-0.1704
3	0	296.7832	0.000492	3.3131	-0.1813
4	0	296.78303	5.9895E-7	3.3212	-0.1817
5	0	296.78303	9.695E-13	3.3212	-0.1817

ANALYSIS OF MAXIMUM-LIKELIHOOD ESTIMATES					
Effect	Parameter	(2) Estimate	(3) Standard Error	(4) Chi-Square	(5) Prob
INTERCEPT	1	3.3212	0.7763	18.30	0.0000
AY	2	-0.1817	0.0336	29.34	0.0000

Şekil 3.3. Tek Bağımsız Değişkenli Basit Lojistik Regresyon Modeli Analizi.

- (1) -2 (log olabilirlik).
- (2) β_0 ve β_1 parametrelerinin tahminleri.
- (3) parametre tahminlerine ait standart hatalar.
- (4) parametre tahminlerine ait ki-kare değerleri.
- (5) parametre tahminlerinin önemlilik durumu için olasılık değerleri.

Şekil 3.3'te verilen -2(Log olabilirlik) değeri kullanılarak log olabilirlik değeri,

$$-2 \log \text{olabilirlik} = 296.78303$$

ve buradan da,

$$\log \text{olabilirlik} = -148.391515$$

şeklinde bulunur.

3.2.4. Lojistik Regresyon Katsayılarının Önem Testi

Hem basit hem de çoklu lojistik regresyonda katsayılar hesaplandıktan sonra katsayıların önemliliği test edilir. Katsayıların önem testi, 1) *Olabilirlik oran* (likelihood ratio) testi, 2) *Wald* testi, 3) *Skor* (score) testi olmak üzere üç farklı yöntemle yapılabilir.

1) Olabilirlik Oran Testi

Bu işlem, genelde modelde bulunan bağımsız değişkenlerin sonuç değişkeniyle "önemli" bir ilişki içinde olup olmadığına dair hipotezleri test etmeyi gerektirir. Bunun için, incelenmek istenen herhangi bir değişkeni içeren modelin (*tüm model*), cevap değişkeni hakkında o değişkeni içermeyen modele (*indirgenmiş model*) göre daha fazla bilgi verip vermediği araştırılır. Gözlenen ve tahmin edilen

değerleri karşılaştırma işlemi *log olabilirlik fonksiyonu* ile yapılır. Tüm model, bağımsız değişken sayısı kadar (p tane) parametre içeren modeldir.

Olabilirlik fonksiyonunu kullanarak, gözlenen değerlerle tahmin edilen değerlerin karşılaştırılması işlemi aşağıdaki ifade ile yapılmaktadır.

$$D = -2 \log \left(\frac{\text{İndirgenmiş modelin olabilirliği}}{\text{Tüm modelin olabilirliği}} \right) \quad (3.13)$$

Parantez içindeki ifade “*olabilirlik oranı*”nı göstermektedir. Logaritmanın –2 katının alınmasının nedeni matematiksel olduğu kadar dağılımı bilinen bir değer elde edilebilmesidir. Sonuçta D değeri, hipotez testi amacıyla kullanılabilir. Böyle bir teste “*olabilirlik oran testi*” adı verilir. Eşitlik (3.13) ile verilen *olabilirlik oranı* eşitlik (3.10)’daki *log olabilirlik fonksiyonu* cinsinden yazılacak olur ise,

$$D = -2 \sum_{i=1}^n \left[y_i \log \left(\frac{\hat{P}_i}{y_i} \right) + (1 - y_i) \log \left(\frac{1 - \hat{P}_i}{1 - y_i} \right) \right] \quad (3.14)$$

elde edilir. Burada $\hat{P}_i = \hat{P}(x_i)$ ’dir (Hosmer ve Lemeshow, 1989). D istatistiği bazı araştırmacılar tarafından *sapma değeri* (deviance) olarak isimlendirilir ve uyum iyiliğine karar verirken önemli bir rol oynar. Lojistik regresyon için sapma değeri, lineer regresyondaki HataKT ile aynı rolü üstlenmektedir. Aslında denklem (3.13)’de verilen sapma, bir tane bağımsız değişkenin bulunduğu basit lojistik regresyon için hesaplanırsa lineer regresyondaki HataKT’na eşit olduğu görülür.

Bağımsız bir değişkenin önemine karar vermek amacıyla, denklemde bu bağımsız değişkenin olduğu ve olmadığı durumlardaki D değerleri karşılaştırılır. Bağımsız değişkenin bulunup bulunmamasından dolayı ortaya çıkan D’deki değişim, (3.14) eşitliğinden elde edilen *sapmalar* (D) kullanılarak,

$$G = D(\text{İndirgenmiş model için}) - D(\text{Tüm model için}) \quad (3.15)$$

şeklinde ifade edilir. Bu istatistik, lineer regresyonda kullanılan *F* testindeki pay kısmı ile aynı rolü üstlenir. G’yi hesaplamak için farkı alınacak D değerlerinin her ikisi için de tüm modelin olabilirlikleri ortak olduğu için, G istatistiği,

$$G = -2 \log \left(\frac{\text{İndirgenmiş modelin olabilirliği}}{\text{Tüm modelin olabilirliği}} \right) \quad (3.16a)$$

$$= -2 (\text{İndirgenmiş modelin log olabilirliği} - \text{Tüm modelin log olabilirliği}) \quad (3.16b)$$

olarak da elde edilir.

Tek bağımsız değişkenli **basit lojistik regresyon özel durumunda** değişkenin modelde olmadığı zamanki β_0 ’ın maksimum olabilirlik tahmini $\log(n_1/n_0)$ ’dır [$n_1 = \sum y_i$ ve $n_0 = \sum (1 - y_i)$]. G istatistiği ise

$$G = -2 \log \left[\frac{\left(\frac{n_1}{n} \right)^{n_1} \left(\frac{n_0}{n} \right)^{n_0}}{\prod_{i=1}^n \hat{P}_i^{y_i} (1 - \hat{P}_i)^{(1-y_i)}} \right] \quad (3.17)$$

veya

$$G = 2 \left\{ \sum_{i=1}^n [y_i \log(\hat{P}_i) + (1 - y_i) \log(1 - \hat{P}_i)] - [n_1 \log(n_1) + n_0 \log(n_0) - n \log(n)] \right\} \quad (3.18)$$

şeklinde de elde edilebilir.

$H_0: \beta_1 = 0$ hipotezi altında, G istatistiği 1 serbestlik dereceli ki-kare dağılımına sahip olacaktır. Denklem (3.18)'deki ilk terim, değişken modeldeyken elde edilen log olabilirlik değeridir. Denklemin kalan kısmı kolayca n_0 ve n_1 değerleri denklemde yerine koyularak elde edilir. Ek 12'den $n_0 = 207$, $n_1 = 77$, $n_0 + n_1 = 284$ ve Şekil 3.3'den log olabilirlik değeri kullanılarak (3.18) eşitliğine göre G değeri,

$$\begin{aligned} G &= 2 \{ -148.391515 - [207 \log(207) + 77 \log(77) - 284 \log(284)] \} \\ &= 2 \{ -148.391515 - (-166.942905) \} \\ &= 37.102 \end{aligned}$$

şeklinde hesaplanabilir. Hesap yoluyla elde edilen G değeri 1 serbestlik dereceli ki-kare cetvel değerinden ($37.102 > \chi^2_{1,0.05} = 3.841$) büyüktür. Yani H_0 hipotezi reddedilmiştir.

İkinci bir yol olarak, tek bağımsız değişkenli basit durum için, ilk önce sadece sabit terimi içeren ve daha sonrada sabit terimle birlikte bağımsız değişkeni içeren yeni bir model oluşturulur. Bağımsız değişkenin modele eklenmesi log olabilirlik değerinde artışa neden olur. Log olabilirlik oran testi, her iki modelin log olabilirlik değerleri arasındaki farkın -2 katına eşittir. Şekil 3.3'de verilen tek bağımsız değişkenli basit durum için, sadece sabit terimi içeren (YAS yani x_{i1} 'in bulunmadığı) modelin analizi (Ek-3) ile elde edilen log olabilirlik değeri $333.88581/(-2) = -166.942905$ (Şekil 3.4), tek bağımsız değişken olan YAS modeldeyken log olabilirlik değeri ise -148.391515 idi. Buna göre olabilirlik oran testi,

$$\begin{aligned} \text{Log olabilirlik oranı} &= -2[(-166.942905) - (-148.391515)] \\ &= -2[-18.55139] \\ &= 37.102 \end{aligned}$$

olarak hesaplanabilir.

MAXIMUM-LIKELIHOOD ANALYSIS					
Iteration	Sub Iteration	-2 Log Likelihood	Convergence Criterion	Parameter Estimates	
0	0	393.7076	1.0000	0	
1	0	334.16399	0.1512	-0.9014	
2	0	333.88587	0.000832	-0.9701	
3	0	333.88581	1.8348E-7	-0.9712	
4	0	333.88581	1.004E-14	-0.9712	

ANALYSIS OF MAXIMUM-LIKELIHOOD ESTIMATES					
Effect	Parameter	Estimate	Standard Error	Chi- Square	Prob
INTERCEPT	1	-0.9712	0.1329	53.36	0.0000

Şekil 3.4. Yalnızca Sabit Terimi İçeren Basit Lojistik Regresyon Modeli Analizi.

2) Wald Testi

Bu test için gerekli varsayımlar *log olabilirlik oran testindeki* varsayımlarla aynıdır. *Wald* testi, eğim parametresinin en çok olabilirlik tahmini olan $\hat{\beta}_1$ 'in kendi standart hatasına bölünmesiyle elde edilir. Elde edilen oran standart normal dağılışı gösterir ve $\hat{\beta}_1 = 0$ şeklindeki hipotezi test eder. Örneğin Şekil 3.3'deki lojistik regresyon analizi için Wald testi,

$$W = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \sim Z_{(\alpha \text{ veya } \alpha/2)} \quad (3.19)$$

şeklindedir ve hesap değeri,

$$W = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{-0.1817}{0.0336} = -5.407$$

olarak bulunmuştur. Wald testinde de YAS'a ait katsayı önemlidir ($|-5.407| > Z_{0.05/2} = 1.96$). Ancak bazı durumlarda olabilirlik oran testinde önemli bulunan bir katsayı, Wald testinde önemsiz çıkabilmektedir. Bu nedenle olabilirlik oran testinin kullanılması önerilir. Dikkat edilirse hem olabilirlik oran testi (G) hem de Wald testi (W) için, katsayıların en çok olabilirlik tahminlerinin bilinmesi gerekmektedir. Tek değişkenli durumda bu değerlerin hesaplanması zor değildir ancak değişken sayısı arttıkça hesaplamalar zorlaşır.

3) Skor Testi

Skor testi, bu hesaplamaları gerektirmez. Bu testin avantajı, hesaplamaların diğer testlere göre daha az olması, dolayısıyla kolay hesaplanabilmesidir. Skor testi, 3.7 eşitliğinin şartlı dağılışına bağlıdır. Bu test için gerekli varsayımlar log olabilirlik

oran testindeki varsayımlarla aynıdır. Tek değişkenli durumda(basit lojistik regresyon durumu), skor testi için eşitlik

$$ST = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sqrt{\bar{y}(1 - \bar{y}) \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (3.20)$$

şeklindedir.

Burada, $\bar{y} = n_1/n$ 'e eşittir ve skor testi standart normal dağılış gösterir.

3.2.5. Desen Değişkenleri Tanımlanması

Lojistik regresyonda bağımsız değişken, x , *adlandırma* ya da *sıralama* ölçeğinde kesikli bir değişken ise “*desen değişkeni*” (gölge, kukla, dummy, design) tanımlamak gerekir. Kesikli x bağımsız değişkeninin sınıf sayısı r ile gösterilsin. Buna göre, tanımlanacak desen değişkeni sayısı $r-1$ tanedir.

Eğer kesikli x bağımsız değişkeni iki sınıflı ($r=2$) ise,

- a) kısmi
- b) marjinal

olmak üzere iki türlü desen değişkeni belirleme yöntemi vardır. Kısmi yöntemde sınıflardan herhangi biri için 0, diğeri için 1 kodlanırken, marjinal yöntemde kodlama -1, 1 şeklinde yapılır. Örneğin, doğan çocukların cinsiyetleri (x) inceleniyor ise desen değişkeni (D) Çizelge 3.4’deki gibi tanımlanır.

Çizelge 3.4. $r=2$ İçin Desen Değişkenlerinin Kısmi ve Marjinal Yöntemle Tanımlanması

Cinsiyet	Araştırmacı kodlaması	Kısmi yöntem	Marjinal yöntem
		D	D
Erkek	1	0	-1
Dişi	2	1	1

Eğer kesikli x bağımsız değişkeni ikiden fazla ($r>2$) sınıflı ise,

- a) Referans hücre yöntemi,
- b) Lojitten sapma yöntemi,
- c) Ortogonal polinomlar

şeklinde üç yöntemle desen değişkenleri belirlenebilir. Referans yöntemi ile lojitten sapma yöntemi gerek adlandırma gerekse sıralama ölçekli değişkenlerde kullanılabilirken, ortogonal polinomlar yalnızca en azından sıralama ölçeğindeki bir x değişkeni için kullanılabilir.

Örneğin gelen hastaların ikamet ettiği şehirler (Adana, K.Maraş, Mersin, Hatay) inceleniyor olsun. Adlandırma ölçeğindeki bu bağımsız değişken için $4-1=3$

tane (D_1, D_2, D_3) desen değişkeni tanımlanabilir. Referans hücre yönteminde referans alınan sınıf (örn. Adana) için Bütün D'ler 0 olarak kodlanır. Kalan üç tane sınıf için de köşegende 1, köşegen dışında 0 olacak şekilde kodlama yapılır (Çizelge 3.5). Marjinal yöntemde ise sınıflardan bir tanesi (örn. Adana) için tüm desen değişkenleri -1 ile kodlanır. Kalan üç sınıf için de frekans yöntemindeki kodlama şekli uygulanır.

Çizelge 3.5. $r>2$ İçin Desen Değişkenlerinin Kısmi ve Marjinal Yöntemle Tanımlanması

Yörelere	Araştırmacı kodlaması	Referans hücre yöntemi			Marjinal yöntem		
		D_1	D_2	D_3	D_1	D_2	D_3
Adana	1	0	0	0	-1	-1	-1
K.Maraş	2	1	0	0	1	0	0
Hatay	3	0	1	0	0	1	0
Mersin	4	0	0	1	0	0	1

Örneğin bir araştırmada işletme tipi (küçük, orta, büyük) ya da deneklerin yaş grupları (11-20, 21-30, 31-40, 41-50) inceleniyorsa bu özellikler sıralama ölçeğindedir. Yani belli bir büyüklük sırasına dizmek mümkündür. Bu gibi durumlarda referans hücre ve marjinal yöntemlerine ek olarak ortogonal polinomlar yöntemi de kullanılabilir. Ancak burada bu yöntemler ele alınmayacaktır.

3.2.6. Çoklu Lojistik Regresyon Modeli

Önceki bölümlerde lojistik regresyon modeli, tek değişkenli, basit lojistik regresyon modeli olarak tanıtıldı. Doğum ağırlığı verilerine ilişkin bir örnek te verildi. Bu örnekte, ikil bağımlı y değişkeni olarak doğum ağırlığı (düşük=1, yüksek=0) ile bağımsız değişken olarak da YAS ele alınmıştı. Ancak uygulamada, bağımlı bir değişkeni etkileyen birden çok bağımsız (açıklayıcı) değişken söz konusudur. Örneğin doğum ağırlığını etkileyen, YAS dışında son adet dönemi anne ağırlığı, bebeğin ultrasonda tespit edilen cinsiyeti ve yüksek tansiyon gibi başka bağımsız değişkenler vardır. Bu türlü bağımsız değişkenlerin ikil bağımlı y üzerine birlikte etkisinin incelenmesi "*çoklu lojistik regresyon*" yolu ile yapılır.

Bu bölümde lojistik regresyon birden fazla bağımsız değişken için geliştirilecek, modeldeki katsayıların tahmini ve önem testleri ele alınacaktır. Aynı zamanda kesikli ve adlandırma (nominal) ölçekli bağımsız değişkenleri modele dahil etmek için desen değişkenlerinin kullanımı tanıtılacaktır.

$x = (x_1, x_2, \dots, x_p)$ şeklinde p tane bağımsız değişken için (3.6) nolu eşitlikten yararlanarak *lojistik model*,

$$\ell(\mathbf{x}_i) = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad (3.21)$$

şeklinde yazılabilir.

Lojistik regresyonda, bağımsız x değişkeni adlandırma ya da sıralama ölçeğinde kesikli bir değişken ise desen değişkeni kullanmak gerekir. x 'in r tane sınıfı var ise $r-1$ tane desen değişkeni oluşturulur. Desen değişkenlerinin oluşturulması bölüm 3.2.5'de anlatılmıştı.

Eğer p tane x bağımsız değişkeninden m -inci sıradaki, adlandırma ölçeğinde kesikli bir değişken ise desen değişkenli *lojit model*,

$$\begin{aligned} \ell(\mathbf{x}_i) &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{m-1} x_{i(m-1)} + \beta_{m1} D_{im1} + \beta_{m2} D_{im2} + \dots \\ &\dots + \beta_{mr-1} D_{imr-1} + \beta_{m+1} x_{i(m+1)} + \dots + \beta_p x_{ip} \\ &= \beta_0 + \sum_{k=1}^{m-1} \beta_k x_{ik} + \sum_{j=1}^{r-1} \beta_m D_{imj} + \sum_{k=m+1}^p \beta_k x_{ik} \end{aligned} \quad (3.22)$$

ve $P(\mathbf{x}_i) = P(y=1 | \mathbf{x})$, y 'nin gerçekleşmesine ait şartlı olasılık olmak üzere, *lojit fonksiyon*,

$$P(\mathbf{x}_i) = \frac{e^{\ell(\mathbf{x}_i)}}{1 + e^{\ell(\mathbf{x}_i)}} \quad (3.23)$$

şeklinde yazılabilir.

Parametrelerin Tahmini

(x_i, y_i) , $i = 1, 2, \dots, n$ şeklinde n tane bağımsız gözlem çifti bulunan bir örnek olduğu varsayalım. Çoklu lojistik regresyonda *en çok olabilirlik* yöntemi ile katsayıları tahmin edilebilir. Ayrıca *iteratif olmayan ağırlıklandırılmış EKK* yöntemi ile *diskriminant fonksiyonu* yöntemleri de vardır. Burada *en çok olabilirlik* yöntemi dikkate alınacaktır.

Çoklu lojistik regresyonda, basit lojistik regresyona göre yalnızca x 'lerin sayısı ve dolayısı ile lojit model, $\ell(\mathbf{x}_i)$, farklılığı söz konusudur. Olabilirlik fonksiyonu, $L(y | \mathbf{x}) = P(y | \mathbf{x}_i)$, elde edildikten sonra p tane β için tek tek kısmi türev alınıp sıfıra eşitlenerek,

$$\sum_{i=1}^n [y_i - P(\mathbf{x}_i)] = 0 \quad (3.24)$$

ve

$$\sum_{i=1}^n x_{ik} [y_i - P(\mathbf{x}_i)] = 0, \quad k=1, 2, \dots, p \quad (3.25)$$

şeklinde $p+1$ tane olabilirlik denklemi elde edilir. Bu denklem sisteminin çözümü ile $\hat{\beta}$ tahminleri elde edilebilir. Bu tahminler kullanılmak sureti ile lojit fonksiyonun tahmini, $\hat{\ell}(\mathbf{x}_i)$, de elde edilebilir.

Önem testlerinin yapılabilmesi ve güven aralıklarının bulunabilmesi için katsayıların standart hatalarının bilinmesi gerekir. $\hat{\beta}_k$ için standart hata,

$$SE(\hat{\beta}_k) = [\hat{\sigma}_k^2]^{1/2} \quad (3.26)$$

şeklinde dir. Burada $\hat{\sigma}_k^2$, $\hat{\beta}_k$ değerinin varyansıdır. Bu varyanslar, “bilgi” (information) matrisinin tersinin köşegen elemanlarıdır. Bilgi matrisi $\mathbf{I} = \mathbf{X}'\mathbf{C}\mathbf{X}$ şeklindedir ve $(p+1) \times (p+1)$ boyutludur. Burada $n \times (p+1)$ 'lik $\mathbf{X}$ matrisi ile $n \times n$ 'lik $\mathbf{C}$ matrisi,

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}$$

ve

$$\mathbf{C} = \begin{bmatrix} \hat{P}_1(1-\hat{P}_1) & 0 & \dots & 0 \\ 0 & \hat{P}_2(1-\hat{P}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{P}_n(1-\hat{P}_n) \end{bmatrix}$$

şeklinde dir.

3.2.7. Modelin ve Katsayıların Önem Testi

Tek değişkenli lojistik regresyon modelinde olduğu gibi, çok değişkenli lojistik regresyon modelinde de modeldeki değişkenlerin katsayılarının önem testleri yapılır. Katsayıların önemlilik testleri, en iyi modeli mümkün olan en az değişkenle oluşturma amacına hizmet eder. Öncelikle modele alınan tüm değişkenlerin katsayıları ($\hat{\beta}_k; k=1, 2, \dots, p$ ve $\hat{\beta}_{kj}; j=1, 2, \dots, r_k-1$) birlikte test edilir. Bunun için tüm model ile modelde yalnızca $\hat{\beta}_0$ 'ın bulunduğu indirgenmiş modelin log olabilirlikleri yardımı ile, olabilirlik oran testi, (3.16b) eşitliği ile,

$$G = -2[\text{indirgenmiş modelin log olabilirliği} - \text{tüm modelin log olabilirliği}]$$

şeklinde hesaplanabilir.

Olabilirlik oran testi, G , $\chi^2_{SD,\alpha}$ dağılışı gösterir. Burada $SD = v_{\text{tüm}} - v_{\text{ind}}$, χ^2 için serbestlik derecesidir. Burada $v_{\text{tüm}} = (\text{tüm modeldeki } \hat{\beta} \text{ sayısı} + 1)$, $v_{\text{ind}} = (\text{indirgenmiş modeldeki } \hat{\beta} \text{ sayısı} + 1)$ şeklindedir. Modelde desen değişkeni tanımlanmış kesikli bağımsız bir değişken varsa, her bir desen değişkeni de sayılmalıdır. Yalnızca β_0 'ın dikkate alındığı indirgenmiş model için lojistik regresyon analizi uygulandığında $v_{\text{ind}} = 0 + 1$ 'dir.

Modelin önemliliği test edildikten sonra önemsiz olan değişkenleri eleyerek en az değişkenle aynı sonucu verebilecek (en iyi uyum sağlayan) modeli oluşturmak istenir. Hangi değişkenin modele alınacağına, hangisinin alınmayacağına karar verebilmek için tek tek değişkenlerin katsayılarının önem testleri yapılır.

Değişkenlerin **tek tek önemlilik testleri basit lojistik regresyon yolu** ile yapılır. Yalnızca etkileri araştırılacak herhangi bir X_k değişkeni kullanılarak basit lojistik regresyon analizi yapıldıktan sonra X_k 'ya ait $\hat{\beta}_1$ katsayısının (dolayısıyla X_k 'nın) önem testi a) olabilirlik oran testi, b) wald testi ve c) skor testi ile yapılabilir. Bu çalışmada yalnızca olabilirlik oran testi üzerinde durulacaktır. Basit lojistik regresyon yolu ile $\hat{\beta}_1$ 'in önemliliğine ait *olabilirlik oran testi*, G , $SD = (1+1) - (0+1) = 1$ serbestlik dereceli χ^2 dağılışı göstermektedir. Test edilen hipotez ise, β_0 ve β_1 'i içeren tüm model ile yalnızca β_0 'ı içeren indirgenmiş modelin aynı olduğu şeklindedir. Kısaca $H_0: \beta_1 = 0$ olarak gösterilebilir.

Her bir β_k 'nin basit lojistik regresyon yolu ile tek tek önemlilik testleri yapıldıktan sonra, önemsiz olanlardan başlanarak birer birer modelden bir değişkenin çıkarıldığı model *indirgenmiş*, çıkarılmadan önceki model *tüm* kabul edilip her iki modelin aynılığına dair H_0 hipotezi, G ile test edilir. H_0 hipotezi kabul edilirse ilgili X_k değişkeni modelden çıkarılır. Bu teste p tane X_k için devam edilir. Daha sonra önemsiz bulunan herhangi bir X_k ile ikinci bir X_k birlikte çıkarılarak ($k \neq k'$) bu indirgenmiş model ile tüm modelin aynılığı test edilebilir. Test işlemine benzer şekilde devam edilebilir.

Kategorik olarak ölçeklendirilmiş bir değişken modelden çıkarıldığı zaman, onun bütün desen değişkenleri de modelden çıkarılmalıdır. Örneğin kategorik bir değişkenin r düzeyi varsa, bu değişkenin modelden çıkarılmasının olabilirlik oran testi için serbestlik derecesine katkısı $k-1$ olacaktır.

3.2.8. Basit Lojistik Regresyonda Katsayıların Yorumlanması

Buraya kadar lojistik modelin oluşturulması, katsayıların önemlilik testleri ile modelin önemlilik testleri üzerinde duruldu. Mümkün olan en az değişken ile en uygun model oluşturulmaya çalışıldı.

Model oluşturulduktan sonra katsayıların yorumlanması aşamasına gelinir. Bilindiği gibi bağımsız bir X_k değişkeninin katsayısı (β_k) "*eğim*" parametresidir ve

X_k 'da meydana gelen bir birim değişimin bağımlı değişken y üzerinde meydana getirdiği değişimin miktarını ve yönünü verir.

Bir modeldeki bağımsız değişkenler ile bağımlı değişken arasındaki lineer ilişkiyi veren fonksiyona “*link fonksiyonu*” denir. Lineer regresyonda link fonksiyonu, I , identity matrisidir. Lojistik regresyon modelinde ise link fonksiyonu lojit dönüşümdür. Lojit dönüşüm ile elde edilen tek değişkenli (tek bir bağımsız x değişkeni içeren) basit lojistik model hatırlanacağı gibi,

$$\begin{aligned}\ell(x) &= \log[P(x)/(1 - P(x))] \\ &= \beta_0 + \beta_1 x_1\end{aligned}$$

şeklinde idi. Buna göre β_1 katsayısı X 'deki bir birim artışın lojitte sağlayacağı değişim demektir ve,

$$\hat{\beta}_1 = \hat{\ell}(x+1) - \hat{\ell}(x) \quad (3.27)$$

olarak ifade edilir.

Lojistik regresyonda katsayıların yorumlanması için “*odds*”lar ve “*odds oranı*”ndan yararlanılır. X 'in herhangi bir grubu ($X=0$) için $y=1$ olmasına ilişkin olasılık “*odds*” olarak adlandırılır. “*odds*”, lojitin doğal logaritması alınmamış halidir ve örneğin $X=1$ için odds değeri,

$$\frac{P(x=1)}{1-P(x=1)} \quad \text{veya} \quad \frac{P(1)}{1-P(1)}$$

olarak, $X=0$ için odds değeri,

$$\frac{P(x=0)}{1-P(x=0)} \quad \text{veya} \quad \frac{P(0)}{1-P(0)}$$

şeklinde gösterilir.

İki Sınıflı Bağımsız Değişken Durumu

Basit lojistik regresyondaki tek bağımsız değişken, kesikli ve iki sınıflı olduğunda kısmi yöntemle göre $D1=0$ ve $D2=1$, marjinal yöntemle göre de $D1=-1$ ve $D2=1$ olarak kodlanıyordu. X 'in desen değişkeni olarak **kodlama yöntemine göre** $P(x)$ ve $1-P(x)$ 'in alacağı değerler de değişmektedir. Buna göre $e^{\beta_0 + \beta_1 x}$ veya desen değişkeni kullanılıyor ise $e^{\beta_0 + \beta_1 D}$ değeri,

$$X=0 \text{ (ya da } D=0) \text{ için } e^{\beta_0}$$

$X=1$ (ya da $D=1$) için $e^{\beta_0+\beta_1}$

ve

$X=-1$ (ya da $D=-1$) için $e^{\beta_0-\beta_1}$

olacaktır. X 'in 0, 1 veya -1, 1 şeklinde kodlanması durumunda $y=1$ veya $y=0$ sonuçlarının X 'e göre koşullu olasılıkları Çizelge 3.6.'da verilmiştir.

Çizelge 3.6. İki Sınıflı Bağımsız Değişkene Sahip Basit Lojistik Regresyonda $y=1$ ve $y=0$ İçin Koşullu Olasılıklar

Kısmi yöntemle kodlanma durumu		Bağımsız Değişken X	
		$x=1$	$x=0$
Bağımlı Değişken Y	$y=1$	$P(1) = \frac{e^{\beta_0+\beta_1}}{1 + e^{\beta_0+\beta_1}}$	$P(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
	$y=0$	$1 - P(1) = \frac{1}{1 + e^{\beta_0+\beta_1}}$	$1 - P(0) = \frac{1}{1 + e^{\beta_0}}$
Toplam		1.0	1.0
Marjinal yöntemle kodlanma durumu		Bağımsız Değişken X	
		$X=1$	$x=-1$
Bağımlı Değişken Y	$y=1$	$P(1) = \frac{e^{\beta_0+\beta_1}}{1 + e^{\beta_0+\beta_1}}$	$P(-1) = \frac{e^{\beta_0-\beta_1}}{1 + e^{\beta_0-\beta_1}}$
	$y=0$	$1 - P(1) = \frac{1}{1 + e^{\beta_0+\beta_1}}$	$1 - P(-1) = \frac{1}{1 + e^{\beta_0-\beta_1}}$
Toplam		1.0	1.0

Bundan sonra *odds*'ların hesaplanmasına geçilebilir. Kodlama kısmi yöntem (0 ve 1) ile yapılmışsa, $x=1$ olan denekler arasında bağımlı değişkeninin görülme ($y=1$) *odds* değeri $P(1)/[1 - P(1)]$ olarak tanımlanır. Benzer olarak $x=0$ olan denekler arasında bağımlı değişkeninin görülme ($y=1$) *odds* değeri $P(0)/[1 - P(0)]$ olarak tanımlanır. *Odds* değerinin doğal logaritması önceden tanımlandığı gibi "*lojit*" olarak tanımlanır ve

$$\ell(1) = \log\{P(1)/[1 - P(1)]\}$$

ve

$$\ell(0) = \log\{P(0)/[1 - P(0)]\}$$

şeklinde verilebilir.

Odds oranı Ω ile gösterilecektir. Odds oranı $x=1$ için hesaplanan odds değerinin $x=0$ için hesaplanan odds değerine oranı şeklindedir. Buna göre odds oranı,

$$\Omega(1,0) = \frac{P(1)/[1 - P(1)]}{P(0)/[1 - P(0)]} \quad (3.28)$$

olarak yazılabilir.

Odds oranının doğal logaritması, “*log-odds oranı*” veya “*log-odds*” şeklinde ifade edilir ve bu da “*lojit fark*”a eşittir.

$$\log[\Omega(1,0)] = \log\left[\frac{P(1)/[1 - P(1)]}{P(0)/[1 - P(0)]}\right] = \ell(1) - \ell(0) \quad (3.29)$$

Çizelge 3.6’da lojistik regresyon için ifade edilen $P(1)$ ve $P(0)$ değerleri (3.28) eşitliğinde yerine koyulduğunda *odds oranı*,

$$\Omega(1,0) = \frac{\left(\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}\right) / \left(\frac{1}{1 + e^{\beta_0 + \beta_1}}\right)}{\left(\frac{e^{\beta_0}}{1 + e^{\beta_0}}\right) / \left(\frac{1}{1 + e^{\beta_0}}\right)} = e^{\beta_1}$$

şeklinde bulunur.

Buna göre lojistik regresyonda bağımsız değişkenin ikili olması ve 0,1 şeklinde kodlanması durumunda *odds oranı*,

$$\Omega = e^{\beta_1} \quad (3.30)$$

ve bunun doğal logaritması alınarak bulunacak olan *lojit fark* veya *log-odds* ise,

$$\log(\Omega) = \log(e^{\beta_1}) = \beta_1 \quad (3.31)$$

şeklindedir.

Aynı *lojit fark* veya *log-odds değeri* (3.29) sayılı eşitliğin son kısmı kullanılarak da elde edilebilir. Bu yolla,

$$\ell(1) - \ell(0) = \beta_0 + \beta_1(x=1) - [\beta_0 + \beta_1(x=0)]$$

$$= \beta_0 + \beta_1 - \beta_0 - 0$$

$$= \beta_1$$

bulunur.

$x=1$ olan denekler arasında sonuç değişkeninin $y=1$ olması, $x=0$ olan deneklere göre ne kadar olası (veya ne kadar olası değil) olduğunu vurguladığı için, özellikle biyolojik çalışmalarda *odds oranı* çok yaygın kullanılan bir ilişki ölçüsüdür. Örneğin, y değişkeni deneğin düşük doğum ağırlığında bebek doğurup doğurmadığını belirleyen, x değişkeni de bebeğin cinsiyetini gösteren değişkenler olduğu varsayalım. Eğer $\Omega=1.4$ ise, incelenen ana kütlede düşük doğum ağırlığının gerçekleşmesi durumunun, cinsiyetin kız olması durumunda, erkek olması durumuna göre 1.4 kat daha fazla görüldüğü söylenebilir.

Odds oranının tahmini ($\hat{\Omega}$) eğik bir dağılıma sahiptir. Teorik olarak örnek genişliği yeteri kadar büyük olduğu zaman $\hat{\Omega}$ 'nın dağılımı normal olacaktır. Odds oranı için %100 $(1-\alpha)$ güven aralığının tahmini, β_1 katsayısı için güven aralığının alt ve üst noktalarının hesaplanması ve sonra bu değerlerin üstelini ($\exp$) alınmasıyla

$$\exp[\hat{\beta}_1 \pm Z_{1-\alpha/2} \times SE(\hat{\beta}_1)] \quad (3.32)$$

şeklinde elde edilir.

Sayısal bir örnek ile odds oranının pekiştirilmesi yararlı olacaktır. Bir çalışmada, bağımlı değişkenin (Y) ve bağımsız değişkenin (X) incelendiği varsayalım. Bu durumda basit lojistik regresyonla ilgili odds oranlarını inceleyelim. X ve Y ikiye sınıflı ve 0,1 şeklinde kodlanmış olması durumunda bu iki değişkenin çapraz sınıflaması Çizelge 3.7'deki gibi olsun.

Çizelge 3.7. İkişer Sınıflı X ve Y'nin Çapraz Sınıflamasına Bir Örnek

Y	X		Toplam
	1	0	
0	10	19	29
1	4	32	36
Toplam	14	51	65

Çizelge'den 10 deneğin ($x=1, y=1$), 4 deneğin ($x=1, y=0$), 19 deneğin ($x=0, y=1$), 32 deneğin ($x=0, y=0$) değerlerini aldığı anlaşılmaktadır. Bu sebeple bu veri seti için (3.4) eşitliği ile verilen olabilirlik fonksiyonu,

$$L(\beta) = P(1)^{10} \times [1 - P(1)]^4 \times P(0)^{19} \times [1 - P(0)]^{32}$$

şeklinde hesaplanabilir. β_0 ve β_1 katsayılarının tahminleri sırasıyla $\hat{\beta}_0 = -0.5213$ ve $\hat{\beta}_1 = 1.4376$ olarak bulunmuş olsun. Buna göre,

Odds oranının en çok olabilirlik tahmini, $\hat{\Omega} = e^{1.4376} = 4.211$ 'dir. Lojistik regresyon analizi yapılmaksızın (3.28) nolu eşitlikten, Çizelge 3.7'deki çapraz sınıflandırma kullanılarak da aynı *odds oranı* ve $\hat{\beta}_1$ katsayısı,

$$\hat{\Omega} = \frac{10 / 4}{19 / 32} = 4.211$$

ve buradan da

$$\hat{\beta}_1 = \log(\hat{\Omega}) = \log(4.211) = 1.4376$$

şeklinde hesaplanabilir. Hesaplanan *odds oranı* için *güven aralıkları* ise,

$$\exp(1.4376 \pm 1.96 \times 0.6587) = (1.158, 15.312)$$

şeklinde bulunabilir.

Marjinal yöntem kullanılarak 1, -1 kodlaması yapıldığında ise *log odds*'un *tahmini*, *lojit farktan* gidilerek,

$$\log[\hat{\Omega}(1, -1)] = \ell(1) - \ell(-1)$$

$$\log[\hat{\Omega}(1, -1)] = \ell(x = 1) - \ell(x = -1)$$

$$= [\hat{\beta}_0 + \hat{\beta}_1(1)] - [\hat{\beta}_0 + \hat{\beta}_1(-1)]$$

$$= 2\hat{\beta}_1 \quad (3.33)$$

şeklinde ve *odds oranının tahmini* $\hat{\Omega}(1, -1) = \exp(2\hat{\beta}_1)$ şeklinde bulunur.

Kodlama yöntemi aynı zamanda odds oranı için güven aralığının alt ve üst sınırlarının hesaplanmasını da etkiler. Marjinal yöntem kullanıldığında güven aralığının tahmini için gerekli standart hata $SE(2\hat{\beta}_1)$ veya $2 \times SE(\hat{\beta}_1)$ değerine eşittir. Böylece güven aralığının alt ve üst sınırları,

$$\exp[2\hat{\beta}_1 \pm Z_{1-\alpha/2} 2 \times SE(\hat{\beta}_1)] \quad (3.34)$$

şeklinde verilebilir.

Dikkat edilecek olursa, basit lojistik regresyonda, X için verilen kodlar arası farkın mutlak değeri odds oranı tahminine çarpımsal bir katsayı olmaktadır. Yani 0, 1 kodlamasında $|0-1|=1$ ve -1, 1 kodlaması ise $|(-1)-(1)|=2$ olup (3.31) ve (3.33) sayılı eşitliklerdeki *odds oranı tahminlerinin* katsayılarını oluşturmuşlardır.

Genelleştirilecek olursa, basit lojistik regresyonda iki sınıflı X bağımsız değişkeni a ve b şeklinde kodlanırsa *odds oranı tahmini*,

$$\hat{\Omega} = e^{a-b|\beta_1}$$

ve *log odds* veya *lojit fark*,

$$\log(\hat{\Omega}) = a - b |\hat{\beta}_1|$$

şeklinde yazılabilir.

Odds oranı için güven aralığının alt ve üst sınırları,

$$\exp[(a-b)\hat{\beta}_1 \pm Z_{1-\alpha/2} |a-b| \times SE(\hat{\beta}_1)] \quad (3.35)$$

şeklindedir.

İkiden Fazla Sınıflı Bağımsız Değişken Durumu

Adlandırma ölçekli bağımsız değişkenin ikiden fazla düzeyinin ($r > 2$) olması durumunda desen değişkenleri kullanılarak modele dahil edilmelidir. Bu bölümde, ikiden fazla düzeye sahip adlandırma veya sıralama ölçekli bağımsız değişkenler için, desen değişkenlerini oluşturma yöntemleri üzerinde durulacaktır. Bir çalışmada X'in dört seviyede ($r=4$) kodlandığı varsayılın. Y ve ikiden fazla sınıflı X'in çapraz sınıflamasıyla ilgili bir örnek Çizelge 3.8'de verilmiştir.

Çizelge 3.8. Y ve İkiden Fazla Sınıflı X'in Çapraz Sınıflamasına Örnek

Süt Verimi	X-1. Grup	X-2. Grup	X-3. Grup	X-4. Grup	Toplam
1	14	7	3	5	29
0	1	8	16	11	36
Toplam	15	15	19	16	65
Odds oranı ($\hat{\Omega}$)	1	0.0625	0.0134	0.0325	
%95 Güven Aralığı		(0.006, 0.603)	(0.001, 0.144)	(0.003, 0.32)	
Log($\hat{\Omega}$)	0	-2.7725	-4.3125	-3.4265	

Çizelge 3.8'de *odds oranları* ($\hat{\Omega}$) ve her 4 gruba ait tahmin edilen katsayı ($\text{Log}(\hat{\Omega})$) değerleri verilmiştir. Her bir desen değişkenine ait odds oranı ve tahmin edilen katsayılara ait standart hata değerleri basit bir hesaplamayla elde edilebilir. Çizelgeden örneğin, X-1. grup içinde Y=1 olan deneklerin sayısının 14, Y=0 olan deneklerin sayısının 1 olduğu; X-4. grup içinde ise Y=1 olan deneklerin sayısının

5, Y=0 olan denklemlerin sayısının 11 gözlemlendiği anlaşılmaktadır. Örneğin X-3. grup için *odds değeri*, $(3 \times 1) / (14 \times 16) = 0.0134$ (1. grup ile çapraz çarpım) ve tahmin edilen katsayıya ait *odds oranı*, $\text{Log}(\hat{\Omega}) = -4.3125$, standart hata ise, hücre frekansları kullanılarak $(1/1 + 1/14 + 1/3 + 1/16)^{1/2} = 1.2113$ şeklinde hesaplanır.

Odds oranlarının aynı değerleri desen değişkenleri oluşturularak lojistik regresyon analizi yapan bir istatistik paket programından da elde edilebilir. **Kısmi yöntem** kullanılarak, X değişkeni Çizelge 3.9'daki gibi desen değişkeni haline getirilmiştir.

Çizelge 3.9. X'in Kısmi Yöntemle Desen Değişkenine Dönüştürülmesi

Gruplar	Desen Değişkenleri		
	D1	D2	D3
X-1. Grup	0	0	0
X-2. Grup	1	0	0
X-3. Grup	0	1	0
X-4. Grup	0	0	1

Kısmi yöntem bazı kaynaklarda “*referans hücre kodlaması*” olarak da isimlendirilmektedir. Kullanılan lojistik regresyon paket programları, en küçük değerde kodlanan grubu referans grup olarak işleme almaktadırlar.

3.grubun 1.gruba göre karşılaştırılması,

$$\begin{aligned}
 \log[\hat{\Omega} (3.\text{grup}, 1.\text{grup})] &= \hat{\ell} (3.\text{grup}) - \hat{\ell} (1.\text{grup}) \\
 &= [\hat{\beta}_0 + \hat{\beta}_{11}x(D1=0) + \hat{\beta}_{12}x(D2=1) + \hat{\beta}_{13}x(D3=0)] \\
 &\quad - [\hat{\beta}_0 + \hat{\beta}_{11}x(D1=0) + \hat{\beta}_{12}x(D2=0) + \hat{\beta}_{13}x(D3=0)] \\
 &= \hat{\beta}_{12}
 \end{aligned} \tag{3.36}$$

şeklindedir.

Odds oranı için $\%(1-\alpha)100$ 'lük güven sınırlarının tahmini, iki sınıflı bağımsız değişkenlerde kullanılan yaklaşıma benzer bir yaklaşımla elde edilir. Lojistik regresyon katsayısı için güven aralığının alt ve üst sınırları,

$$\hat{\beta}_{ij} \pm Z_{1-\alpha/2} \times SE(\hat{\beta}_{ij})$$

şeklinde elde edilir.

Odds oranı için güven sınırları ise sınırların üsteli (exp) alınarak,

$$\exp(\hat{\beta}_{ij} \pm Z_{1-\alpha/2} \times SE(\hat{\beta}_{ij})) \quad (3.37)$$

şeklinde bulunur.

Kısmi yöntemin yaygın olarak kullanılmasının başlıca sebeplerinden biri, diğer grupların (2.grup, 3.grup, 4.grup) riskini, kontrol grubunun (1.grup) riskine göre tahmin etmesidir. Hesaplamaların diğer yöntemlere göre daha kolay olması da tercih unsurudur.

Desen değişkenlerini kodlamada kullanılan ikinci yöntem ise **marjinal yöntemdir** ve “*ortalamadan sapma*” yöntemi olarak adlandırılır. Bu, genel ortalamadan grup ortalamasının sapması gibi bir etkiyi ifade eder. Bu durumda lojistik regresyonda “grup ortalaması” grup için lojit ve “genel ortalama” da ortalama lojittir. Ortalamadan sapma yönteminde, sınıflardan birisi için bütün desen değişkenleri -1’e eşitlenirken, diğer sınıflara 0, 1 kodlaması yapılarak desen değişkenleri elde edilir. Çizelge 3.10’da ortalamadan sapma yöntemi kullanılarak desen değişkenleri oluşturulmuştur.

Çizelge 3.10. X’in Ortalamadan Sapma Yöntemi İle Desen Değişkenine Dönüştürülmesi

Gruplar	Desen Değişkenleri		
	D1	D2	D3
X-1. Grup	-1	-1	-1
X-2. Grup	1	0	0
X-3. Grup	0	1	0
X-4. Grup	0	0	1

Kısmi yöntemde olduğu gibi, bu yöntemde de Çizelge 3.8’deki veriler kullanılarak katsayı tahminleri yapılabilir. Bunun için X değişkeninin 4 grubunun her biri için çapraz çarpımlar yolu ile,

$$\hat{\ell}_1 = \log[(14/15)/(1/15)] = 2.63905,$$

$$\hat{\ell}_2 = \log[(7/15)/(8/15)] = -0.13353,$$

$$\hat{\ell}_3 = \log[(3/19)/(16/19)] = -1.67397,$$

$$\hat{\ell}_4 = \log[(5/16)/(11/16)] = -0.78845,$$

ve lojitlerin ortalaması $\bar{\ell} = \sum \hat{\ell}_i / r = [2.63905 + \dots + (-0.78845)] / 4 = 0.01077$ ’dir ($i=1, 2, \dots, r$; örnek için $r=4$). Desen değişkeni D1 için tahmin edilen katsayı,

$$\hat{\ell}_2 - \bar{\ell} = -0.13353 - 0.01077 = -0.14430$$

şeklinde hesaplanabilir.

Her bir desen değişkeni için katsayılar tahmin edilirken $j=1, 2, 3=r-1$ olmak üzere, $\hat{\ell}_{j+1} - \bar{\ell}$ şeklinde bir ilişki vardır. Yani D1 için $\hat{\ell}_2$, D2 için $\hat{\ell}_3$ ve D3 için de $\hat{\ell}_4$ 'ün ortalama lojitten farkları kullanılır.

Tahmin edilen katsayıların yorumu, referans grup (kısmi) yönteminde olduğu kadar kolay ve açık değildir. Tahmin edilen katsayıların üsteli, odds'ların geometrik ortalamasına göre, belirli bir grup için odds'un oranını verir. Örneğin, Çizelge 3.10'daki ikinci grup için (1. Desen değişkeni için),

$$\begin{aligned} \text{Exp}(\beta_{11}) &= \exp(\hat{\ell}_2 - \bar{\ell}) \\ &= \exp(-0.13353 - 0.01077) \\ &= \exp(-0.14430) \\ &= 0.8656 \end{aligned}$$

şeklinde hesaplanabilir. Bulunan bu değer, "ortalama odds"a göre odds değeridir.

Ortalamadan sapma yöntemiyle bulunan katsayılar, bir grubun referans bir gruba göre odds oranını tahmin etmek için kullanılabilir. Bununla beraber tahminler için elde edilen eşitlikler, referans hücre kodlamasına göre daha karmaşıktır.

Çizelge 3.10'daki kodlama göz önünde bulundurulduğunda, 2.grubun 1.gruba göre *log odds*'lerinin tahmini (*lojit fark*),

$$\begin{aligned} \log[(\hat{\Omega}(2.\text{grup}, 1.\text{grup}))] &= \hat{\ell}(2.\text{grup}) - \hat{\ell}(1.\text{grup}) \\ &= [\hat{\beta}_0 + \hat{\beta}_{11} \times (D1=1) + \hat{\beta}_{12} \times (D2=0) + \hat{\beta}_{13} \times (D3=0)] \\ &\quad - [\hat{\beta}_0 + \hat{\beta}_{11} \times (D1=-1) + \hat{\beta}_{12} \times (D2=-1) + \hat{\beta}_{13} \times (D3=-1)] \\ &= 2\hat{\beta}_{11} + \hat{\beta}_{12} + \hat{\beta}_{13} \end{aligned} \quad (3.38)$$

şeklinde dir.

Katsayı tahminlerinin değerleri yukarıdaki eşitlikte yerine koyulduğunda, 2.grubun 1.gruba göre *odds* değeri elde edilir.

Odds oranı için güven aralığı, güven sınırlarının üsteli (exp) alınarak elde edilir. Güven sınırlarının hesaplanmasında gerekli olacak standart hatanın (SE) bulunması için öncelikle, katsayıların toplam varyansının bulunması gerekir. Toplam varyans ise,

$$\begin{aligned}
Var\{\log[\hat{\Omega}(2.grup,1.grup)]\} &= Var(2\hat{\beta}_{11} + \hat{\beta}_{12} + \hat{\beta}_{13}) \\
&= 4Var(\hat{\beta}_{11}) + Var(\hat{\beta}_{12}) + Var(\hat{\beta}_{13}) + 4Kov(\hat{\beta}_{11}, \hat{\beta}_{12}) \\
&\quad + 4Kov(\hat{\beta}_{11}, \hat{\beta}_{13}) + 2Kov(\hat{\beta}_{12}, \hat{\beta}_{13}) \quad (3.39)
\end{aligned}$$

şeklinde ifade edilir.

Varyans bulunduğundan sonra standart hata,

$$SE\{\log[\hat{\Omega}(2.grup,1.grup)]\} = \sqrt{Var\{\log[\hat{\Omega}(2.grup,1.grup)]\}}$$

olarak bulunur.

Sürekli Bağımsız Değişken Durumu

Lojistik regresyon modelinin sürekli bir bağımsız değişken içermesi durumunda, tahmin edilen katsayıların nasıl yorumlanacağı değişkenin modele nasıl girdiğine bağlıdır. Bunun için a) kartil bölünme b) $x[\log(x)]$ değişkeni ekleme yöntemleri kullanılabilir.

Çoklu lojistik regresyon modelinde, önemli bulunan sürekli değişkenlerin modele nasıl gireceğinin belirlenmesi amacıyla ilk etapta söz konusu sürekli bağımsız değişken, x , 4 gruba (kartiller kullanılarak) ayrılır ve birinci grup referans grup olarak belirlenir. Kartil bölünme (Ek-13) için öncelikle o değişkene ait gözlem değerleri küçükten büyüğe doğru sıralanır. Kartil bölünmede sınır değerinin her iki gruba düşmesi durumunda, sınır değeri sayısal olarak en çok hangi grupta ise o gruba dahil edilir.

Kartil bölünmeye göre desen değişkenleri oluşturulduktan sonra (D1, D2, D3), söz konusu sürekli değişkenin yerine modele alınır. Desen değişkenlerini içeren çok değişkenli lojistik regresyon analizi sonucunda eğer, grupların odds oranları arasında lineer bir artış veya azalış varsa, veya gruplara ait tahmin edilen katsayılar giderek artan veya azalan bir eğilim izliyorsa, söz konusu sürekli bağımsız değişkenin lojitle lineer olduğu varsayılır.

Sürekli bağımsız, x , değişkeninin modele nasıl gireceğinin belirlenmesinde ikinci bir yol yeni bir $x[\log(x)]$ değişkeni eklenmesidir. Buna göre söz konusu değişkeni kapsayan modele $(x)[\log(x)]$ değişkeninin eklenmesi sonucunda, $(x)[\log(x)]$ değişkeninin katsayısı önemsiz çıkarsa x değişkeninin lojitle lineer olduğu varsayılır. Her iki yoldan biri ile x 'in lojitle lineer olduğu varsayımı kabul edilir ise modele sürekli bir değişken olarak girmesine karar verilir.

Gruplara ait tahmin edilen katsayılar giderek artan veya azalan bir eğilim izlemiyorsa veya yeni eklenen $x[\log(x)]$ değişkeninin katsayısı (eğimi) önemli çıkarsa söz konusu değişkenin lojitle lineer olmadığına karar verilir ve desen değişkenleri kullanılarak söz konusu değişkenin yerine modele dahil edilir. Bu durumda katsayıların yorumu, her bir desen değişkeni için (D1, D2, D3) kesikli değişkenlerde olduğu gibidir.

Araştırmacı, modele girecek desen değişkenlerinin kesim noktalarını (sınıf limitlerini) belirlerken aşağıdaki gibi üç farklı yol izleyebilir.

1. Kartil bölünmede kullanılan kesim noktaları kullanılabilir.
2. Gruplara ait tahmin edilen katsayılar veya odds oranları arasında birbirine çok yakın değerler varsa bunlar birleştirilebilir.
3. Konu uzmanları tarafından bildirilen kesim noktaları kullanılabilir.

Eğer sürekli bağımsız bir x değişkeninin lojitle lineer olduğu varsayıp aynen kalmasına karar verirse basit lojistik regresyondaki eğim katsayısı (β_1), x 'deki bir birimlik artışın *log odds* değerinde meydana getireceği değişimi gösterir. Burada X 'in herhangi bir değeri için $\beta_1 = \ell(x+1) - \ell(x)$ 'dir.

$$\exp[a\hat{\beta}_1 \pm Z_{1-\alpha/2} \times aSE(\hat{\beta}_1)] \quad (3.40)$$

Burada ' a 'nın seçimi oldukça keyfidir. Ancak seçilen ' a 'nın, sorudaki değişken ile risk sonuçlarının mevcut durumunun nasıl değiştiğini açıklar nitelikte olmasına dikkat edilmelidir. Daha önce yaş değişkeninin bulunduğu tek değişkenli lojistik regresyon sonuçlarına göre Şekil 3.3'de tahmin edilen *lojit* $\hat{\ell}(YAS_i) = 3.3212 - 0.1817(YAS_i)$ şeklinde verilmişti. Örneğin burada, 1 yıllık artış için tahmin edilen *odds oranı* $\hat{\Omega}(1) = \exp(1 \times (-0.1817)) = 0.83$ olur. Bu, yaştaki her 1 yıllık artış için 2500 gr'dan düşük doğum yapma riskinin ($2500 > DA$) 0.83 defa azaldığını (katsayı işareti negatif olduğu için) gösterir. Şekil 3.3'de verilen tek değişkenli lojistik regresyon örneğinde, YAS değişkeni için %95 güven sınırları,

$$\exp(1 \times (-0.1817) \pm 1.96 \times 1 \times 0.0336)$$

$$= (0.78, 0.89)$$

şeklinde hesaplanır.

3.2.9. Çoklu Lojistik Regresyonda Katsayıların Yorumlanması

Daha önceki bölümlerde tahmin edilen lojistik regresyon modelinde bir bağımsız değişken olması durumunda katsayı tahminleri üzerinde durulmuştu. Genellikle verilerin analizinin tek değişkenli modeller oluşturularak yapılması uygun olmamaktadır, çünkü bağımsız değişkenler genellikle birbirleriyle ilişki içinde değişimlerdir ve sonuç değişkeninin farklı düzeyleri içinde bağımsız değişkenin dağılımları farklı olabilmektedir. Bu nedenle çok değişkenli analiz verilerin modellenmesi için daha uygun olmaktadır.

Modeldeki her bir değişkenin dağılımlarındaki farklılıkları, tahmin edilen etkileri ve diğer bağımsız değişkenlerle arasındaki ilişkiyi istatistiksel olarak ayarlamak çok değişkenli analizin amacıdır. Bu durum çok değişkenli lojistik regresyon analizine uygulandığı zaman, tahmin edilen katsayının her biri modele dahil edilen diğer tüm değişkenler için ayarlama yaparak *log odds*'un tahminini sağlar. Çok değişkenli lojistik regresyon analizi sonucunda, sürekli ve kesikli

değişkenlere ait tahmin edilen katsayıların yorumlanması işlemi daha önceki bölümlerde anlatıldığı gibidir.

Modelde İnteraksiyon Olması Durumu

Modelde interaksiyon olması halinde *odds oranları* tahmini Hosmer ve Lemeshow (1989)'a göre aşağıda verilmiştir. Risk faktörü ve diğer bir değişken arasında interaksiyon olduğu zaman, risk faktörü için tahmin edilen *odds oranı* interaksiyon içinde olduğu değişkenin değerine bağlıdır. Bu durumda odds oranı hesaplanırken, interaksiyon içinde bulunan değişkenler arasındaki korelasyon hesaba katılmalıdır.

Bir risk faktörü (F), kovaryet (X) ve bunların interaksiyonu ($F \times X$)'i içeren bir model göz önünde bulundursun. Bu model için *lojistik* ($F=f$ ve $X=x$) aşağıdaki gibidir.

$$\ell(f, x) = \beta_0 + \beta_1 f + \beta_2 x + \beta_3 fx$$

$X=x$ 'de sabit tutulduğunda X 'i içine alan $F=f_1$ 'in $F=f_0$ 'a karşı *log odds* değeri,

$$\begin{aligned} \log[\hat{\Omega}(F = f_1, F = f_0, X = x)] &= \ell(f_1, x) - \ell(f_0, x) \\ &= \beta_1 (f_1 - f_0) + \beta_3 x(f_1 - f_0) \end{aligned} \quad (3.41)$$

şeklindedir.

Tahmin edilen parametreler, yukarıdaki eşitlikte (3.41) yerine koyulduğunda tahmin edilen *log odds* değeri elde edilir. (3.41) eşitliği için varyans tahmini,

$$\begin{aligned} Var\{\log[\hat{\Omega}(F = f_1, F = f_0, X = x)]\} &= Var(\hat{\beta}_1)(f_1 - f_0)^2 \\ &+ Var(\hat{\beta}_3)[x(f_1 - f_0)]^2 + 2Kov(\hat{\beta}_1, \hat{\beta}_3)x(f_1 - f_0)^2 \end{aligned} \quad (3.42)$$

şeklindedir.

Çoğu lojistik regresyon programları modeldeki tahmin edilen parametrelerin varyans ve kovaryanslarını vermektedir. Tahmin edilen değerler elde edildikten sonra denklem (3.42)'de bu değerler yerine koyularak tahmin edilen odds oranının varyansı tahmin edilir. $\Omega(F = f_1, F = f_0, X = x)$ için % $(1-\alpha)100$ 'lük güven sınırları ise

$$\exp([\hat{\beta}_1 (f_1 - f_0) + \hat{\beta}_3 x(f_1 - f_0)] \pm z_{1-\alpha/2} SE\{\log[\hat{\Omega}(F = f_1, F = f_0, X = x)]\}) \quad (3.43)$$

şeklinde verilebilir.

Risk faktörü (F) ikili bir değişken olduğu zaman bu ifade daha basit bir şekil alır. Eğer $f_1=1, f_0=0$ olursa *log odds* değeri,

$$\log[\hat{\Omega}(F = 1, F = 0, X = x)] = \hat{\beta}_1 + \hat{\beta}_3 x$$

şeklinde olur. Bu durumda varyans tahmini,

$$\begin{aligned} & Var\{\log[\hat{\Omega}(F=1, F=0, X=x)]\} \\ & = Var(\hat{\beta}_1)(f_1 - f_0)^2 + Var(\hat{\beta}_3)[x(f_1 - f_0)]^2 + 2Kov(\hat{\beta}_1, \hat{\beta}_3)x(f_1 - f_0)^2 \quad (3.44) \end{aligned}$$

şeklindedir. Odds oranı için güven sınırları ise aşağıdaki gibidir.

$$\exp((\hat{\beta}_1 + \hat{\beta}_3 x) \pm z_{1-\alpha/2} SE\{\log[\hat{\Omega}(F=1, F=0, X=x)]\}) \quad (3.45)$$

3.2.10. Model Yapılandırma

Önceki bölümlerde lojistik regresyon modelindeki katsayıların tahmin edilmesi ve yorumlanması üzerinde durulmuştur. Bağımsız değişken az sayıda olduğunda model oluşturulması kolay olmaktadır. Ancak bağımsız değişken sayısı arttıkça modele girecek değişkenlerin seçilmesi ve modelin oluşturulması oldukça zor olmaktadır. Modele ne kadar çok değişken girerse tahmin edilen standart hata da o kadar büyük olur ve gözlenen veri kümesine daha çok bağımlı olunur.

Araştırmacının hedefi en iyi modeli oluşturacak değişkenleri seçmektir. Yani mümkün olan en az sayıda değişkenle, mevcut verileri en iyi şekilde temsil edecek modeli oluşturmaktır. Bu hedefe ulaşabilmek için, modeldeki değişkenlerin teker teker incelenmesi ve modele dahil edilip edilmeyeceğine karar verilmesi gerekir.

Değişken Seçimi

Lojistik regresyonda değişken seçiminde farklı yöntemler kullanılmaktadır. Bunlar 1) basit lojistik regresyon ile katsayıların tek tek testi, 2) adımsal ve 3) en iyi alt küme seçimi yöntemleridir (Hosmer ve Lemeshow, 1989).

1) Bu yöntemlerden birincisi, değişkenlerin önce teker teker basit lojistik regresyon analizleri ile önemliliklerinin belirlenmesi ve önemli bulunan değişkenlerin çok değişkenli modele dahil edilmesini içeren iki basamaktan oluşur. Bu işlem lineer regresyondaki model oluşturmaya oldukça benzemektedir.

1. Basamak: Değişken seçme işlemine her bir değişken için ayrı ayrı tek değişkenli basit lojistik regresyon analizlerinin yapılmasıyla başlanır. Bu yöntemde bağımsız değişkenin istatistiki olarak önemli olmasının yanında biyolojik olarak da önemli olması gerekir ve dolayısıyla biyolojik çalışmalarda sıklıkla kullanılmaktadır. Olabilirlik oran ki-kare testi $r-1$ serbestlik derecesiyle, tek bağımsız değişkeni içeren basit lojistik regresyon modelindeki $r-1$ desen değişkenin katsayılarının önemi için olabilirlik oran testinin değeriyle tamamen birbirine eşittir. Pearson ki-kare testi asimtotik olarak olabilirlik oran ki - kare testine eşit olması nedeni ile aynı şekilde kullanılabilir. Bunlara ek olarak, en azından orta derecede ilişki gösteren değişkenler için, düzeylerden birini referans grup olarak kullanıp bireysel odds oranlarını ve güven aralıklarını tahmin etmekte iyi bir yöntemdir.

Herhangi bir olasılık çizelgesinin hücrelerinden birinin ya da daha fazlasının gözlenen frekansı sıfır olduğu zaman dikkat edilmelidir. Bu durum, odds oranlarının biri için tek değişkenli nokta tahminini sıfır ya da sonsuz yapar. Böyle bir değişken herhangi bir lojistik regresyon modeline dahil edilirse istenmeyen sayısal sonuçlar ortaya çıkar. Sıfırlı hücreleri önlemek için birkaç yöntem vardır. Bunlar, bağımsız değişkenin kategorilerini sıfırlı hücreyi ortadan kaldıracak şekilde birleştirmek, o kategoriye tamamen iptal etmek, eğer değişken sıralı ölçekliyse o değişkeni sürekli değişken olarak modellemek şeklinde sıralanabilir.

Sürekli değişkenler için en çok kullanılan analiz, tek değişkenli lojistik regresyon modeli oluşturarak tahmin edilen katsayısı, katsayının standart hatasını, katsayının önemi için olabilirlik oran testini ve tek değişkenli Wald istatistiğini elde etmektir. Alternatif bir analiz de iki grup t - testi olabilir.

t - testi esasına dayanan tek değişkenli analiz, değişkenin modele girip girmeyeceğine karar verme işleminde yararlı olabilir. Çünkü bu testten elde edilen değeri Wald istatistiği, Score testi ya da olabilirlik oran testinden elde edilen değeriyle aynı büyüklükte olabilmektedir.

Sürekli değişkenler için tek değişkenli lojistik uyuma karar verilmesi dağılım grafiği yardımıyla olabilir. Bu çizim lojit ölçek üzerinde yapıldığında, sadece değişkenin potansiyel önemini araştırmaya yardımcı olmakla kalmaz, aynı zamanda değişkenin uygun ölçekte olup olmadığını belirler.

2. *Basamak*: Tek değişkenli analizlerin tamamlanmasından sonra çok değişkenli analiz için değişkenler seçilebilir. Biyolojik olarak önemli olduğu düşünülen değişkenler arasında tek değişkenli test sonunda p değeri 0.25'den küçük (<0.25) olan değişkenler çok değişkenli model için aday olarak seçilebilir. Değişkenler tanımlandıktan sonra, seçilen değişkenlerin tümünü içeren model oluşturulur. Tek değişkenli yaklaşımla ilgili bir problem değişkenlerin toplam ilişkisinin göz ardı edilmesidir. Değişkenlerin her biri sonuç değişkeniyle zayıf bir ilişki içinde olsa da birlikte alındıkları zaman sonuç değişkeninin önemli bir tahmin edicisi olabilirler. Bu nedenle önemlilik düzeyi yeterince büyük seçilmelidir ki, şüpheli değişkenlerin çok değişkenli modele dahil edilmesi mümkün olabilsin.

Genel olarak mümkün olan bütün değişkenleri içeren çok değişkenli modelle işe başlamanın uygunluğu, toplam denek sayısına ve modele girmeye aday değişkenlerin sayısına göre her bir sonuç grubundaki denek sayısına bağlıdır. Veriler böyle bir analizi desteklemeye uygun olduğu zaman, bu noktadan çok değişkenli modellemeye başlamak yararlı olabilir. Bununla beraber veriler uygun olmadığı zaman bu yaklaşım sayısal olarak güvenilir çok değişkenli model oluşturabilir. Bu durumda sonuçların güvenilir olmamasından dolayı Wald istatistiği değişkenleri seçmek için kullanılmamalıdır.

2) Değişken seçimi için diğer bir yaklaşımda adımsal yöntemi kullanmaktır. Adımsal yöntemde değişkenlerin modele dahil yada hariç edilmesi tamamen istatistiksel kriterlere bağlı olarak belirli bir düzen içerisinde olmaktadır. Dolayısıyla bu yöntem biyolojik çalışmalarda kullanıldığında, biyolojik olarak önemli olmayan değişkenlerin modele girmesine ve yanlış tahminlere neden olabilmektedir. Adımsal yöntemin iki ana versiyonu vardır. Bunlar, geriye doğru eleme testiyle ileriye dönük seçim ve ileriye doğru seçim testiyle takip edilen geriye doğru eleme işlemleridir.

3) Lojistik regresyonda çok sık kullanılmayan alternatif bir seçim yöntemi de en iyi alt küme seçimidir. Bu işlemle bir, iki, üç ve bu şekilde artan sayıda değişken içeren modellerin bazı belirlenen kriterlere göre “en iyi” olduğu test edilir.

Çok değişkenli modelin oluşturulmasıyla modele dahil edilen her bir değişkenin önemi test edilmelidir. Bu durum, her bir değişkenin Wald istatistiğiyle test edilmesi ve her tahmin edilen katsayının, yalnız o değişkeni içeren tek değişkenli modelden elde edilen katsayıyla karşılaştırılmasını kapsar. Bu kriterlere göre modele katkıda bulunamayan değişkenler model dışı bırakılarak, kalan değişkenlerle yeni bir model kurulur. Yeni model, *olabilirlik oran testi* kullanılarak eski modelle karşılaştırılır. Aynı zamanda geriye kalan değişkenler için tahmin edilen kat sayılar tüm modelden elde edilenlerle karşılaştırılmalıdır. Özellikle katsayıları önemli derecede değişen değişkenler üzerinde durulmalıdır.

Eğer tek değişkenli analiz sonunda çok fazla sayıda bağımsız değişken bulunursa o zaman adımsal ya da en iyi alt küme yöntemlerini kullanmak iyi olabilir.

Önemli değişkenlerin modelde olduğuna emin olduktan sonra, değişkenler arasındaki interaksiyon terimleri incelenmelidir. Kesikli değişkenlerin kategorilerinin uygunluğu tek değişkenli analiz sırasında yapılmalıdır. Değişken seçme aşamasında lojitle lineerlik varsayımı genel bir uygulamadır ve değişkenin modelde kalıp kalmayacağına karar verme işlemiyle de uyumludur. Herhangi bir değişkeni önemli olarak tanımladıktan sonra, model iyileştirme basamağında bu değişken için doğru bir parametrik ilişki ya da ölçek elde edilebilir.

Kolaylık açısından tek değişkenli model kullanarak ölçek seçmek için bir yaklaşım açıklanacaktır. Çok değişkenli model için de konunun içeriği aynıdır. Tek değişkenli *lineer regresyonda* kurulan modelin gerçekten değişkenle lineer olup olmadığına karar vermek için, bağımlı değişkene karşı bağımsız değişkenin dağılım grafiği üzerinde tahmin edilen regresyon çizgisini çizdikten sonra, çizgiden belirgin sistematik bir sapma olup olmadığına bakılmalıdır. Bu yaklaşımın bir başka şekli, bağımsız değişkeni belirli aralıklara bölerek gruplandırmak ve her grup için bağımlı değişkenin ortalama değerine karşı grupların orta noktasını çizmektir. Bu yaklaşımın lojistik regresyona uygulanabilmesi için dikey eksen lojit ölçeğe dönüştürmek gerekmektedir. Böylece her bir grup için lojistik regresyonda, grup ortalamalarının lojitine karşı grubun orta noktası çizilir.

Sonuçta bulunan eğrinin şekline göre çizim değerlendirilir.

Bu işlem kolay ve doğru gözükse de, gerçekleştirilmesi bir çok işlemsel adımlar gerekmektedir. Verilerin gruplandırılması, her bir grup ortalamasının hesaplanması, grup ortalamalarının dönüştürülmesi ve bu verilerin bir paket programa girildikten sonra çizilmesi gerekmektedir. Bunun yanı sıra herhangi bir lojistik regresyon programıyla kolayca gerçekleştirilebilen benzer bir işlem vardır. İki farklı grup için lojitler arasındaki fark, gruplanmış değişkeni kategorik olarak kabul edip lojistik regresyon modelini kurduktan sonra elde edilen desen değişkenlerinden birisi için tahmin edilen katsayının değerine eşittir. Böylece gruplanmış bağımsız değişkenin, en küçük grubu referans grup olarak kategorik bir değişkenmiş gibi kabul edilmiş olunur. Bu aşamadan sonra, grupların orta noktasına karşı tahmin edilen katsayılar çizilir ve parametrik model için en mantıklı şekil seçilir. Çoğu uygulamalarda göz kararıyla sonuca ulaşmak yeterince kesin

olmaktadır. Bununla beraber, daha kesin bir karar vermek gerekirse o zaman ortogonal desen değişkenlerine başvurulabilir. Bunun içinde x 'in enazından sıralama ölçeğinde olması gerekir.

Lojistik regresyonda ölçek tanımlamak için daha az bilinen alternatif bir işlem Box-Tidwell dönüşümüdür. Box-Tidwell yaklaşımı $x \log(x)$ şeklinde bir terimini modele dahil eder. Eğer bu değişkenin katsayısı önemli olarak bulunursa, değişkenin lojitle lineer olmayan bir ilişki içerisinde olduğu anlaşılır. Fakat bu işlem lineerlikten küçük sapmaları belirlemede yetersiz kalabilmektedir. Kay ve Little (1987) sonuç değişkeninin grupları içinde birlikte değişkenin marjinal dağılımını inceleyerek uygun ölçeği bulabilmek için nasıl yardımcı olunabileceğini göstermiştir. Örneğin, eğer x farklı ortalama ve varyansla normal dağılıma sahipse, o zaman lojitte bir tane lineer (x) ve bir tane ikinci dereceden terimin (x^2) bulunması gerekli olacaktır. Eğer x , sonuç değişkeninin grupları içerisinde farklı parametreler ile beta dağılımına sahipse, x 'i doğru olarak modelleyebilmek için $\ell(x)$ ve $\ell(1-x)$ terimlerini modele dahil etmek gerekecektir. Eğer x gama dağılımına (sağa eğik) sahipse, o zaman x ve $\ell(x)$ terimlerinin modele dahil edilmesi gerekmektedir. Sonuç değişkeninin her bir düzeyindeki dağılıma kesin bir yaklaşım elde edebilmek için, her bir düzeyde yeterli sayıda gözlem olduğu zaman sürekli değişkenler için Kay ve Little (1987)'in yaklaşımı en yararlı yöntem olabilir.

Sürekli değişkenlerin hepsinin doğru ölçeğe sahip olduğundan emin olduktan sonra, modeldeki interaksiyonlar kontrol edilmelidir. Modeldeki herhangi iki değişken arasındaki interaksiyon, değişkenlerden birinin etkisinin diğerinin düzeyleri üzerinde sabit olmamasıyla belirlenir. İnteraksiyon terimi elde edildikten sonra modele dahil edilir ve daha sonra olabilirlik oran testi kullanılarak önemliliğine karar verilir. Önemlilikle vurgulanmak istenen interaksiyonun modele katkısının olup olmadığıdır. Örneğin, modele interaksiyon teriminin girmesi modeldeki değişkenlerin nokta tahminlerini değiştirmeden yalnızca tahmin edilen standart hataları arttırıyorsa, interaksiyon teriminin bir yararı yoktur. Genel olarak interaksiyon teriminin hem nokta hem de aralık tahminlerini değiştirebilmesi için en azından orta derecede bir istatistiksel öneme sahip olması gerekmektedir. Sonuç, olarak interaksiyon teriminin modele dahil edilip edilmemesi kararı istatistiksel olduğu kadar biyolojik esaslara da dayanmalıdır. Yani interaksiyon terimi aynı zamanda biyolojik açıdan da bir anlama sahip olmalıdır.

3.2.11. Uyum İyiliğinin Belirlenmesi

Oluşturulan modelin bağımlı değişkeni ne kadar iyi tanımladığının belirlenmesi uyum iyiliği testleri ile yapılır. Uyum iyiliği denemede gözlenen $\mathbf{y}' = [y_1, y_2, y_3, \dots, y_n]$ ile, kurulan modelden tahmin edilen $\hat{\mathbf{y}}' = [\hat{y}_1, \hat{y}_2, \hat{y}_3, \dots, \hat{y}_n]$ değerlerinin birbirine ne kadar benzediğinin bir ölçüsüdür. Modelin uyumlu olduğunu söyleyebilmek için,

- $\mathbf{y}'$ ve $\hat{\mathbf{y}}'$ vektörlerinin karşılıklı elemanları $(y_i, \hat{y}_i)$ arasındaki uzaklığı belirleyen ölçünün küçük olması,
- b) hata'nın $(e_i = y_i - \hat{y}_i)$, dağılışının sistematik olmaması gerekir.

Uyum iyiliği hakkında fikir veren Pearson ki-kare ve sapma (deviance) test istatistikleri yanında doğrudan uyum iyiliğini test eden testler de vardır (Hosmer ve Lemeshow, 1980; Lemeshow ve Hosmer,1982; Hosmer ve ark.,1989). Bu çalışmada yalnızca Hosmer ve Lemeshow (1980)'de sözü edilen Hosmer-Lemeshow'un $\hat{C}$ ve $\hat{H}$ testleri ele alınacaktır.

$\hat{C}$ testi:

$\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ değerleri hatırlanacağı gibi lojistik regresyon modelinin en çok olabilirlik parametre tahminleriydi ve “*son model*”de yer alıyordu. Ayrıca i -inci denek için $y=1$ olması olasılığının tahmini,

$$P(\mathbf{x}_i) = \frac{e^{\ell(\mathbf{x}_i)}}{1 + e^{\ell(\mathbf{x}_i)}} = \frac{e^{\beta_0 + \sum \beta_k x_{ikj}}}{1 + e^{\beta_0 + \sum \beta_k x_{ikj}}} \quad (3.46)$$

şeklindeydi ($i=1, 2, \dots, n$; $j=1, 2, \dots, r-1$; $k=1, 2, \dots, p$). Buna göre, denemede var olan n tane $(y_i, \mathbf{x}_i)$ gözlemi için $P(\mathbf{x}_i)$ olasılıkları hesaplanır. Küçükten büyüğe (ya da tersi) şeklinde büyüklük sırasına dizilir. Tüm denekler $P(\mathbf{x}_i)$ 'ye göre sıralandıktan sonra, $t > p+1$ olacak şekilde t tane gruba ayrılır. Genellikle önerilen $p+1 < 10$ ise *onluk* (decile) risk gruplarıdır. Gruplardaki denek sayıları toplam denek sayısının (n) grup sayısına (t) bölünmesi ile n/t şeklinde bulunur. Gruplardaki gözlem sayıları her zaman tam eşit çıkmayabilir. Bu durumda rastgele belirlenen bazı gruplarda 1 denek eksik ya da fazla olacak şekilde bir dengeleme yapılır. Örneğin 284 deneklik bir çalışmada onluk risk grupları ($t=10$) oluşturmak istenirse her bir gruptaki denek sayısı $284/10=28.4 \approx 28$ tane olacaktır. Bölümün sonucu tam sayı olmadığı için bazı gruplarda da (4 tanesinde) 29'ar gözlem bulunması gerekecektir.

Her hangi bir m -inci onluk grup (R_m) için gözlenen (g) ve beklenen (b) frekanslar bulunmalıdır. Herhangi bir m -inci onluk gruptaki (R_m için);

$y=1$ olan deneklerin *gözlenen* frekansı,

$$g_{1m} = \sum_{y_i \in R_m} y_i \quad (3.47a)$$

$y=0$ olan deneklerin *gözlenen* frekansı,

$$g_{0m} = \sum_{y_i \in R_m} (1 - y_i) \quad (3.47b)$$

$y=1$ olan deneklerin *beklenen* frekansı,

$$b_{1m} = \sum_{P(\mathbf{x}_i) \in R_m} P(\mathbf{x}_i) \quad (3.47c)$$

$y=0$ olan deneklerin *beklenen* frekansı,

$$b_{0m} = \frac{\sum [1 - P(x_i)]}{P(x_i) \in R_m} \quad (3.47d)$$

şeklinde hesaplanabilir. Buna göre gözlenen ve beklenen frekanslar Çizelge 3.11'deki gibi özetlenebilir.

Çizelge 3.11. Sabit Denek Sayılı Onlu Risk Grupları İçin Gözlenen ve Beklenen Frekanslar

İkil Bağımlı Y Değişkeninin Değeri	Onlu Risk Grupları (R_m)					Toplam
	R_1	R_2	R_3		R_{10}	
$y=1$	g_{11} b_{11}	g_{12} b_{12}	g_{13} b_{13}		g_{110} b_{110}	n_1
$y=0$	g_{01} b_{01}	g_{02} b_{02}	g_{03} b_{03}		g_{010} b_{010}	n_0
Toplam	$n/10$	$n/10$	$n/10$		$n/10$	n

Çizelge 3.11'deki beklenen ve gözlenen frekanslar kullanılarak uyum iyiliğinin testi için $\hat{C}$ test istatistiği,

$$\hat{C} = \sum_{m=1}^t \left[\frac{(g_{1m} - b_{1m})^2}{b_{1m}} + \frac{(g_{0m} - b_{0m})^2}{b_{0m}} \right] \quad (3.48)$$

şeklinde hesaplanabilir. $\hat{C}$ test istatistiği $t-2$ serbestlik dereceli χ^2 dağılışı göstermektedir.

Yine Lemeshow ve Hosmer (1982) tarafından bildirilen **ikinci bir** gruplandırma şekli $P(x_i)$ 'leri sabit aralıklı t tane gruba ayırmaktır. $P(x_i)$ olasılıkları $0 \leq P(x_i) \leq 1$ aralığında değiştiği için örneğin $t=10$ gruba bölünmek istenirse $1/t=1/10=0.1$ genişliğinde 10 tane risk grubu oluşturulabilir. Herhangi bir m -inci risk grubu (R_m) için beklenen ve gözlenen frekanslar 3.47a, b, c, d eşitliklerinde olduğu gibi hesaplanır ve sabit $P(x_i)$ aralıklı onlu risk grubu Çizelge 3.12'deki gibi oluşturulur.

Çizelge 3.12. Sabit $P(x_i)$ Aralıklı Onlu Risk Grupları İçin Gözlenen ve Beklenen Frekanslar

İkil Bağımlı Y Değişkeninin Değeri	Onlu Risk Grupları (R_m)					Toplam
	R_1	R_2	R_3		R_{10}	
	0-0.1	0.11-0.20	0.21-0.30		0.91-1.0	
$y=1$	g'_{11}	g'_{12}	g'_{13}		g'_{110}	n_1
	b'_{11}	b'_{12}	b'_{13}		b'_{110}	
$y=0$	g'_{01}	g'_{02}	g'_{03}		g'_{010}	n_0
	b'_{01}	b'_{02}	b'_{03}		b'_{010}	
Toplam	N_1	n_2	n_3		n_{10}	n

Çizelge 3.11 ve Çizelge 3.12'deki en önemli farklılığın risk gruplarına düşen denek sayılarında olduğuna dikkat edilmelidir. Çizelge 3.11'de her bir risk grubunda $n/10$ tane (yaklaşık eşit) denek varken Çizelge 3.12'deki her bir risk grubuna düşecek denek sayısı önceden bilinmez.

Sabit $P(x_i)$ aralıklı yöntemle risk grupları çizelgesi oluşturulduktan sonra Çizelge 3.12'den

$$\hat{H} = \sum_{m=1}^t \left[\frac{(g'_{1m} - b'_{1m})^2}{b'_{1m}} + \frac{(g'_{0m} - b'_{0m})^2}{b'_{0m}} \right] \quad (3.49)$$

şeklinde hesaplanır. Bu test istatistiği de $t-2$ serbestlik dereceli χ^2 dağılışı göstermektedir.

4. BULGULAR VE TARTIŞMA**Lojistik Regresyon Analizinin Çocuklarda Doğum Ağırlığı Gerçek Verilerine Uygulanması**

Diğer memelilerde olduğu gibi insanlarda da doğum ağırlığı ile, bebeğin yaşama gücü, hastalıklara karşı dayanıklılığı ve ileri dönemlerdeki fiziksel gelişimi arasında büyük bir ilişki vardır. Her şeyden önce normal ağırlıktaki bir bebeğin (2500 gr <doğum ağırlığı≤ 4500 gr) doğumu, anne karnındaki gelişiminin normal olduğunun bir göstergesidir.

Bebeğin doğum ağırlığını etkileyen bir çok faktör vardır. Bunlardan bazıları kontrol altına alınabilir faktörlerdir. Doğum ağırlığını etkileyen veya düşük doğum ağırlığına neden olan risk faktörlerinin, doğum ağırlığını hangi düzeylerde etkilediğinin bilinmesi gerekmektedir. Bu yolla kontrol altına alınabilir risk faktörleri dikkate alınarak, düşük doğum ağırlığı oranı en aza indirilebilir.

Analiz için kullanılan veri seti, K. Maraş ilindeki özel bir klinikten temin edilmiş olup, yaklaşık 1486 adet hasta kayıt formu (antenatal form) incelenmiş ve bunlardan 284'ü analiz için seçilmiştir (EK-12).

Aday risk faktörlerinden sürekli değişkenlerin düşük ve normal doğum ağırlıkları için ortalama ve standart sapmaları Çizelge 4.1'de verilmiştir. Kesikli değişkenlere göre deneklerin dağılımları ise Çizelge 4.2., 4.3., 4.4, 4.5., 4.6., 4.7.'de verilmiştir.

Çizelge 4.1. Sürekli Değişkenlerin Ortalama ve Standart Sapma Değerleri ($\bar{x} \pm S$)

Sürekli Değişkenler	Doğum Ağırlığı	
	DA >2500 kg	DA ≤2500 kg
SADA	62.868 ± 6.892	57.384 ± 7.414
YAS	25.733 ± 4.985	21.769 ± 4.617
HB	2.247 ± 0.973	2.230 ± 0.836
Denek sayısı	206	78

Çizelge 4.2. Deneklerin Sigara İçme Durumuna Göre Dağılımları

Sigara İçme Durumu	Doğum Ağırlığı	
	DA >2500 kg	DA ≤2500 kg
İçiyor	180	54
İçmiyor	26	24
Denek sayısı	206	78

Çizelge 4.3. Deneklerin Gebelik Durumuna Göre Dağılımları

Doğum Ağırlığı	Gebelik Sayısı			
	1.Gebelik	2.Gebelik	3.Gebelik	4.Gebelik
DA >2500 kg	76	86	40	4
DA ≤2500 kg	61	15	2	0
Denek sayısı	137	101	42	4

Çizelge 4.4. Deneklerin Daha Önce Doğum Yapıp Yapmama Durumuna Göre Dağılımları

Doğum Ağırlığı	Daha Önce Doğum	
	Yok	Var
DA >2500 kg	91	116
DA ≤2500 kg	65	12
Denek sayısı	156	128

Çizelge 4.5. Deneklerin Akrabalık Durumuna Göre Dağılımları

Akrabalık Durumu	Doğum Ağırlığı	
	DA >2500 kg	DA ≤2500 kg
Yok	180	26
Var	71	7
Denek sayısı	251	33

Çizelge 4.6. Deneklerin Hipertansiyona Göre Dağılımları

Değişken	Doğum Ağırlığı	
	DA >2500 kg	DA ≤2500 kg
Yok	193	15
Var	52	24
Denek sayısı	245	39

Çizelge 4.7. Deneklerin Bebek Cinsiyetine Göre Dağılımları

Cinsiyet	Doğum Ağırlığı	Doğum Ağırlığı
	DA >2500 kg	DA ≤2500 kg
Erkek	120	86
Kız	26	52
Denek sayısı	146	138

Çok değişkenli lojistik regresyon modeline girecek değişkenleri belirlemek amacıyla, aday değişkenlerin her biri için ayrı ayrı yapılan tek değişkenli basit lojistik regresyon analizi (Ek-4) sonuçları Çizelge 4.8’de özetlenmiştir.

Çizelge 4.8. Düşük Doğum Ağırlığı İle İlişkili Olabileceği Düşünülen Değişkenlerin Basit Lojistik Regresyon Analizi

Etki	$\hat{\beta}$	$SE(\hat{\beta})$	$\hat{\Omega}$	%95 Güven Aralığı	Ki-Kare	-2Log Olabilirlik	G	P
SABİT	-0.971	0.1329			53.36	333.885		
SIG	1.019	0.3229	2.772	(1.47, 5.22)	9.97	324.193	9.692	0.0016*
YAS	-0.181	0.0336	0.833	(0.78, 1.12)	29.34	296.783	37.10	0.0000*
DGS1	-1.526	0.3284	0.217	(0.14, 0.41)	21.61	289.595	44.29	0.0000*
DGS2	-2.871	0.7432	0.057	(0.02, 0.24)	14.93	289.595	44.29	0.0001*
AK	-0.3819	0.4483	0.682	(0.28, 1.64)	0.73	333.118	0.767	0.3942
D	-1.843	0.3347	0.158	(0.08, 0.30)	30.35	296.005	37.88	0.0000*
SADA	-0.116	0.0221	0.889	(0.85, 0.92)	27.81	300.264	33.62	0.0000*
C	1.026	0.2787	2.790	(1.61, 4.81)	13.56	319.639	14.25	0.0002*
HT	1.733	0.3634	5.659	(2.77, 11.5)	22.74	310.407	23.48	0.0000*
HB	0.144	0.1060	1.115	(0.93, 1.42)	1.87	331.995	1.89	0.1715

* $\alpha=0.10$ yanılma düzeyinde istatistiksel olarak önemli.

Çizelge 4.8’in birinci sütununda modelde yalnız o değişken bulunurken tahmin edilen eğim katsayısı ($\hat{\beta}$), ikinci sütununda tahmin edilen eğim katsayısının standart hatası ($SE(\hat{\beta})$), üçüncü sütununda tahmin edilen odds oranları ($\hat{\Omega}$), dördüncü sütununda tahmin edilen odds oranı için %95 güven aralığı, beşinci sütunda mevcut model için $-2 \log$ olabilirlik değerleri, altıncı sütunda eğim katsayısının sıfıra eşit olup olmadığını test eden oran test istatistiği (G), yedinci sütunda G istatistiğine ait P değerleri verilmiştir.

Kesikli bir değişken olan gebelik sayısının (GS) ikiden fazla düzeyinin olması nedeni ile, desen değişkeni kullanılarak tek değişkenli modele alınmıştır (Çizelge 4.9). Burada 4. gebeliğinde olan denek sayısı çok az ($n=4$) olduğu için üçüncü ve dördüncü gebelikler birleştirilmiştir.

Çizelge 4.9. GS değişkeni için desen değişkenlerinin oluşturulması ve kodlanması

Gruplar	Grup Adı	Desen Değişkenleri	
		DGS1	DGS2
GS=1	GS 1.Grup	0	0
GS=2	GS 2.Grup	1	0
GS=3 ve 4	GS 3.Grup	0	1

Olabilirlik oran testi sonucunda, Çizelge 4.8'in son sütununda verilen önem düzeyi 0.10'un altında bulunan değişkenler çok değişkenli modele girecek olan aday değişkenlerdir. Çizelge 4.8'de görüldüğü gibi HB ve AK değişkenleri istatistiki olarak önemsiz bulunduğu için çok değişkenli modele dahil edilmeyecektir. Tek değişkenli basit lojistik regresyon analizi sonunda önemli bulunan değişkenlerle oluşturulan çok değişkenli modelin analiz (Ek-5) sonuçları Çizelge 4.10'da verilmiştir.

Çizelge 4.10. Çoklu Lojistik Regresyon Analizi

Effect	$\hat{\beta}$	$SE(\hat{\beta})$	Ki-Kare	P
SABİT	3.1615	1.5451	4.20	0.0404*
SIG	1.4406	0.4228	11.61	0.0007*
YAS	-0.0745	0.0403	3.42	0.0645*
DGS1	-1.3053	0.6783	3.70	0.0543*
DGS2	-2.7419	1.0921	6.30	0.0120*
D	0.2549	0.7458	0.12	0.7325
SADA	-0.0483	0.0254	3.61	0.0575*
C	1.0745	0.3355	10.26	0.0014*
HT	1.6385	0.4745	11.92	0.0006*

* $\alpha=0.10$ yanılma düzeyinde istatistiksel olarak önemli.

Çizelge 4.10'daki çok değişkenli lojistik regresyon analizi sonucunda D değişkeninin artık istatistiksel olarak bir öneme sahip olmadığı görülmektedir. Bu durumda D (annenin daha önce doğum yapıp yapmadığı) değişkeninin modele katkısını test etmek amacıyla, D değişkenini içeren ve içermeyen modeller olabilirlik oran test istatistiği ile karşılaştırılacaktır. Çizelge 4.11'de D değişkeni model dışı bırakıldıktan sonra kalan değişkenlerle oluşturulan çok değişkenli (çoklu) lojistik regresyon analizi (Ek-6) sonuçları verilmiştir.

Çizelge 4.11. Doğum Durumu (D) Değişkeni İndirgendikten Sonra Geriye Kalan Değişkenler İçin Çoklu Lojistik Regresyon Analizi

Etkiler	$\hat{\beta}$	$SE(\hat{\beta})$	Ki-Kare	P
SABİT	3.1284	1.5395	4.13	0.0421*
SIG	1.4426	0.4232	11.62	0.0007*
YAS	-0.0722	0.0397	3.31	0.0687*
DGS1	-1.1203	0.3988	7.89	0.0050*
DGS2	-2.4958	0.8196	9.27	0.0023*
SADA	-0.0484	0.0254	3.63	0.0568*
C	1.0714	0.3352	10.22	0.0014*
HT	1.6068	0.4626	12.06	0.0005*

* $\alpha=0.10$ yanılma düzeyinde istatistiksel olarak önemli.

D değişkenini içeren ve içermeyen modelleri karşılaştıran *olabilirlik oran* test istatistiği,

$$G = 233.516 - 233.397 = 0.119$$

şeklinde hesaplanabilir. Bu değer, $SD = v_{\text{tam}} - v_{\text{ind}} = (8+1) - (7+1) = 1$ serbestlik dereceli ki-kare dağılımıyla karşılaştırılır. Hesaplanan G istatistiği, bir serbestlik dereceli ki-kare dağılımından küçük olduğu için ($0.119 < \chi^2 = 3.841$), D değişkeninin modele herhangi bir katkısı olmadığı sonucuna varılır ve D değişkeni model dışı bırakılır.

İnteraksiyon terimlerinin incelenmesine geçmeden önce, modelde bulunan sürekli değişkenlerin lojitle lineer bir ilişki içinde olup olmadıkları ve bu arada modele doğru ölçekle girip girmedikleri kontrol edilecektir. Çizelge 4.11'deki modelimizde, YAS ve SADA değişkenleri sürekli değişkenlerdir. Lojitle lineerliği test etmek için YAS değişkeni dört eşit gruba bölündükten sonra (kartiller kullanılarak) birinci grup referans grup olarak belirlenip, sürekli YAS değişkeni yerine üç desen değişkeni ile modele dahil edilecektir. Çizelge 4.12'de YAS değişkeni için desen değişkeninin oluşturulması ve kodlama şekli verilmiştir.

Çizelge 4.12. YAŞ'ın Kısmi Yöntemle Desen Değişkenine Dönüştürülmesi

Gruplar	Grup Adı	Desen Değişkenleri		
		DYAS1	DYAS2	DYAS3
$YAŞ \leq 20$	YAŞ 1.Grup	0	0	0
$20 < YAŞ \leq 24$	YAŞ 2.Grup	1	0	0
$24 < YAŞ \leq 27$	YAŞ 3.Grup	0	1	0
$27 < YAŞ$	YAŞ 4.Grup	0	0	1

Grup limitleri ile belirtilen aralıklar, karşılık gelen kod değerleri ile (0, 1, 2, 3) kodlandıktan sonra oluşturulan yeni yaş değişkeni (DYAS), YAS değişkeni yerine Çizelge 4.11'deki modele dahil edilir. Buna göre çok değişkenli lojistik regresyon analizi (Ek-7) sonuçları Çizelge 4.13'deki gibi bulunur.

Çizelge 4.13. Kartil Bölünmüş ve Desen Değişkenine Dönüştürülmüş Yaş Değişkeni İçin Çoklu Lojistik Regresyon Analizi

Etkiler	$\hat{\beta}$	$SE(\hat{\beta})$	Ki-Kare	P
SABİT	1.2251	1.5390	0.63	0.4260
SIG	1.4851	0.4433	11.22	0.0008*
DYAS1	-0.7950	0.4334	3.37	0.0666*
DYAS2	-1.0330	0.5266	3.85	0.0498*
DYAS3	-1.7780	0.5729	9.63	0.0019*
DGS1	-0.8734	0.4081	4.58	0.0323*
DGS2	-2.2440	0.8370	7.19	0.0073*
SADA	-0.0347	0.0259	1.80	0.1799
C	1.0873	0.3429	10.05	0.0015*
HT	1.5538	0.4654	11.15	0.0008*

* $\alpha=0.10$ yanılma düzeyinde istatistiksel olarak önemli.

Eğer lojit, YAS değişkeni ile lineer ise, bu durumda tahmin edilen katsayılar da lineer artan ya da azalan bir eğilim olması beklenir. YAS için kartil analizi sonunda, 4'üncü grubun 3'üncü gruptan daha düşük odds oranına sahip olduğu, 3'üncü grubun odds oranının ise 2'inci grubun odds oranından daha düşük olduğu (azalan bir eğim; 0.45, 0.36, 0.17) görülür (Çizelge 4.14). YAS sürekli değişkenini içeren modele YAS[log(YAS)] değişkeninin eklenmesi sonucunda, YAS[log(YAS)] değişkeninin önemli bir eğime sahip olmaması YAS değişkeninin lojitle lineer olduğu varsayımını desteklemektedir. Bu nedenle YAS değişkeninin modele sürekli bir değişken olarak girmesinin bir sakıncası olmadığına karar verilir.

Çizelge 4.14. Çizelge 4.15. YAS Değişkeni İçin , Katsayı Tahminleri, Standart Hata, Ki-Kare ve Odds Oranları

Kartil	1	2	3	4
Kesim noktası	$YAS \leq 20$	$20 < YAS \leq 24$	$24 < YAS \leq 27$	$27 < YAS$
N	82	68	48	86
$\hat{\beta}$	-	-0.7950	-1.0330	-1.7780
$SE(\hat{\beta})$	-	0.4334	0.5266	0.5729
Ki-Kare	-	3.37	3.85	9.63
$\hat{\Omega}$	-	0.45	0.36	0.17

Aynı işlemler diğer sürekli değişken SADA için de yapılarak, lojitle lineer olup olmadığı test edilmeli ve modele nasıl gireceğine karar verilmelidir. SADA değişkeni Çizelge 4.15'deki gibi dört eşit gruba ayrıldıktan sonra, desen değişkeni (DSADA) kullanılarak Çizelge 4.11'de oluşturulan modele dahil edilecektir.

Çizelge 4.15. SADA Değişkeninin Desen Değişkenine Dönüştürülmesi

Gruplar	Grup Adı	Desen Değişkenleri		
		DSADA1	DSADA2	DSADA3
SADA≤55	SADA 1.Grup	0	0	0
55<SADA≤60	SADA 2.Grup	1	0	0
60<SADA≤65	SADA 3.Grup	0	1	0
65<SADA	SADA 4.Grup	0	0	1

Son adet dönemi ağırlık değişkeni SADA, yerine desen değişkenleri Çizelge 4.11'deki modele dahil edildiğinde çoklu lojistik analizi (Ek-8) sonuçları Çizelge 4.16'daki gibi elde edilir.

Çizelge 4.16. Kartil Bölünmüş ve Desen Değişkenine Dönüştürülmüş SADA Değişkeni İçin Çoklu Lojistik Regresyon Analizi

Etkiler	$\hat{\beta}$	$SE(\hat{\beta})$	Ki-Kare	P
SABİT	0.9634	0.9144	1.11	0.2921
SIG	1.3688	0.4273	10.26	0.0014*
YAS	-0.0800	0.0395	4.10	0.0428*
GS1	-1.0969	0.4008	7.49	0.0062*
GS2	-2.4211	0.8209	8.70	0.0032*
DSADA1	-0.6565	0.4479	2.15	0.1428
DSADA2	-0.6167	0.4519	1.86	0.1724
DSADA3	-1.1032	0.5094	4.69	0.0303*
C	1.0850	0.3351	10.48	0.0012*
HT	1.5168	0.4653	10.63	0.0011*

* $\alpha=0.10$ yanılma düzeyinde istatistiksel olarak önemli.

Son adet dönemi ağırlık için kartil analizi sonunda, 2'inci ve 3'üncü grubun odds oranlarının yaklaşık olarak birbirine eşit olduğu, 4'üncü grubun odds oranını ise bir miktar düştüğü görülmektedir (Çizelge 4.17). Tahmin edilen katsayılar giderek artan veya azalan bir eğilim izlemediğinden (0.518, 0.539, 0.332) dolayı desen değişkeni lojitle lineer değildir. SADA değişkeni modele desen değişkeni kullanılarak dahil edilecektir ve bu yeni değişken "DSADA" olarak gösterilecektir.

Çizelge 4.17. SADA Değişkeni İçin, Katsayı Tahminleri, Standart Hata, Ki-Kare ve Odds Oranları

Kartil	1	2	3	4
Kesim noktası	SADA≤55	55<SADA≤60	60<SADA≤65	65<SADA
N	65	68	70	65
$\hat{\beta}$	-	-0.6565	-0.6167	-1.1032
$SE(\hat{\beta})$	-	0.4479	0.4519	0.5094
Ki-Kare	-	2.15	1.86	4.69
$\hat{\Omega}$	-	0.518	0.539	0.332

Her üç gurubun odds değerlerinin birbirine çok yakın olması nedeniyle, her üç grup birleştirilecek (Hosmer ve Lemeshow, 1989) ve kartil analizinde kullanılan kesim noktalarının yerine Çizelge 4.18’de verilen kesim noktaları kullanılacaktır.

Çizelge 4.18. SADA Değişkeninin Kesikli Değişken Olarak Kodlanması

Gruplar	Kod	DSADA
SADA ≤55	1	1
SADA >55	0	0

Yeni oluşturulan desen değişkeni (DSADA), SADA değişkeni yerine Çizelge 4.11’deki modele dahil edilir. Oluşturulan yeni çoklu lojistik regresyon modelinin analiz (Ek-9) sonuçları Çizelge 4.19’da verilmiştir.

Çizelge 4.19. DSADA Değişkeni Kullanılarak Elde Edilen Katsayı Tahminleri , Standart Hata, Ki-Kare ve Önemlilikleri

Etkiler	$\hat{\beta}$	$SE(\hat{\beta})$	Ki-Kare	P
SABİT	0.3089	0.8943	0.12	0.7298
SIG	1.3921	0.4219	10.89	0.0010*
YAS	-0.0486	0.0382	4.92	0.0266*
DGS1	-1.1033	0.3987	7.66	0.0057*
DGS2	-2.4919	0.8175	9.29	0.0023*
DSADA	0.7526	0.3674	4.20	0.0405*
C	1.0903	0.3343	10.64	0.0011*
HT	1.5036	0.4632	10.54	0.0012*

* $\alpha=0.10$ yanılma düzeyinde istatistiksel olarak önemli.

C ve HT kesikli değişkenleri iki sınıflı olduğu için ayrıca incelemeye gerek yoktur. Böylece ana etkilerin incelenmesi tamamlanmış ve hangi ana etkilerin ne şekilde (kesikli mi, desen değişkenleri ile mi...vs) modele gireceğine karar verilmiştir.

Ana etkiler modelini (Çizelge 4.19) elde ettikten sonra, modeldeki değişkenler arasındaki interaksiyonlar incelemeye alınır. Bu aşamada her bir interaksiyon teriminin ana etkiler modeline katkısı incelenecektir. Bu nedenle muhtemel her bir interaksiyon terimi için (Ek-10) $-2 \log$ olabilirlik değeri, *olabilirlik oran* test istatistiği (G), serbestlik derecesi ve P- değeri Çizelge 4.20’de verilmiştir.

Çizelge 4.20. Ana Etkiler Modeline İnteraksiyon Terimlerinin Tek Tek Eklenmesiyle Elde Edilen Yeni Modellerin $-2 \log$ Olabilirlik, Olabilirlik Oran Testi (G) ve P- Değerleri

İnteraksiyon	-2 Log Olabilirlik	S.D.	G	P
Ana Etkiler Modeli	233.140			
SIG×YAS	232.246	1	0.894	0.3390
SIG×DGS1	233.115	2	0.025	0.8722
SIG×DSADA	226.644	1	6.496	0.0111*
SIG×C	232.634	1	0.506	0.4759
SIG×HT	232.9641	1	0.176	0.9331
YAS×DGS1	231.812	2	1.328	0.2582
YAS× DSADA	225.516	1	7.624	0.0056*
YAS×C	230.956	1	2.184	0.1417
YAS×HT	232.850	1	0.29	0.5793
DGS1×DSADA	232.338	2	0.802	0.3693
DGS1×C	233.116	2	0.024	0.8773
DGS1×HT	232.444	2	0.696	0.4026
DSADA×C	232.980	1	0.16	0.6887
DSADA×HT	232.466	1	0.674	0.4071
C×HT	230.658	1	2.482	0.1144
YAS×DSADA + SIG×DSADA	219.265	2	13.875	0.0196*

* $\alpha=0.10$ yanılma düzeyinde istatistiksel olarak önemli.

Çizelge 4.20’de görüldüğü gibi yalnızca iki interaksiyon terimi önemli bulunmuştur ($P<0.10$). Bunlar YAS×DSADA ve SIG×DSADA interaksiyonlarıdır. Çizelge 4.21’de önemli bulunan interaksiyon terimlerini içeren yeni model için çoklu lojistik regresyon analizi (Ek-11) sonuçları verilmiştir. İki interaksiyon teriminin modele dahil edilmesiyle yaş, son adet dönemi ağırlık ve sigara içme durumunun, düşük doğum ağırlığı üzerine etkisinin daha iyi tanımlanabildiği söylenebilir.

Çizelge 4.21. Önemli İnteraksiyon Terimlerini İçeren Model İçin Çoklu Lojistik Regresyon Analizi

Etkiler	$\hat{\beta}$	$SE(\hat{\beta})$	Ki-Kare	P
SABİT	2.1536	1.1735	3.37	0.0665*
SIG	2.4229	0.5745	17.78	0.0000*
YAS	-0.1758	0.0533	10.89	0.0010*
DGS1	-1.1256	0.4184	7.24	0.0071*
DGS2	-2.4266	0.8877	7.47	0.0063*
DSADA	-3.5627	1.8314	3.78	0.0517*
C	1.2425	0.3510	12.53	0.0004*
HT	1.4287	0.4923	8.42	0.0037*
YAS×DSADA	0.2125	0.0785	7.32	0.0068*
SIG×DSADA	-2.1255	0.8540	6.19	0.0128*

* $\alpha=0.10$ yanılma düzeyinde istatistiksel olarak önemli.

Çizelge 4.21'deki ana ve interaksiyon etkilerinden oluşan “son model” den herhangi bir denek için *lojistik regresyon modeli tahmini*,

$\hat{\ell}$ (SIG, YAS, DGS1, DGS2, DSADA, C, HT)

$$\begin{aligned}
 &= \hat{\beta}_0 + \hat{\beta}_1 (\text{SIG}) + \hat{\beta}_2 (\text{YAS}) + \hat{\beta}_3 (\text{DGS1}) + \hat{\beta}_4 (\text{DGS2}) + \hat{\beta}_5 (\text{DSADA}) \\
 &+ \hat{\beta}_6 (\text{C}) + \hat{\beta}_7 (\text{HT}) + \hat{\beta}_8 (\text{YAS} \times \text{DSADA}) + \hat{\beta}_9 (\text{SIG} \times \text{DSADA}) \quad (4.1)
 \end{aligned}$$

şeklinde yazılabilir. Burada,

$$\text{SIG} \in \{0, 1\}$$

$$\text{YAS} \in \{13 < \text{YAS} < 43\}$$

$$\text{DSADA} \in \{0, 1\}$$

$$\text{C} \in \{0, 1\}$$

$$\text{HT} \in \{8.2 < \text{HT} < 13.9\}$$

$$\text{DGS} \in \{0, 1, 2\}$$

şeklindedir. Son Modelde SIG, DSADA ve YAS ana etkilerinin interaksiyon etkileri de söz konusudur. Bunların dışındaki DGS1, DGS2, C, HT ana etkilerinin ise herhangi bir interaksiyona karışmadığı görülmektedir. Modeldeki SIG, DSADA ve YAS değişkenleri dışındaki etkiler $\hat{\beta}z$ ile gösterildiğinde ifade,

$\hat{\ell}$ (DSADA, SIG, YAS, z)

$$= \hat{\beta}_0 + \hat{\beta}_1(\text{SIG}) + \hat{\beta}_2(\text{YAS}) + \hat{\beta}_5(\text{DSADA}) + \hat{\beta}_8(\text{YAS} \times \text{DSADA}) + \hat{\beta}_9(\text{SIG} \times \text{DSADA}) + \hat{\beta}_z \quad (4.2)$$

şeklini alır.

Daha önce de ifade edildiği gibi, tahmin edilen *lojistik regresyon modelinden* çıkarılacak yorumlar, modeldeki risk faktörleri için tahmin edilen *odds oranları* ile yapılır. İnteraksiyon halinde olmayan DGS1, DGS2, C ve HT risk faktörleri için *odds oranları*, tahmin edilen katsayıların üsteli (exponansiyeli) alınarak elde edilir. Örneğin C değişkeni için *odds oranı*,

$$\hat{\Omega} = \exp(1.2425) = 3.464$$

şeklindedir. Bu katsayı, doğacak bebeğin kız olması durumunda düşük doğum ağırlığında olması olasılığının, erkek olması durumuna göre 3.464 kat daha yüksek olduğunu gösterir. Aynı şekilde DGS1 değişkeni için *odds oranı*,

$$\hat{\Omega} = \exp(-1.1256) = 0.324$$

şeklindedir. Buna göre herhangi bir kadında ikinci gebelikte 2500 gr'dan daha yüksek ağırlıkta (GS1'e ait katsayı işareti negatif) doğum yapma olasılığı, birinci gebeliğe göre 0.324 kat daha fazladır. Hipertansiyon durumu için *odds oranı* ise,

$$\hat{\Omega} = \exp(1.4287) = 4.173$$

olup, yüksek tansiyonu olan bir kadının sağlıklı bir kadına göre, düşük ağırlıkta bebek doğurma olasılığının 4.173 kat daha fazla olacağı söylenebilir.

YAS, DSADA ve SIG değişkenleri interaksiyon içinde olduklarından dolayı, odds oranları elde edilirken, DSADA değişkeni için odds oranı hem sigara içen hem de içmeyenler için, YAS değişkeni kontrol altında tutularak tahmin edilecektir (Çizelge 4.22). Çizelge 4.23'deki beşinci odds oranı, son adet dönemi ağırlığı 55 kg ve daha düşük olan ve sigara içen kadınların (DSADA=1, SIG=1), son adet dönemi ağırlığı 55 kg'dan yüksek olan ve sigara içmeyen (DSADA=0, SIG=0) kadınlara göre YAS değişkeni kontrol altında tutulduktan sonra tahmin edilen odds oranıdır.

DSADA ve SIG için bütün kombinasyonlara ait *lojitler* ve *lojit farklar* Çizelge 4.22'deki gibidir.

Çizelge 4.22. SIG ve DSADA'nın Dört Farklı Kombinasyonu İçin Lojit ve Lojit Farkları

SIG	DSADA		Lojit Fark
	0	1	
0	$\hat{\beta}_0 + \hat{\beta}_2(YAS) + \hat{\beta}z$	$\hat{\beta}_0 + \hat{\beta}_2(YAS) + \hat{\beta}_5 + \hat{\beta}_8(YAS) + \hat{\beta}z$	$\hat{d}_1 = \hat{\beta}_5 + \hat{\beta}_8(YAS)$
1	$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2(YAS) + \hat{\beta}z$	$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2(YAS) + \hat{\beta}_5 + \hat{\beta}_8(YAS) + \hat{\beta}_9 + \hat{\beta}z$	$\hat{d}_5 = \hat{\beta}_5 + \hat{\beta}_8(YAS) + \hat{\beta}_9$
Lojit Fark	$\hat{d}_3 = \hat{\beta}_1$	$\hat{d}_4 = \hat{\beta}_1 + \hat{\beta}_9$	$\hat{d}_5 = \hat{\beta}_1 + \hat{\beta}_5 + \hat{\beta}_8(YAS) + \hat{\beta}_9$

Çizelgedeki herhangi bir hücrede yer alan lojit, o hücrenin kod değeri örneğin 1. hücre için SIG=0 ve DSADA=0 (4.2) sayılı *lojit model tahmin eşitliğinde* yerine koyularak elde edilir. 1. hücre için *lojit tahmini*,

$$\hat{\ell}(\text{SIG}, \text{YAS}, \text{DSADA}, z)$$

$$= \hat{\beta}_0 + \hat{\beta}_1(0) + \hat{\beta}_2(YAS) + \hat{\beta}_5(0) + \hat{\beta}_8(YAS(0)) + \hat{\beta}_9(0(0)) + \hat{\beta}z$$

$$= \hat{\beta}_0 + \hat{\beta}_2(YAS) + \hat{\beta}z$$

ve hemen altındaki hücre için SIG=1, DSADA=0 olduğundan *lojit tahmini*,

$$\hat{\ell}(\text{DSADA}, \text{SIG}, \text{YAS}, z)$$

$$= \hat{\beta}_0 + \hat{\beta}_1(1) + \hat{\beta}_2(YAS) + \hat{\beta}_5(0) + \hat{\beta}_8(YAS(0)) + \hat{\beta}_9(1(0)) + \hat{\beta}z$$

$$= \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2(YAS) + \hat{\beta}z$$

şeklinde elde edilir. Her hücre için *lojit tahminler* hesaplandıktan sonra *lojit farklar* bulunur. Örneğin, DSADA=0 grubundaki deneklerde SIG=1 ve SIG=0 için *lojit fark*,

$$\hat{d}_3 = (\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2(YAS) + \hat{\beta}z) - (\hat{\beta}_0 + \hat{\beta}_2(YAS) + \hat{\beta}z)$$

$$\hat{d}_3 = \hat{\beta}_1$$

şeklinde elde edilir. Bu şekilde bütün *lojit farklar* alındıktan sonra Çizelge 4.23'deki gibi özetlenebilir ve *odds oranı* tahminleri elde edilebilir.

Çizelge 4.23. SIG ve DSADA İçin Lojit Farklar ve Odds Oranı Tahminleri
($\hat{d}_1$, $\hat{d}_2$ ve $\hat{d}_5$; 35 Yaşlı Denekler İçin Hesaplanmıştır.)

Arasında	Lojit Fark	$\hat{\Omega}$
DSADA=0 Grubu içinde SİG için	$\hat{d}_3 = 2.4229$	11.28
DSADA=1 Grubu içinde SİG için	$\hat{d}_4 = 2.4229 - 2.1255 = 0.2974$	1.346
SIG =0 Grubu içinde DSADA için	$\hat{d}_1 = -3.5627 + 0.2125 \times 35 = 3.875$	48.17
SIG =1 Grubu içinde DSADA için	$\hat{d}_2 = -3.5627 - 2.1255 + 0.2125 \times 35 = 1.749$	5.75
DSADA + SIG için	$\hat{d}_5 = -3.5627 + 2.4229 - 2.1255 + 0.2125 \times 35 = 4.172$	64.85

Yukarıdaki Çizelgede verilen 1'inci ve 2'inci satırlar sırasıyla şu şekilde yorumlanabilir; son adet dönemi ağırlığı 55 kg'dan yüksek olan kadınlar arasında sigara içenlerin düşük ağırlıkta doğum yapma olasılığının, sigara içmeyenlere göre 11.28 kat daha fazla olduğunu söyleyebiliriz. Buna karşın, son adet dönemi ağırlığı 55 kg ve daha düşük olan kadınlar arasında sigara içenlerin düşük ağırlıkta doğum yapma olasılığının, sigara içmeyenlere göre 1.346 kat daha fazla olduğu söylenebilir.

Benzer şekilde 3'üncü ve 4'üncü satırlar; sigara içmeyen bayanlar arasında son adet dönemi ağırlığı 55 kg ve daha düşük olanların düşük ağırlıkta doğum yapma olasılığının, 55 kg'dan yüksek olanlara göre 48.17 kat daha fazla olduğunu söyleyebiliriz. Buna karşın, sigara içen bayanlar arasında son adet dönemi ağırlığı 55 kg ve daha düşük olanların düşük ağırlıkta doğum yapma olasılığının, 55 kg'dan yüksek olanlara göre 5.75 kat daha fazla olduğu söylenebilir.

Modelin Uyum İyiliği

(4.1) eşitliğinde verilen ve *son modelden* elde edilen herhangi bir denek için *lojit tahmininden*, sırasıyla her bir denek için *lojit tahmini son modeldeki* (Çizelge 4.21) ana etkiler eşitlikte yerine koyularak elde edilir. Son modelde yer alan ana etkiler (SİG, YAŞ, DGS, SADA, C, HT, YAŞ×DSADA, SİG×DSADA) ve her bir deneye ait *lojit tahmini*, $\hat{\ell}(x_i)$, Ek-14'de verilmiştir. Daha sonra her bir denek için

$$P(\mathbf{x}_i) = \frac{e^{\hat{\ell}(\mathbf{x}_i)}}{1 + e^{\hat{\ell}(\mathbf{x}_i)}} \text{ eşitliği kullanılarak sırası ile } P(\mathbf{x}_i) \text{ ve } 1 - P(\mathbf{x}_i) \text{ değerleri elde edilir.}$$

$P(\mathbf{x}_i)$ 'lere göre küçükten büyüğe doğru sıralanır.

Onlu risk grubu oluşturmak için $n/t=284/10=28.4 \cong 28$ 'er (bazıları da 29'ar) deneklik gruplar oluşturulur. 3.47a, b, c, d nolu eşitliklerden g_{1m} , g_{0m} , b_{1m} , b_{0m} gözlenen ve beklenen frekansları elde edilir. Örneğin birinci gruptaki $y=1$ olan deneklerin *gözlenen* frekansı (3.47a), 0, *beklenen* frekansı ise birinci gruptaki $P(\mathbf{x}_i)$ 'lerin toplamı (3.47c) yani , 0.267'dir (Ek-14). Benzer şekilde birinci gruptaki $y=0$ olan deneklerin *gözlenen* frekansı (3.47b), 28, *beklenen* frekansı ise birinci gruptaki $1-P(\mathbf{x}_i)$ 'lerin toplamı (3.47d) yani , 27.873'tür (Ek-14). Bu şekilde her bir onlu risk grubuna ait gözlenen ve beklenen frekanslar elde edildikten sonra Çizelge 4.24'deki gibi düzenlenir.

Çizelge 4.24. Sabit Denek Sayılı Onlu Risk Grupları İçin Gözlenen ve Beklenen Frekanslar

İkil Bağımlı Y Değişkeninin Değeri	Onlu Risk Grupları (R_m)										Σ
	1	2	3	4	5	6	7	8	9	10	
$y=1$	0 0.267	1 0.543	0 0.986	0 2.140	5 4.410	7 6.470	9 8.995	18 12.350	18 17.679	20 24.160	78 78
$y=0$	28 27.873	28 28.457	28 27.310	29 26.850	23 23.586	22 22.526	19 19.000	11 16.140	10 10.303	8 3.955	206 206
Σ	28	29	28	29	28	29	28	29	28	28	284

Çizelgeden (3.46) eşitliğine göre,

$$\hat{C} = \frac{(0 - 0.267)^2}{0.267} + \frac{(28 - 27.873)^2}{27.873} + \dots + \frac{(20 - 24.160)^2}{24.160} + \frac{(8 - 3.955)^2}{3.955}$$

$$= 13.21413$$

bulunur. $\hat{C} = 13.21413 < X_{t-2, \alpha}^2 = X_{8, 0.05}^2 = 15.507$ olduğundan oluşturulan son modelin verilere iyi uyduğu söylenebilir.

Eğer modelin uyum iyiliği, sabit $P(x_i)$ aralıklı onlu risk grubu oluşturarak $\hat{H}$ testi ile test edilmek istenirse, yine aynı şekilde lojit tahmini, $\hat{\ell}(x_i)$, $P(x_i)$ ve $1 - P(x_i)$ değerleri elde edilir ve $P(x_i)$ 'lere göre küçükten büyüğe doğru sıralanır. $1/t=1/10=0.1$ genişliğinde 10 tane risk grubu oluşturulur (Ek-14). Sabit $P(x_i)$ aralıklı onlu risk grupları Çizelge 4.25'deki gibi düzenlenir.

Çizelge 4.25. Sabit $P(x_i)$ Aralıklı Onlu Risk Grupları İçin Gözlenen ve Beklenen Frekanslar

İkil Bağımlı Y Değişkeninin Değeri	Onlu Risk Grupları (R_m)										Σ
	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	
	0-.1	.11-.20	.21-.30	.31-.40	.41-.50	.51-.60	.61-.70	.71-.80	.81-.90	.91-1	
$y=1$	1 3.493	5 5.258	9 7.705	11 9.317	10 8.368	7 5.797	10 11.602	6 7.319	14 14.029	4 5.653	77 79
$y=0$	109 106.5	29 28.74	24 25.295	16 17.68	9 10.63	4 5.202	8 6.3977	4 2.688	2 1.971	2 0.347	207 205
Σ	110	34	33	27	19	11	18	10	16	6	284

Çizelgeden (3.49) eşitliğine göre,

$$\hat{H} = \frac{(1-3.493)^2}{3.493} + \frac{(109-106.5)^2}{106.5} + \dots + \frac{(4-5.653)^2}{5.653} + \frac{(2-0.347)^2}{0.347}$$

$$= 13.355$$

bulunur. $\hat{H}$ testi ile de oluşturulan son modelin verilere uyduğu ($\hat{H} = 13.355 < X^2_{t-2,\alpha} = X^2_{8,0.05} = 15.507$) söylenebilir.

5. SONUÇLAR VE ÖNERİLER

Sonuç olarak, doğacak bebeğin kız olması durumunda düşük doğum ağırlığında olması olasılığının, erkek olması durumuna göre 3.464 kat daha yüksek, ikinci gebelikte 2500 gr'dan daha yüksek ağırlıkta doğum yapma olasılığı, birinci gebeliğe göre 0.324 kat daha fazla ve yüksek tansiyonu olan bir kadının sağlıklı bir kadına göre, düşük ağırlıkta bebek doğurma olasılığının 4.173 kat daha fazla olacağı söylenebilir. Bu duruma göre annenin gebelik sayısı önemli bir risk faktörü olarak görülmezken, bebeğin cinsiyeti ve yüksek tansiyon durumu önemli bir risk faktörü olarak görülmektedir.

Benzer şekilde, son adet dönemi ağırlığı 55 kg'dan yüksek olan kadınlar için sigaranın önemli bir risk faktörü ($\hat{\Omega}=11.28$) olduğunu, son adet dönemi ağırlığı 55 kg ve daha düşük olan kadınlar için sigaranın çok önemli bir risk faktörü ($\hat{\Omega}=1.346$) olmadığını söyleyebiliriz. Diğer taraftan, sigara içmeyen kadınlar için son adet dönemi ağırlığın 55 kg ve daha düşük olmasının önemli bir risk faktörü olduğunu ($\hat{\Omega}=48.17$), sigara içen kadınlar için son adet dönemi ağırlığın 55 kg'dan yüksek olmasının ise önemli bir risk faktörü ($\hat{\Omega}=5.75$) olmadığını söyleyebiliriz.

Lojistik regresyon, özellikle normallik varsayımı gerektirmemesi, bağımsız değişkenlerin hem kesikli hem de sürekli olabilmesi, esnek ve kolay yorumlanabilir olması gibi avantajları nedeniyle ayırım, kümeleme, çapraz tablolar, ana bileşenler gibi analizlere alternatif olmaktadır. Burada, yalnız ikil (dichotomous) bağımlı değişken durumu incelenmesine rağmen ikiden fazla sınıflı (polychotomous) durum için de kullanılabilir.

Parametrelerin tahmini en çok olabilirlik yöntemi ile ve iteratif bir yolla yapıldığından elde çözümlerine değinilmemiştir. Çünkü özellikle iki ve daha fazla bağımsız değişken içeren çoklu lojistik regresyon modelinin katsayılarının tahmini oldukça zor ve karmaşıktır. Bu amaçla en çok olabilirlik tahmini yapabilen istatistik paket programlar kullanılabilir.

Tek değişkenli lojistik regresyon analizinde, istatistiksel olarak önemsiz bulunan bir değişkenin model dışı bırakılmasına karar verilmeden önce, aynı zamanda biyolojik olarak da önemsiz olup olmadığı araştırılmalıdır. İstatistiksel olarak önemsiz bulunan bir değişkenin, *biyolojik olarak* önemli olduğu biliniyor ise değişken çoklu lojistik regresyon modeline alınmalıdır.

Dikkat edilmesi gereken diğer bir husus ta, değişkenin lojitle lineer olmadığı durumlarda, değişken modele alınırken kullanılacak kesim noktalarının belirlenmesidir. Kesim noktaları belirlenirken öncelikle konu uzmanlarının görüşlerine başvurulmasında yarar vardır.

- AKKAYA, Ş., PAZARLIOĞLU, M.V., 1998. *Ekonometri. Erkam matbaacılık, İzmir, 519s.*
- ALHO, J.M., 1990. Logistic regression in capture-recapture models. *Biometrics.*, 46: 623-635.
- ALİ, M., KHAN, A.H., 1989. On order statistics from the log-logistic distribution. *Jou. of Statis. Plan.*, 17: 103-108.
- ANONİM, 1988. SAS Guide for Personal Computers, Version 6.xx., SAS Institute Inc., NY.
- BAŞARIR, G., 1990. Çok değişkenli verilerde ayımsama sorunu ve lojistik regresyon analizi. H.Ü. Doktora Tezi, Ankara, 122s (yayınlanmamış).
- BAZZAZ, F.A., HARPER, J.L., (1977). Demographic analysis of the growth of *linum usitatissimum*. *New Phytologist*, 78:193-208.
- BERGER, P.J., CUBAS, A.C., KOEHLER, K.J., HEALEY, M.H., 1992. Factors affecting dystocia and early calf mortality in angus cows and heifers. *Journal of animal Science.*, 70: 1775-1796.
- BERNING, L.M., SHOOK, G.E., 1992. Prediction of mastitis using milk somatik cell count, N-acetyl-beta-D-glucosaminidase, and lactose. *Journal of Dairy Science.*, 75: 1840-1848.
- BOER, A., WELLER, J.I., GIPSON, T.A., GROSSMAN, M., 1989. Multiphasic analysis of milk and fat yield curves of israeli holsteins. *J. Dairy sci.*, 72: 2143-2152.
- BONNEY, G.E., 1987. Logistic regression for dependent binary observations. *Biometrics.*, 43:951-973.
- BRESLOW, N.E., CAIN, K.C., 1988. Logistic regression for two-stage case-control data. *Biometrics.*, 75:11-20.
- BUESCHER, P.A., LARSON, L.C., NELSON, M.D., LENIHAN, A.J., 1993. Prenatal WIC participation can reduce low birth weight and newborn medical costs: a cost benefit analysis of wic participation in North Carolina. *Journal of the American Dietetic Association (USA)*. 93:163-166.
- CORROL, R.J., WAND, M.P., 1990. Semiparametric estimation in logistic measurement error models. *J. R. Statist. Soc.*, 53: 573-585.
- CONSTABLE, G.A., GLAESON, A.C., (1977). Growth and distribution of dry matter in cotton. *Australian Journal of Agricultural Research*, 28:249-256.
- CORNFIELD, J., 1962. Joint dependence of the risk of coronary heart disease on serum cholesterol and systolic blood pressure: A discriminant function analysis. *Federation Proceedings.*, 21: 58-61.
- COX, B.D., WHICHELOW, M.J., PREVOST, A.T., 1998. The development of cardiovascular disease in relation to anthropometric indices and hypertension in British adults. *International Journal of Obesity*. 22: 97330-10966.
- CHATFIELD, C., J. COLLINS, A., 1992. *Introduction to Multivariate Analysis. Chapman&Hall, London, 245s.*
- DALE, R.F., COELHO, D.T., GALLO, K.P., (1980). Prediction of daily green leaf area index for corn. *Agronomy Journal*, 72:999-1005.
- DIETZ, A.B., COHEN, N.D., TIMMS, L., KEHRLI, M.E., 1997. Bovine lymphocyte antigen class II alleles as risk factors for high somatic cell counts in milk of lactating dairy cows. *Journal of Dairy Science.*, 80: 406-412.

- DOMECQ, J.J., SKIDMORE, A.L., LLOYD, J.W., KANEENE, J.B., 1997. Relationship between body condition scores and conception at first artificial insemination in a large dairy herd of high yielding Holstein cows. *Journal of Dairy Science.*, 80: 113-120.
- DOĞAN, S., 1985. Zooteknide kategorik verilerin log-linear modeller yardımı ile analizi. Ç.Ü. Yüksek lisans tezi, Adana, 172s,(yayınlanmamış).
- DRESCHER, K., SCHILL, W., 1991. Attributable risk estimation from case-control data via logistic regression. *Biometrics.*, 47:1247-1256.
- DUFFY, D.E., 1990. On continuity-corrected residuals in logistic regression. *Biometrika.*, 77:287-293.
- EFE, E., 1990. Büyüme eğrileri. Ç.Ü. Doktora tezi, Adana, 160s,(yayınlanmamış).
- EFE, E., 1995. Ki-kare dağılışının gamma, normal ve standart normal dağılışlarla olan ilişkisi ve rxc tabloları üzerine bir çalışma. Ç.Ü. Ziraat Fakültesi Dergisi, 10(3).
- ELHAN, A., 1997. Lojistik Regresyon Analizinin İncelenmesi ve Tıpta Bir Uygulaması. A.Ü. Yüksek Lisans Tezi, Ankara, 132s (yayınlanmamış).
- EMANUELSON, U., BENDIXEN, P.H., 1991. Occurrence of cystic ovaries in dairy cows in Sweden. *Preventive Veterinary Medicine.* 10:261-271.
- ENEVOLDSEN, C., SORENSEN, J.T., 1992. Effects of dry period length on clinical mastitis and other major clinical health disorders. *Journal of dairy science (USA).* 75:1007-1014.
- EVERITT, B.S., 1977. *The Analysis of Contingency Tables.* London, Chapman and Hall.
- FEARS, T.R., BROWN, C.C., 1986. Logistic regression models for retrospective case - control studies using complex sampling procedures. *biometrics.*, 42: 955-960.
- GANGULY, S.S., NIMBALKAR, U.N., 1992. Use of polytomous logistic model in matched case-control studies. *Biom. J.*, 2:209-217.
- GARTSIDE, P.S., GLUECK, C.J., 1995. The Important Role of Modifiable Dietary and Behavioral Characteristic in the Causation and Prevention of Coronary Heart Disease Hospitalization and Mortality. *Journal of the American College of the Nutrition.*, 14:71-79.
- GIBBONS, R.D., HEDEKER, D., 1996. Random effects probit and logistic regression models for three-level data. *Biometrika.*, 42:121-130.
- GROSSMAN, M., ANDERSON, V.L., 1985. Logistic growth curve of chickens: heritability of parameters. *the journal of heredity.*, 76:459-462.
- GROSSMAN, M., BOHREN, B.B., ANDERSON, V.L., 1985. Logistic growth curve of chickens: A comparison of techniques to estimate parameters. *The Journal of Heredity.*, 76: 397-399.
- HEISE, M.A., MYERS, R.H., 1996. Optimal designs for bivariate logistic regression. *Biometrics.*, 14: 613-623.
- HOSMER, D.W., JOVANOVIC, B., LEMESHOW, S., 1989. Best subsets logistic regression. *Biometrics.*, 45: 1265-1270.
- HOSMER, W.D., LEMESHOW, S., 1980. A goodness of fit test for the multiple logistic regression model. *Communications in Statistics.*, A10: 1043-1069.

- HOSMER, W.D., LEMESHOW, S., KLAR, J., 1988. Goodness of fit testing for multiple logistic regression analysis when the estimated probabilities are small. *Biometrical Journal*, 30: 911-924.
- HOSMER, W.D., LEMESHOW, S., 1989. *Applied logistic regression*. John Wiley & Sons, America, 307s.
- HOUBEN, E.H.P., DIJKHUIZEN, A.A., ARENDONK, J.A.M., HUIRNE, R.B.M., 1993. Short and long term production losses and repeatability of clinical mastitis in dairy cattle. *Journal of dairy science (USA)*. 76:2561-2578.
- HOUCK, W., 1988. Choice of scale and asymmetric logistic regression. *Biom. J.*, 1:79-86.
- HSU, J.S., LEONARD, T., 1995. Hierarchical bayesian semiparametric procedures for logistic regression. *Biometrika*, 84:85-93.
- HUANG, R., TU, J., 1998. Effects of hydrogen ion concentration and temperature on attachment of zoospores of *pythium aphanidermatum* to tomato roots. *Biometrika*, 63: 855-859.
- HUMBLOT, P., GRIMARD, B., RIBON, O., KHIREDINE, B., DERVISHI, V., THIBIER, M., 1996. Sources of variation of post partum cyclicity, ovulation and pregnancy rates in primiparous Charolais cows treated with norgestomet implants and PMSG. *Theriogenology (USA)*. 46:1085-1096.
- JUDGE, L.J., ERSKINE, R.J., BARTLETT, P.C., 1997. Recombinant bovine somatotropin and clinical mastitis: incidence, discarded milk following therapy, and culling. *Journal of dairy science (USA)*. 80: 3212-3218.
- KAY, R., LITTLE, S., 1987. Transformation of the explanatory variables in the logistic regression model for binary data. *Biometrika*, 74: 495-501.
- KLOIBER, L.L., WINN, N.J., SHAFFER, S.G., HASSANEIN, R.S., 1996. Late hyponatremia in very low birth weight infants: incidence and associated risk factors. *Journal of the American Dietetic Association (USA)*. 96:880-884.
- KOOPS, W.J., 1986. Multiphasic growth curve analysis. *Growth*, 59: 169-177.
- LA VECCHIA, C., FERRARONI, M., FRANCESCHI, S., MEZZETTI, M., DECARLI, A., NEGRI, E., 1997. Fibers and breast cancer risk. *Nutrition and cancer (USA)*, 28:264-269.
- LEMESHOW, S., HOSMER, W.D., 1982. The use of goodness of fit statistics in the development of logistic regression models. *American Journal of Epidemiology*, 115: 92-106.
- LE, C.T., 1984. Logistic models for cross-over designs. *Biometrika*, 71: 216-217.
- LEE, B.W., WAIN, J.C., KELSEY, K.T., WIENCKE, J.K., CHRISTIANI, D.C., 1998. Association between diet and lung cancer location. *American Journal of Respiratory and Critical Care Medicine*. 158: 1197-1203.
- LEES, V.W., MEEK, A.H., ROSENDAL, S., 1990. Epidemiology of *Haemophilus somnus* in young rams. *Canadian Journal of Veterinary Research*. 54:331-336.
- LUST, G., WILLIAMS, A.J., BURTON, W.N., PIJANOWSKI, G.J., BECK, K.A., RUBIN, G., 1993. Joint laxity and its association with hip dysplasia in Labrador Retrievers.
- MAJOR, D.J., 1980. Effect of simulated frost injury induced by paraquat on kernel growth and development in corn. *Canadian Journal of Plant Science*, 60: 419-426.

- MARGARET, A.C., LIANG, K., 1988. Conditional logistic regression models for correlated binary data. *Biometrika.*, 75: 501-506.
- MEYER, F., BAIRATI, I., FRADET, Y., MOORE, L., 1997. Dietary energy and nutrients in relation to preclinical prostate cancer. *Nutrition and cancer (USA).* 29:120-126.
- MOHAMMED, H.O., WHITE, M.E., GUARD, C.L., SMITH, M.C., MECHOR, G.D., BOOKER, C.W., 1991. A case control study of the association between blood selenium and cystic ovaries in lactating dairy cattle. L.D., Dascanio, J.J., Kenney, D.G. *Journal of dairy science (USA).* 74: 2180-2185.
- MORIN, D.E., CONSTABLE, P.D., MCCOY, G.C., 1998. Use of clinical parameters for differentiation of Gram positive and Gram negative mastitis in dairy cows vaccinated against lipopolysaccharide core antigens. *Journal of the American Veterinary Medical Association.* 212:1423-1431.
- MORRIS, A.K., SILK, W.K., 1992. Use of a flexible logistic function to describe axial growth of plants. *Bulletin of mathematical biology.*, 54:1069-1081.
- MUSCATI, S.K., KOSKI, K.G., GRAY, D.K., 1996. Increased Energy Intake in Pregnant Smokers Does Not Prevent Human Fetal Growth Retardation.
- NIELEN, M., SCHUKKEN, Y.H., BRAND, A., DELUYKER, H.A., MAATJE, K., 1995. Detection of subclinical mastitis from on line milking parlor data. *Journal of dairy science (USA).* 78:1039- 1049.
- OETZEL, G.R., 1991. Meta analysis of nutritional risk factors for milk fever in dairy cattle. *Journal of dairy science (USA).* 74:3900-3912.
- O'NEIL, T.J., BARRY, S.C., 1995. Truncated logistic regression. *Biometrics.*, 51: 533-541.
- PEOPLES, M.D., SIEGEL, E., SUCHINDRAN, C.M., ORIGASA, H., WARE, A., BARAKAT, A., 1991. Characteristics of maternal employment during pregnancy: effects on low birthweight. *American journal of public health (USA).* 81:1007-1012.
- PERZANOWSKI, M.S., SPORIK, R., SQUILLACE, S.P., GELBER, L.E., CALL, R., CARTER, M., PLATTS, M.T., 1998. Association of sensitization to *Alternaria* allergens with asthma among school age children. *Journal of Allergy and Clinical Immunology.* 101: 626-632.
- PINARD, M.H., JANSS, L.L.G., MAATMAN, R., NOORDHUIZEN, J.P.T.M., ZIJPP, A.J., 1993. Effect of divergent selection for immune responsiveness and of major histocompatibility complex on resistance to Marek's disease in chickens. *Poultry science (USA).* 72:391-402.
- ROBERTS, G., RAO, N.K., KUMAR, S., 1987. Logistic regression analysis of sample survey data. *Biometrika.*, 74: 1-12.
- RUIZ-VELESCO, S., 1989. Asymptotic efficiency of logistic regression relative to linear discriminant analysis. *Biometrika.*, 35: 58-79.
- SABLE, M.R., HERMAN, A.A., 1997. The relationship between prenatal health behavior advice and low birth weight. *Public health reports (USA).* 112:332-339.
- SANTOS, I.S., VICTORA, C.G., HUTTLY, S., CARVALHAL, J.B., 1998. Caffeine intake and low birth weight: a population based case control study. *American Journal of Epidemiology.* 147: 6275-7620.

- SCHUMANN, F.J., TOWNSEND, H.G.G., NAYLOR, J.M., 1990. Risk factors for mortality from diarrhea in beef calves in Alberta. *Canadian Journal of Veterinary Research*. 54:366-372.
- SCOTT, A., WILD, C.J., 1990. Fitting logistic regression models in stratified case-control studies. *Biometrics*, 47:497-510.
- SHRESTHA, H.K., GHIMIRE, S.C., 1996. Efficacy of sodium thiosulphate against aconite poisoning in migratory sheep and goats. Pokhara, Kaski (Nepal). Lumle Agricultural Research Centre. 13:
- SINGH, R.B., NIAZ, M.A., RASTOGI, S.S., SARITA, B., ZHANG, G., ZHU, S., BAJAJ, S., 1998. Current zinc intake and risk of diabetes and coronary artery disease and factors associated with insulin resistance in rural and urban populations of North India. *SM Journal of the American College of Nutrition*. 17: 564-570.
- STEFANI, E., BOFFETTA, P., CARZOGLIO, J., MENDILAHARSU, S., DENEIO, P.H., DE STEFANI, E., 1998. Tobacco smoking and alcohol drinking as risk factors for stomach cancer: a case control study in Uruguay. *Cancer Causes and Control*. 9:321-329.
- STEVENS, S.C., MCANARNEY, E.R., 1992. Determinants of weight gain in pregnant adolescents. *Journal of the American Dietetic Association (USA)*. 92:1348-1351.
- TATLIDİL, H., 1996. *Uygulamalı çok değişkenli İstatistiksel analiz. Cem ofset, Ankara, 424s.*
- THOMPSON, J.A., MARSH, W.E., CALVIN, J.A., ETHERINGTON, W.G., MOMONT, H.W., KINSEL, M.L., 1994. Pregnancy attrition associated with pregnancy testing by rectal palpation. *Journal of dairy science (USA)*. 77:3382-3387.
- VACH, W., SCHUMACHER, M., 1991. Logistic regression with incompletely observed categorical covariates: A comparison of three approaches. *Biometrika*, 80:353-362.
- YOĞURTÇUGİL, M.K., 1978. *Ki-Kare üzerine bir deneme. Sermet Matbaası, İstanbul, 133 s.*
- YURTSEVER, N., 1984. *Deneyisel istatistik metotları. Toprak ve gübre araştırma enstitüsü müdürlüğü yayınları, Ankara, 623s.*

ÖZGEÇMİŞ

28-01-1973 tarihinde Yozgat'ta doğdum. İlk, orta ve lise öğrenimimi Yozgat'ta tamamladım. 1992 yılında Çukurova Üniversitesi Ziraat Fakültesi Zootečni Bölümünde Lisans eğitime başladım. Lisans düzeyi bitirme tezini "Adana bölgesindeki kanatlı kesim haneleri" üzerine yaptım. 1996 yılında lisans eğitimimi tamamladıktan sonra aynı yıl Kahramanmaraş Sütçü İmam Üniversitesi Ziraat Fakültesi Zootečni Bölümünde, Biyometri ve Genetik Ana Bilim Dalında yüksek lisans eğitime başladım. Evliyim, bir kızım var.



Ek-1

Hatanın (ε_i) Bekleneni ve Varyansı:

$$E(\varepsilon_i) = P(y_i=0)(-\sum \beta x_i) + P(y_i=1)(1-\sum \beta x_i) = 0$$

$$-\sum \beta x_i + P(y=1) \sum \beta x_i + \underbrace{P(y=1) - P(y=1)}_{\sum \beta x_i} \sum \beta x_i = 0$$

$$Var(\varepsilon_i) = E(\varepsilon_i^2) = P(y=0)(-\sum \beta x_i)^2 + P(y=1)(1-\sum \beta x_i)^2$$

$$= [1 - P(y=1)](-\sum \beta x_i)^2 + P(y=1)[1 - 2\sum \beta x_i + (\sum \beta x_i)^2]$$

$$= (-\sum \beta x_i)^2 + P(y=1)(-\sum \beta x_i)^2 + P(y=1) - \underbrace{2P(y=1)\sum \beta x_i}_{\sum \beta x_i} + P(y=1)(\sum \beta x_i)^2$$

$$= (-\sum \beta x_i)^2 + \sum \beta x_i - 2(\sum \beta x_i)^2$$

$$= (\sum \beta x_i)^2 + \sum \beta x_i - 2(\sum \beta x_i)^2$$

$$= \sum \beta x_i - (\sum \beta x_i)^2$$

$$= \sum \beta x_i - (\sum \beta x_i) \sum \beta x_i$$

$$= \sum \beta x_i (1 - \sum \beta x_i) \text{ yani,}$$

$$Var(\varepsilon_i) = \sum_{k=0}^p \beta_k x_{ik} \left[1 - \sum_{k=0}^p \beta_k x_{ik} \right]$$

Ek-2. Şekil 3.3'deki Analiz İçin SAS'ta Veri ve Komut Dizimi.

```
DATA TEZ;  
INPUT YAS DA;  
CARDS;  
14 0  
15 0  
21 0  
.  
20 1  
39 1  
40 1  
;  
DATA DIFFS;  
SET TEZ;  
RUN;  
PROC CATMOD DATA=DIFFS;  
DIRECT YAS;  
RESPONSE CLOGITS;  
MODEL DA=YAS;  
RUN;
```

Ek-3. Şekil 3.4'deki Analiz İçin SAS'ta Veri ve Komut Dizimi.

```
DATA TEZ;  
INPUT YAS DA;  
CARDS;  
14 0  
15 0  
21 0  
.  
20 1  
39 1  
40 1  
;  
DATA DIFFS;  
SET TEZ;  
PROC CATMOD DATA=DIFFS;  
DIRECT ;  
RESPONSE CLOGITS;  
MODEL DA=;  
RUN;
```

Ek-4. Çizelge 4.8'deki Analiz İçin SAS'ta Veri ve Komut Dizimi.

```

DATA TEZ;
INPUT SIG YAS GS DGS D AK SADA C DA HT HB;
CARDS;
1 14 1 0 0 0 50.0 0 0 0 13.1
0 15 1 0 0 1 58.0 0 0 0 10.2
0 21 1 0 0 1 61.0 0 0 0 11.0
. . . . .
0 39 1 0 0 0 54.0 0 1 0 10.9
0 40 1 0 0 0 53.0 0 1 0 13.5
;
DATA DIFFS; SET TEZ;
SELECT (DGS);
  WHEN (0) DO; DGS1=0; DGS2=0;END;
  WHEN (1) DO; DGS1=1; DGS2=0;END;
  WHEN (2) DO; DGS1=0; DGS2=1;END;
END;
END;
PROC CATMOD DATA=DIFFS;
  DIRECT ; RESPONSE CLOGITS;
  MODEL DA= ;RUN;
PROC CATMOD DATA=DIFFS;
  DIRECT SIG; RESPONSE CLOGITS;
  MODEL DA= SIG;RUN;
PROC CATMOD DATA=DIFFS;
  DIRECT YAS ; RESPONSE CLOGITS;
  MODEL DA=YAS ;RUN;
PROC CATMOD DATA=DIFFS;
  DIRECT DGS1 DGS2; RESPONSE CLOGITS;
  MODEL DA= DGS1 DGS2;RUN;
PROC CATMOD DATA=DIFFS;
  DIRECT AK; RESPONSE CLOGITS;
  MODEL DA=AK;RUN;
PROC CATMOD DATA=DIFFS;
  DIRECT D; RESPONSE CLOGITS;
  MODEL DA=D;RUN;
PROC CATMOD DATA=DIFFS;
  DIRECT SADA; RESPONSE CLOGITS;
  MODEL DA= SADA;RUN;
PROC CATMOD DATA=DIFFS;
  DIRECT C; RESPONSE CLOGITS;
  MODEL DA= C;RUN;
PROC CATMOD DATA=DIFFS;
  DIRECT HT; RESPONSE CLOGITS;
  MODEL DA=HT ;RUN;
PROC CATMOD DATA=DIFFS;
  DIRECT HB; RESPONSE CLOGITS;
  MODEL DA=HB;
RUN;

```

Ek-5. Çizelge 4.10'daki Analiz İçin SAS'ta Veri ve Komut Dizimi.

```

DATA TEZ;
INPUT SIG YAS GS DGS D AK SADA C DA HT HB;
CARDS;
1 14 1 0 0 0 50.0 0 0 0 13.1
0 15 1 0 0 1 58.0 0 0 0 10.2
0 21 1 0 0 1 61.0 0 0 0 11.0
. . . . .
0 39 1 0 0 0 54.0 0 1 0 10.9
0 40 1 0 0 0 53.0 0 1 0 13.5
;

DATA DIFFS; SET TEZ;
SELECT (DGS);
  WHEN (0) DO; DGS1=0; DGS2=0;END;
  WHEN (1) DO; DGS1=1; DGS2=0;END;
  WHEN (2) DO; DGS1=0; DGS2=1;END;
END;

PROC CATMOD DATA=DIFFS;
  DIRECT SIG YAS DGS1 DGS2 D SADA C HT;
  RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 D SADA C HT;
RUN;

```

Ek-6. Çizelge 4.11'deki Analiz İçin SAS'ta Veri ve Komut Dizimi.

```

DATA TEZ;
INPUT SIG YAS GS DGS D AK SADA C DA HT HB;
CARDS;
1 14 1 0 0 0 50.0 0 0 0 13.1
0 15 1 0 0 1 58.0 0 0 0 10.2
0 21 1 0 0 1 61.0 0 0 0 11.0
. . . . .
0 39 1 0 0 0 54.0 0 1 0 10.9
0 40 1 0 0 0 53.0 0 1 0 13.5
;

DATA DIFFS; SET TEZ;
SELECT (DGS);
  WHEN (0) DO; DGS1=0; DGS2=0;END;
  WHEN (1) DO; DGS1=1; DGS2=0;END;
  WHEN (2) DO; DGS1=0; DGS2=1;END;
END;

PROC CATMOD DATA=DIFFS;
  DIRECT SIG YAS DGS1 DGS2 SADA C HT;
  RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 SADA C HT;
RUN;

```

Ek-7. Çizelge 4.13'deki Analiz İçin SAS'ta Veri ve Komut Dizimi.

```

DATA TEZ;
INPUT SIG YAS DYAS DGS SADA C DA HT ;
CARDS;
1 14 0 0 50.0 0 0 0
0 15 0 0 58.0 0 0 0
0 21 1 0 61.0 0 0 0
. . . . .
0 39 3 0 54.0 0 1 0
0 40 3 0 53.0 0 1 0
;
DATA DIFFS; SET TEZ;
SELECT (DGS);
  WHEN (0) DO; DGS1=0; DGS2=0;END;
  WHEN (1) DO; DGS1=1; DGS2=0;END;
  WHEN (2) DO; DGS1=0; DGS2=1;END;
END;
END;
SELECT (DYAS);
  WHEN (0) DO; DYAS1=0; DYAS2=0; DYAS3=0;END;
  WHEN (1) DO; DYAS1=1; DYAS2=0; DYAS3=0;END;
  WHEN (2) DO; DYAS1=0; DYAS2=1; DYAS3=0;END;
  WHEN (3) DO; DYAS1=0; DYAS2=0; DYAS3=1;END;
END;
END;
PROC CATMOD DATA=DIFFS;
  DIRECT SIG DYAS1 DYAS2 DYAS3 DGS1 DGS2 SADA C HT;
  RESPONSE CLOGITS;
  MODEL DA=SIG DYAS1 DYAS2 DYAS3 DGS1 DGS2 SADA C HT;
RUN;

```

Ek-8. Çizelge 4.16'daki Analiz İçin SAS'ta Veri ve Komut Dizimi.

```

DATA TEZ;
INPUT SIG YAS DGS SADA DSADA C DA HT ;
CARDS;
1 14 0 50.0 0 0 0 0
0 15 0 58.0 1 0 0 0
0 21 0 61.0 2 0 0 0
. . . . .
0 39 0 54.0 0 0 1 0
0 40 0 53.0 0 0 1 0
/
DATA DIFFS; SET TEZ;
SELECT (DGS);
    WHEN (0) DO; DGS1=0; DGS2=0;END;
    WHEN (1) DO; DGS1=1; DGS2=0;END;
    WHEN (2) DO; DGS1=0; DGS2=1;END; END;
END;
SELECT (DSADA);
    WHEN (0) DO; DSADA1=0; DSADA2=0; DSADA3=0;END;
    WHEN (1) DO; DSADA1=1; DSADA2=0; DSADA3=0;END;
    WHEN (2) DO; DSADA1=0; DSADA2=1; DSADA3=0;END;
    WHEN (3) DO; DSADA1=0; DSADA2=0; DSADA3=1;END; END;
END;
PROC CATMOD DATA=DIFFS;
DIRECT SIG YAS DSADA1 DSADA2 DSADA3 DGS1 DGS2 C HT;
RESPONSE CLOGITS;
MODEL DA=SIG YAS DSADA1 DSADA2 DSADA3 DGS1 DGS2 C HT;RUN;

```

Ek-9. Çizelge 4.19'deki Analiz İçin SAS'ta Veri ve Komut Dizimi.

```

DATA TEZ;
INPUT SIG YAS DGS SADA DSADA C DA HT ;
CARDS;
1 14 0 50.0 1 0 0 0
0 15 0 58.0 0 0 0 0
0 21 0 61.0 0 0 0 0
. . . . .
0 39 0 54.0 1 0 1 0
0 40 0 53.0 1 0 1 0
/
DATA DIFFS; SET TEZ;
SELECT (DGS);
    WHEN (0) DO; DGS1=0; DGS2=0;END;
    WHEN (1) DO; DGS1=1; DGS2=0;END;
    WHEN (2) DO; DGS1=0; DGS2=1;END; END;
END;
PROC CATMOD DATA=DIFFS;
DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
RESPONSE CLOGITS;
MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT;RUN;

```

Ek-10. Çizelge 4.20'deki Analiz İçin SAS'ta Veri ve Komut Dizimi.

```

DATA TEZ;
INPUT SIG YAS DGS SADA DSADA C DA HT ;
CARDS;
1 14 0 50.0 1 0 0 0
0 15 0 58.0 0 0 0 0
0 21 0 61.0 0 0 0 0
. . . . .
0 39 0 54.0 1 0 1 0
0 40 0 53.0 1 0 1 0
;
DATA DIFFS; SET TEZ;
SELECT (DGS);
    WHEN (0) DO; DGS1=0; DGS2=0;END;
    WHEN (1) DO; DGS1=1; DGS2=0;END;
    WHEN (2) DO; DGS1=0; DGS2=1;END; END;
END;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT SIG*YAS;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT SIG*DGS1;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT SIG*DSADA;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT SIG*C;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT SIG*HT;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT YAS*DGS1;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT YAS*DSADA;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT YAS*C;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT YAS*HT;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT DGS1*DSADA;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT DGS1*C;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT DGS1*HT;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT DSADA*C;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT DSADA*HT;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT C*HT;RUN;
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;
    RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT YAS*DSADA
SIG*DSADA;RUN;

```

Ek-11. Çizelge 4.21'deki Analiz İçin SAS'ta Veri ve Komut Dizimi.

```
DATA TEZ;  
INPUT SIG YAS DGS SADA DSADA C DA HT ;  
CARDS;  
1 14 0 50.0 1 0 0 0  
0 15 0 58.0 0 0 0 0  
0 21 0 61.0 0 0 0 0  
.  
0 39 0 54.0 1 0 1 0  
0 40 0 53.0 1 0 1 0  
;  
DATA DIFFS; SET TEZ;  
SELECT (DGS);  
  WHEN (0) DO; DGS1=0; DGS2=0;END;  
  WHEN (1) DO; DGS1=1; DGS2=0;END;  
  WHEN (2) DO; DGS1=0; DGS2=1;END;  
END;  
END;  
PROC CATMOD DATA=DIFFS;DIRECT SIG YAS DGS1 DGS2 DSADA C HT;  
  RESPONSE CLOGITS; MODEL DA= SIG YAS DGS1 DGS2 DSADA C HT  
YAS*DSADA SIG*DSADA;  
RUN;
```

Ek-12. Düşük Doğum Ağırlığı ve Aday Risk Faktörleri İçin Gerçek Veriler

No	DA	SİG	YAŞ	GS	D	AK	SADA	C	HT	HB
1	0	1	14	1	0	0	50.0	0	0	13.1
2	0	0	15	1	0	1	58.0	0	0	10.2
3	0	0	21	1	0	1	61.0	0	0	11.0
4	0	1	25	2	1	1	54.0	0	0	10.8
5	1	0	18	1	0	0	53.0	1	1	11.5
6	0	1	14	3	1	1	50.0	0	0	12.0
7	0	0	30	1	0	0	59.0	1	0	9.60
8	0	0	31	2	1	0	50.0	0	0	13.8
9	1	1	19	1	0	1	61.5	1	1	12.0
10	0	0	23	3	1	0	63.0	0	0	11.7
11	0	0	22	1	0	0	60.0	0	0	13.2
12	1	0	15	1	0	0	57.0	1	1	11.0
13	0	0	20	1	0	1	64.0	0	0	11.3
14	0	0	17	1	0	0	57.0	0	0	12.0
15	0	0	21	1	0	0	71.0	1	1	10.6
16	0	0	30	1	0	0	63.0	1	0	9.80
17	0	0	29	1	0	0	62.0	1	0	13.1
18	1	0	34	2	1	1	66.5	1	0	12.2
19	1	1	20	1	0	1	48.5	1	1	10.8
20	0	0	24	2	1	0	60.0	0	0	13.0
21	1	0	21	2	1	0	61.0	1	0	12.1
22	1	0	20	2	0	0	72.0	0	1	12.1
23	1	0	20	1	0	0	53.0	1	1	10.8
24	1	0	14	1	0	0	50.0	1	1	10.7
25	0	0	24	2	0	0	58.0	1	0	12.0
26	0	0	28	2	1	0	63.5	0	0	13.3
27	1	0	15	1	0	1	58.0	0	0	12.6
28	0	0	30	2	1	0	67.0	1	0	9.70
29	1	0	15	1	0	0	61.5	1	1	13.0
30	1	0	25	1	0	0	62.0	0	1	12.7
31	0	0	22	1	0	0	67.0	0	0	11.0
32	0	0	23	2	1	0	66.5	0	0	10.0
33	1	1	25	1	0	0	61.0	1	0	9.30
34	0	0	31	1	0	0	75.5	0	0	13.7
35	1	0	20	1	0	0	51.0	1	0	12.0
36	0	1	20	2	0	0	54.0	0	0	11.0
37	1	0	20	1	0	0	54.0	1	0	10.3
38	0	0	29	2	1	1	75.0	0	0	12.0
39	1	1	32	1	0	0	80.5	0	0	11.6
40	0	0	33	1	0	1	65.0	0	1	9.00
41	0	0	25	2	1	0	72.0	0	0	13.5
42	0	0	24	3	1	0	63.5	0	0	12.6
43	0	0	28	1	0	0	59.5	0	0	9.60
44	0	1	28	2	1	0	55.0	0	0	10.3
45	0	0	20	1	0	0	54.0	0	0	12.4
46	0	0	21	1	0	0	55.0	0	0	11.8
47	0	0	29	2	1	0	61.0	0	0	10.1
48	0	0	28	2	1	0	58.0	0	0	9.90
49	0	0	23	1	0	0	66.0	0	0	11.6
50	1	1	15	1	0	1	61.5	0	0	10.4
51	0	1	28	2	1	0	65.0	0	0	11.1
52	1	1	24	2	1	0	74.0	0	0	9.50
53	0	0	14	1	0	1	48.5	1	0	12.2

Ek-12. Düşük Doğum Ağırlığı ve Aday Risk Faktörleri İçin Gerçek Veriler (Devam).

No	DA	SİG	YAŞ	GS	D	AK	SADA	C	HT	HB
54	1	1	27	2	1	1	50.0	1	0	8.80
55	0	0	28	2	1	1	70.0	0	0	10.2
56	0	1	25	1	0	1	50.0	1	0	11.7
57	0	0	34	1	0	0	74.0	0	0	10.4
58	0	0	19	2	1	0	63.0	0	0	12.8
59	0	1	20	1	0	0	51.0	0	0	10.8
60	1	0	20	1	0	0	49.0	1	0	13.6
61	0	0	30	2	1	0	63.0	0	0	13.2
62	1	1	22	2	1	0	64.0	0	0	12.2
63	1	1	25	3	1	0	82.0	0	1	13.0
64	0	0	28	1	0	0	50.0	0	0	10.1
65	1	1	20	1	0	0	50.0	1	0	9.40
66	0	0	21	1	0	0	59.0	1	0	10.4
67	0	0	23	1	0	0	58.0	0	0	12.4
68	0	1	25	2	1	0	56.0	1	0	11.5
69	0	0	24	1	0	1	50.0	0	0	12.6
70	0	0	28	3	1	0	50.0	0	0	11.9
71	1	0	19	2	1	0	57.5	1	0	11.2
72	1	0	20	1	0	0	68.0	0	0	11.6
73	0	0	28	1	0	0	54.0	0	0	11.1
74	0	0	23	1	0	0	59.0	0	0	12.3
75	0	1	26	1	0	0	54.0	0	0	12.0
76	0	0	21	2	1	0	60.0	0	0	13.1
77	0	1	19	1	0	0	50.0	1	1	11.6
78	0	0	26	2	0	0	58.0	0	0	11.5
79	0	0	25	1	0	0	55.0	0	0	12.1
80	0	0	31	1	0	0	67.0	1	0	10.3
81	0	1	28	3	1	0	64.5	0	0	10.5
82	1	0	30	2	1	0	57.0	1	1	11.9
83	0	0	24	2	1	0	58.5	0	0	12.5
84	0	0	20	1	0	0	48.5	1	1	10.3
85	0	0	21	2	1	0	65.0	1	0	12.6
86	0	0	28	3	1	1	60.0	0	0	11.8
87	0	0	28	2	1	0	78.0	0	1	11.7
88	0	0	20	1	0	0	61.5	0	0	10.6
89	0	0	20	1	0	0	50.0	0	0	13.4
90	0	0	24	2	1	0	68.0	0	0	10.0
91	0	0	27	2	1	0	70.0	0	0	9.80
92	0	0	30	2	1	0	67.5	1	0	10.3
93	0	0	20	1	0	0	66.0	0	0	13.8
94	1	1	20	1	0	0	54.0	1	0	10.7
95	1	1	20	1	0	0	50.0	0	0	12.1
96	0	0	26	1	0	0	63.5	1	0	10.0
97	0	0	21	1	0	0	51.0	0	0	12.5
98	0	0	30	1	0	0	63.0	0	0	9.70
99	0	0	25	2	0	0	65.0	0	0	11.5
100	0	0	22	2	1	0	59.0	0	0	13.0
101	0	0	35	1	0	0	68.0	1	0	10.6
102	0	0	26	2	1	0	58.0	0	0	12.5
103	0	1	32	2	1	0	79.0	1	0	11.4
104	0	0	26	3	1	0	60.0	0	0	10.0
105	1	0	20	1	0	0	54.0	0	0	12.8
106	0	0	21	1	0	0	57.0	0	0	11.6

Ek-12. Düşük Doğum Ağırlığı ve Aday Risk Faktörleri İçin Gerçek Veriler (Devam).

No	DA	SİG	YAŞ	GS	D	AK	SADA	C	HT	HB
107	0	0	30	2	1	0	65.0	0	0	12.6
108	0	0	33	3	1	0	70.0	1	0	13.1
109	1	1	26	1	0	0	62.0	0	0	8.30
110	1	0	20	1	0	0	57.0	1	0	11.9
111	0	0	21	1	0	0	58.0	0	0	9.20
112	1	0	20	1	0	0	50.0	1	1	14.2
113	1	1	19	1	0	0	57.0	1	1	11.3
114	1	1	26	3	1	0	66.0	1	0	10.2
115	0	0	28	2	1	1	71.0	0	0	11.4
116	0	0	21	1	0	0	66.0	0	0	9.70
117	0	0	31	3	1	0	70.0	1	0	11.5
118	0	0	36	1	0	1	75.0	1	0	12.8
119	0	0	30	2	1	0	69.0	0	0	11.2
120	1	0	19	1	0	0	57.5	1	0	13.1
121	1	0	24	1	0	0	54.0	1	1	10.9
122	1	0	24	1	0	0	55.0	1	1	11.5
123	0	0	25	2	1	0	62.0	1	0	12.1
124	0	0	30	2	1	0	68.0	0	0	13.7
125	0	0	27	1	0	0	62.0	1	0	11.2
126	0	1	31	2	1	0	61.5	1	0	8.80
127	1	0	19	1	0	0	57.0	1	0	10.3
128	1	0	20	1	0	0	51.0	1	0	13.2
129	0	0	23	2	0	1	54.5	1	1	9.50
130	1	0	24	1	0	0	52.0	1	1	13.4
131	0	1	28	3	1	0	61.0	1	0	14.4
132	0	0	27	1	0	0	60.0	1	0	11.8
133	1	0	20	1	0	0	50.5	1	0	13.6
134	0	0	33	2	1	0	61.5	1	0	12.5
135	0	0	25	3	1	0	66.0	0	0	11.9
136	0	0	24	1	0	0	53.5	1	0	10.5
137	1	0	20	1	0	0	51.0	1	1	11.6
138	0	0	34	3	1	0	68.0	1	0	12.8
139	0	0	42	3	1	0	61.5	1	1	13.1
140	0	0	25	2	0	0	63.0	1	0	10.2
141	0	0	27	2	0	0	73.0	0	0	11.6
142	0	0	20	1	0	1	56.5	0	0	9.80
143	0	0	30	1	0	0	71.0	1	0	10.5
144	0	0	35	2	1	0	73.0	1	0	13.2
145	1	0	19	2	1	0	62.0	1	0	12.7
146	0	1	26	3	1	0	71.5	1	0	10.9
147	0	1	24	2	0	0	76.0	1	0	13.2
148	1	0	20	1	0	0	51.5	1	1	11.2
149	0	0	31	2	1	0	66.0	0	0	11.8
150	0	0	24	2	1	0	57.0	0	0	10.9
151	0	0	30	2	1	0	68.0	0	0	9.80
152	0	0	23	1	0	0	61.5	1	0	11.6
153	0	0	26	2	0	1	63.0	1	0	12.3
154	0	1	22	1	0	0	61.0	0	0	12.4
155	1	0	25	1	0	0	56.0	1	0	11.9
156	0	0	30	2	1	0	60.0	0	0	13.6
157	1	0	24	2	0	0	51.5	1	1	10.4
158	1	1	20	1	0	0	56.0	0	0	12.5
159	0	0	29	2	0	0	75.5	0	0	12.0

Ek-12. Düşük Doğum Ağırlığı ve Aday Risk Faktörleri İçin Gerçek Veriler (Devam).

No	DA	SİG	YAŞ	GS	D	AK	SADA	C	HT	HB
160	0	0	27	1	0	0	60.0	0	1	11.7
161	0	0	20	2	1	0	59.0	1	0	10.8
162	0	0	23	1	0	0	61.5	0	0	9.50
163	1	0	20	1	0	0	64.0	0	0	12.1
164	0	1	17	1	0	0	67.0	1	0	12.0
165	0	0	30	1	0	0	63.0	0	0	13.6
166	0	0	23	2	1	0	62.0	0	0	9.80
167	1	0	22	1	0	0	62.0	1	0	10.2
168	0	0	22	1	0	0	51.0	1	0	10.0
169	1	0	27	2	1	0	48.0	1	0	11.3
170	0	0	34	2	1	0	60.0	1	0	10.0
171	0	0	29	3	1	0	64.0	0	0	12.1
172	0	0	31	2	1	0	71.0	0	0	13.4
173	0	0	18	1	0	0	74.0	0	0	11.4
174	0	0	25	2	1	0	58.0	0	0	11.0
175	0	1	27	2	1	1	61.0	0	0	10.0
176	1	0	22	1	0	0	58.0	1	1	13.6
177	0	0	19	1	0	0	57.0	0	0	9.80
178	0	0	15	1	0	0	58.0	1	0	10.7
179	0	0	28	3	1	0	70.0	1	0	10.9
180	0	0	34	2	1	0	73.0	1	0	12.0
181	0	0	18	1	0	0	61.0	0	0	10.5
182	0	0	26	3	1	0	58.0	1	0	11.7
183	0	0	30	2	1	0	64.0	1	0	12.1
184	0	0	30	2	1	0	70.0	0	0	9.60
185	0	0	21	1	0	0	67.0	0	0	11.3
186	1	0	25	1	0	0	65.0	1	0	13.4
187	0	0	31	2	1	0	64.5	1	0	13.1
188	0	1	36	3	1	0	74.0	0	0	11.1
189	0	0	24	2	1	0	57.0	0	0	10.0
190	1	0	18	1	0	0	60.0	0	1	12.6
191	0	0	20	1	0	0	63.0	1	0	13.8
192	0	0	27	3	1	0	66.0	1	0	9.00
193	0	0	29	2	1	0	55.0	1	0	10.8
194	0	0	23	3	1	1	71.0	0	0	12.0
195	1	1	20	2	0	0	48.0	1	0	11.4
196	1	1	20	1	0	0	48.0	1	0	13.6
197	1	0	24	2	1	0	52.0	0	0	12.8
198	0	0	23	2	0	0	61.0	1	0	11.0
199	0	0	27	3	1	0	60.0	1	0	10.4
200	0	0	26	2	1	1	67.0	1	0	11.6
201	0	0	33	1	0	0	71.0	1	0	12.3
202	1	1	19	1	0	0	60.0	0	1	11.0
203	0	1	24	2	1	0	58.0	0	0	10.9
204	1	1	20	1	0	0	47.5	0	0	12.0
205	0	0	26	3	1	0	71.0	1	0	12.5
206	0	0	25	3	1	0	80.0	0	0	13.2
207	0	0	25	2	0	1	57.0	0	0	10.0
208	0	0	33	3	1	0	61.0	1	0	11.5
209	0	0	27	2	1	0	68.0	1	0	11.0
210	0	0	36	3	1	0	71.0	1	0	11.3
211	0	0	20	2	1	0	57.0	1	0	13.7
212	0	0	21	2	1	0	60.0	1	1	12.6

Ek-12. Düşük Doğum Ağırlığı ve Aday Risk Faktörleri İçin Gerçek Veriler (Devam).

No	DA	SiG	YAŞ	GS	D	AK	SADA	C	HT	HB
213	0	0	31	2	1	0	60.0	1	0	10.0
214	0	0	18	1	0	0	71.0	0	1	10.8
215	0	0	22	3	1	0	67.0	0	0	9.00
216	0	0	27	2	0	0	65.0	1	0	11.6
217	0	0	25	3	1	0	57.0	1	0	11.3
218	0	0	23	2	1	0	66.5	1	0	9.60
219	0	0	24	1	0	0	70.0	1	0	10.5
220	0	0	30	2	1	0	65.5	1	0	12.0
221	0	0	29	2	1	0	73.0	0	0	12.1
222	1	0	20	1	0	0	53.5	0	1	13.0
223	1	0	19	1	0	0	50.0	0	0	13.1
224	0	0	20	1	0	0	58.0	0	0	10.2
225	0	0	21	1	0	0	61.0	0	0	11.0
226	0	0	24	2	1	0	54.0	1	0	10.8
227	1	0	20	1	0	0	53.0	1	0	11.5
228	0	1	33	3	1	0	62.0	0	0	12.0
229	0	0	30	2	1	0	59.0	0	0	9.60
230	0	0	31	2	1	0	71.0	0	0	13.8
231	1	0	19	1	0	0	58.5	1	0	12.0
232	0	0	23	3	1	0	63.0	0	1	11.7
233	1	0	22	1	0	0	60.0	0	0	13.2
234	0	0	20	1	0	0	57.0	0	0	11.0
235	0	0	20	1	0	1	64.0	0	0	11.3
236	0	0	17	1	0	0	57.0	0	0	12.0
237	0	0	18	1	0	0	71.0	1	0	10.6
238	0	0	27	3	1	0	64.0	1	0	9.80
239	1	1	29	2	1	0	62.0	1	0	13.1
240	0	1	34	2	1	0	63.5	0	0	12.2
241	1	1	19	1	0	0	48.5	1	0	10.8
242	0	0	24	2	1	0	60.0	0	0	13.0
243	0	0	21	2	1	0	61.0	0	0	12.1
244	0	0	20	2	0	0	72.0	0	1	12.1
245	1	0	19	1	0	0	53.0	0	0	10.8
246	0	0	25	3	1	0	57.0	0	0	10.7
247	0	0	24	2	0	0	58.0	1	0	12.0
248	0	0	28	3	1	0	63.5	0	0	13.3
249	0	0	36	3	1	1	72.0	0	0	12.6
250	0	0	30	3	1	0	67.0	1	0	9.70
251	0	0	18	1	0	0	59.5	0	0	13.0
252	1	0	21	1	0	0	62.0	0	0	12.7
253	0	0	22	1	0	0	67.0	0	0	11.0
254	0	0	23	2	1	0	65.5	0	0	10.0
255	1	1	23	1	0	0	61.0	1	0	9.30
256	0	0	27	3	1	0	75.5	1	0	13.7
257	1	0	20	1	0	0	51.0	1	0	12.0
258	0	0	20	1	0	0	54.0	1	1	11.0
259	1	0	20	1	0	0	57.0	1	0	10.7
260	0	0	29	2	1	1	75.0	0	0	12.0
261	0	0	32	3	1	0	80.5	0	1	11.6
262	0	0	33	4	1	1	61.0	1	0	9.00
263	0	0	25	2	1	0	72.0	0	0	13.5
264	0	0	24	3	1	0	63.5	0	0	12.6
265	0	0	20	1	0	0	68.0	1	0	11.8

Ek-12. Düşük Doğum Ağırlığı ve Aday Risk Faktörleri İçin Gerçek Veriler (Devam).

No	DA	SİG	YAŞ	GS	D	AK	SADA	C	HT	HB
266	0	0	29	4	1	0	66.0	1	0	12.3
267	1	0	22	1	0	0	64.0	1	0	12.3
268	0	0	25	1	0	0	65.0	1	0	9.80
269	1	0	19	1	0	0	60.0	1	0	9.30
270	0	1	27	3	1	0	61.5	0	1	10.9
271	0	0	28	3	1	0	70.0	1	0	13.1
272	1	0	24	2	0	1	54.5	0	1	12.0
273	0	0	28	2	1	0	60.0	1	0	10.0
274	0	1	30	2	1	0	65.0	1	0	10.0
275	1	0	19	1	0	0	49.0	1	0	10.0
276	0	0	32	4	1	0	57.0	1	0	12.0
277	1	0	23	1	0	0	75.0	1	0	13.0
278	0	0	30	2	1	0	50.0	1	0	11.1
279	0	0	19	1	0	0	53.5	1	0	13.0
280	0	0	27	3	1	1	68.0	1	0	12.7
281	0	1	32	4	1	0	59.0	1	0	11.3
282	1	0	20	1	0	0	66.0	1	0	10.8
283	1	0	39	1	0	0	54.0	0	0	10.9
284	1	0	40	1	0	0	53.0	0	0	13.5

Ek-13**Kartil Bölünme:**

$$Q_{h/4} = N \frac{h}{4} + \frac{1}{2}$$

h= kaçınıcı kartil olduđu

N= gözlem sayısı

X sürekli bağımsız değışkenine göre veriler büyüklük sırasına dizilir. Örneğın N=65 olduğında h=1, h=2 ve h=3 için kartil noktaları;

$$Q_{1/4} = 65 \frac{1}{4} + \frac{1}{2} = 16.75 = 17'inci gözlem$$

$$Q_{2/4} = 65 \frac{2}{4} + \frac{1}{2} = 33 = 33'ümcü gözlem$$

$$Q_{3/4} = 65 \frac{3}{4} + \frac{1}{2} = 49.25 = 49'uncu gözlem$$

olarak bulunur.

Ek-14. Sabit Denek Sayılı ve Sabit P(xi) Aralıklı Onlu Risk Grupları İçin Beklenen Frekansların Bulunması.

No	y_i	SlG	YAŞ	DGS	DSADA	C	HT	$\hat{\ell}(x_i)$	Sabit Denek Sayılı	P(xi)	1-P(xi)	Sabit P(xi) Aralıklı
249	0	0	36	2	0	0	0	-6,6018	n=28 $\Sigma P(x_i) = 0.267$ $\Sigma(1-P(x_i)) = 27.873$	0,001356	0,998644	n=110 $\Sigma P(x_i) = 3.493$ $\Sigma(1-P(x_i)) = 106.5$
171	0	0	29	2	0	0	0	-5,3712		0,004627	0,995373	
210	0	0	36	2	0	1	0	-5,3593		0,004682	0,995318	
248	0	0	28	2	0	0	0	-5,1954		0,005511	0,994489	
86	0	0	28	2	0	0	0	-5,0196		0,006564	0,993436	
138	0	0	34	2	0	1	0	-5,0077		0,006642	0,993358	
139	0	0	42	2	0	1	1	-4,9854		0,006791	0,993209	
104	0	0	26	2	0	0	0	-4,8438		0,007816	0,992184	
108	0	0	33	2	0	1	0	-4,8319		0,007908	0,992092	
208	0	0	33	2	0	1	0	-4,8319		0,007908	0,992092	
262	0	0	33	2	0	1	0	-4,8319		0,007908	0,992092	
135	0	0	25	2	0	0	0	-4,668		0,009304	0,990696	
206	0	0	25	2	0	0	0	-4,668		0,009304	0,990696	
246	0	0	25	2	0	0	0	-4,668		0,009304	0,990696	
276	0	0	32	2	0	1	0	-4,6561		0,009414	0,990586	
42	0	0	24	2	0	0	0	-4,4922		0,011072	0,988928	
264	0	0	24	2	0	0	0	-4,4922		0,011072	0,988928	
117	0	0	31	2	0	1	0	-4,4803		0,011203	0,988797	
261	0	0	32	2	0	0	1	-4,4699		0,011319	0,988681	
149	0	0	31	1	0	0	0	-4,4218		0,01187	0,98813	
172	0	0	31	1	0	0	0	-4,4218		0,01187	0,98813	
230	0	0	31	1	0	0	0	-4,4218		0,01187	0,98813	
10	0	0	23	2	0	0	0	-4,3164		0,013172	0,986828	
194	0	0	23	2	0	0	0	-4,3164		0,013172	0,986828	
250	0	0	30	2	0	1	0	-4,3045		0,013328	0,986672	
61	0	0	30	1	0	0	0	-4,246		0,014119	0,985881	
107	0	0	30	1	0	0	0	-4,246		0,014119	0,985881	
119	0	0	30	1	0	0	0	-4,246		0,014119	0,985881	
124	0	0	30	1	0	0	0	-4,246	n=29 $\Sigma P(x_i) = 0.543$ $\Sigma(1-P(x_i)) = 28.457$	0,014119	0,985881	
151	0	0	30	1	0	0	0	-4,246		0,014119	0,985881	
156	0	0	30	1	0	0	0	-4,246		0,014119	0,985881	
184	0	0	30	1	0	0	0	-4,246		0,014119	0,985881	
229	0	0	30	1	0	0	0	-4,246		0,014119	0,985881	
188	0	1	36	2	0	0	0	-4,1789		0,015084	0,984916	
215	0	0	22	2	0	0	0	-4,1406		0,015664	0,984336	
266	0	0	29	2	0	1	0	-4,1287		0,015849	0,984151	
38	0	0	29	1	0	0	0	-4,0702		0,016787	0,983213	
47	0	0	29	1	0	0	0	-4,0702		0,016787	0,983213	
159	0	0	29	1	0	0	0	-4,0702		0,016787	0,983213	
221	0	0	29	1	0	0	0	-4,0702		0,016787	0,983213	
260	0	0	29	1	0	0	0	-4,0702		0,016787	0,983213	
179	0	0	28	2	0	1	0	-3,9529		0,018837	0,981163	
271	0	0	28	2	0	1	0	-3,9529		0,018837	0,981163	
26	0	0	28	1	0	0	0	-3,8944		0,01995	0,98005	
48	0	0	28	1	0	0	0	-3,8944		0,01995	0,98005	

Ek-14. Sabit Denek Sayılı ve Sabit P(xi) Aralıklı Onlu Risk Grupları İçin Beklenen Frekansların Bulunması.

55	0	0	28	1	0	0	0	-3,8944	0,01995	0,98005
115	0	0	28	1	0	0	0	-3,8944	0,01995	0,98005
144	0	0	35	1	0	1	0	-3,8825	0,020183	0,979817
57	0	0	34	0	0	0	0	-3,8236	0,021382	0,978618
192	0	0	27	2	0	1	0	-3,7771	0,022377	0,977623
199	0	0	27	2	0	1	0	-3,7771	0,022377	0,977623
238	0	0	27	2	0	1	0	-3,7771	0,022377	0,977623
256	0	0	27	2	0	1	0	-3,7771	0,022377	0,977623
280	0	0	27	2	0	1	0	-3,7771	0,022377	0,977623
91	0	0	27	1	0	0	0	-3,7186	0,023693	0,976307
141	0	0	27	1	0	0	0	-3,7186	0,023693	0,976307
18	1	0	34	1	0	1	0	-3,7067	0,02397	0,97603
170	0	0	34	1	0	1	0	-3,7067	0,02397	0,97603
180	0	0	34	1	0	1	0	-3,7067	0,02397	0,97603
228	0	1	33	2	0	0	0	-3,6515	0,025296	0,974704
182	0	0	26	2	0	1	0	-3,6013	0,026563	0,973437
205	0	0	26	2	0	1	0	-3,6013	0,026563	0,973437
78	0	0	26	1	0	0	0	-3,5428	0,028119	0,971881
102	0	0	26	1	0	0	0	-3,5428	0,028119	0,971881
134	0	0	33	1	0	1	0	-3,5309	0,028446	0,971554
217	0	0	25	2	0	1	0	-3,4255	0,031508	0,968492
41	0	0	25	1	0	0	0	-3,367	0,033343	0,966657
99	0	0	25	1	0	0	0	-3,367	0,033343	0,966657
174	0	0	25	1	0	0	0	-3,367	0,033343	0,966657
207	0	0	25	1	0	0	0	-3,367	0,033343	0,966657
263	0	0	25	1	0	0	0	-3,367	0,033343	0,966657
34	0	0	31	0	0	0	0	-3,2962	0,035702	0,964298
20	0	0	24	1	0	0	0	-3,1912	0,039498	0,960502
83	0	0	24	1	0	0	0	-3,1912	0,039498	0,960502
90	0	0	24	1	0	0	0	-3,1912	0,039498	0,960502
150	0	0	24	1	0	0	0	-3,1912	0,039498	0,960502
189	0	0	24	1	0	0	0	-3,1912	0,039498	0,960502
242	0	0	24	1	0	0	0	-3,1912	0,039498	0,960502
187	0	0	31	1	0	1	0	-3,1793	0,039952	0,960048
213	0	0	31	1	0	1	0	-3,1793	0,039952	0,960048
98	0	0	30	0	0	0	0	-3,1204	0,042274	0,957726
165	0	0	30	0	0	0	0	-3,1204	0,042274	0,957726
6	0	1	14	2	1	0	0	-3,0245	0,046331	0,953669
32	0	0	23	1	0	0	0	-3,0154	0,046735	0,953265
166	0	0	23	1	0	0	0	-3,0154	0,046735	0,953265
254	0	0	23	1	0	0	0	-3,0154	0,046735	0,953265
28	0	0	30	1	0	1	0	-3,0035	0,047268	0,952732
92	0	0	30	1	0	1	0	-3,0035	0,047268	0,952732
183	0	0	30	1	0	1	0	-3,0035	0,047268	0,952732
220	0	0	30	1	0	1	0	-3,0035	0,047268	0,952732
118	0	0	36	0	0	1	0	-2,9327	0,050561	0,949439
232	0	0	23	2	0	0	1	-2,8877	0,052765	0,947235

n=28 $\Sigma P(x_i) = 0,986$
 $\Sigma(1 - P(x_i)) = 27,310$

Ek-14. Sabit Denek Sayılı ve Sabit $P(x_i)$ Aralıklı Onlu Risk Grupları İçin Beklenen Frekansların Bulunması.

100	0	0	22	1	0	0	0	-2,8396	0,055221	0,944779
70	0	0	28	2	1	0	0	-2,8081	0,056888	0,943112
81	0	1	28	2	0	0	0	-2,7725	0,058828	0,941172
43	0	0	28	0	0	0	0	-2,7688	0,059034	0,940966
101	0	0	35	0	0	1	0	-2,7569	0,059698	0,940302
76	0	0	21	1	0	0	0	-2,6638	0,065144	0,934856
243	0	0	21	1	0	0	0	-2,6638	0,065144	0,934856
273	0	0	28	1	0	1	0	-2,6519	0,065872	0,934128
240	0	1	34	1	0	0	0	-2,5263	0,074035	0,925965
209	0	0	27	1	0	1	0	-2,4761	0,077551	0,922449
216	0	0	27	1	0	1	0	-2,4761	0,077551	0,922449
201	0	0	33	0	0	1	0	-2,4053	0,082769	0,917231
58	0	0	19	1	0	0	0	-2,3122	0,090118	0,909882
153	0	0	26	1	0	1	0	-2,3003	0,091098	0,908902
200	0	0	26	1	0	1	0	-2,3003	0,091098	0,908902
87	0	0	28	1	0	0	1	-2,2899	0,091963	0,908037
281	0	1	32	2	0	1	0	-2,2332	0,096808	0,903192
40	0	0	33	0	0	0	1	-2,2191	0,098048	0,901952
123	0	0	25	1	0	1	0	-2,1245	0,106738	0,893262
140	0	0	25	1	0	1	0	-2,1245	0,106738	0,893262
80	0	0	31	0	0	1	0	-2,0537	0,113679	0,886321
25	0	0	24	1	0	1	0	-1,9487	0,124695	0,875305
247	0	0	24	1	0	1	0	-1,9487	0,124695	0,875305
49	0	0	23	0	0	0	0	-1,8898	0,131267	0,868733
67	0	0	23	0	0	0	0	-1,8898	0,131267	0,868733
74	0	0	23	0	0	0	0	-1,8898	0,131267	0,868733
162	0	0	23	0	0	0	0	-1,8898	0,131267	0,868733
7	0	0	30	0	0	1	0	-1,8779	0,13263	0,86737
16	0	0	30	0	0	1	0	-1,8779	0,13263	0,86737
143	0	0	30	0	0	1	0	-1,8779	0,13263	0,86737
198	0	0	23	1	0	1	0	-1,7729	0,145182	0,854818
218	0	0	23	1	0	1	0	-1,7729	0,145182	0,854818
11	0	0	22	0	0	0	0	-1,714	0,152646	0,847354
31	0	0	22	0	0	0	0	-1,714	0,152646	0,847354
233	1	0	22	0	0	0	0	-1,714	0,152646	0,847354
253	0	0	22	0	0	0	0	-1,714	0,152646	0,847354
17	0	0	29	0	0	1	0	-1,7021	0,154191	0,845809
197	1	0	24	1	1	0	0	-1,6539	0,160583	0,839417
82	1	0	30	1	0	1	1	-1,5748	0,171533	0,828467
3	0	0	21	0	0	0	0	-1,5382	0,176797	0,823203
106	0	0	21	0	0	0	0	-1,5382	0,176797	0,823203
111	0	0	21	0	0	0	0	-1,5382	0,176797	0,823203
116	0	0	21	0	0	0	0	-1,5382	0,176797	0,823203
185	0	0	21	0	0	0	0	-1,5382	0,176797	0,823203
225	0	0	21	0	0	0	0	-1,5382	0,176797	0,823203
252	1	0	21	0	0	0	0	-1,5382	0,176797	0,823203
131	0	1	28	2	0	1	0	-1,53	0,177994	0,822006

$n=29$ $\Sigma P(x_i) = 2,14$
 $\Sigma(1-P(x_i)) = 26,850$

$n=28$ $\Sigma P(x_i) = 4,410$
 $\Sigma(1-P(x_i)) = 23,586$

$n=34$ $\Sigma P(x_i) = 5,258$
 $\Sigma(1-P(x_i)) = 28,74$

Ek-14. Sabit Denek Sayılı ve Sabit P(xi) Aralıklı Onlu Risk Grupları İçin Beklenen Frekansların Bulunması.

36	0	1	20	1	1	0	0	-1,5033	0,181934	0,818066
51	0	1	28	1	0	0	0	-1,4715	0,186715	0,813285
21	1	0	21	1	0	1	0	-1,4213	0,194458	0,805542
85	0	0	21	1	0	1	0	-1,4213	0,194458	0,805542
8	0	0	31	1	1	0	0	-1,397	0,198293	0,801707
13	0	0	20	0	0	0	0	-1,3624	0,203851	0,796149
72	1	0	20	0	0	0	0	-1,3624	0,203851	0,796149
88	0	0	20	0	0	0	0	-1,3624	0,203851	0,796149
93	0	0	20	0	0	0	0	-1,3624	0,203851	0,796149
142	0	0	20	0	0	0	0	-1,3624	0,203851	0,796149
163	1	0	20	0	0	0	0	-1,3624	0,203851	0,796149
224	0	0	20	0	0	0	0	-1,3624	0,203851	0,796149
234	0	0	20	0	0	0	0	-1,3624	0,203851	0,796149
235	0	0	20	0	0	0	0	-1,3624	0,203851	0,796149
125	0	0	27	0	0	1	0	-1,3505	0,205789	0,794211
132	0	0	27	0	0	1	0	-1,3505	0,205789	0,794211
4	0	1	25	1	1	0	0	-1,3198	0,210852	0,789148
175	0	1	27	1	0	0	0	-1,2957	0,21489	0,78511
161	0	0	20	1	0	1	0	-1,2455	0,22348	0,77652
211	0	0	20	1	0	1	0	-1,2455	0,22348	0,77652
44	0	1	28	1	1	0	0	-1,2097	0,229754	0,770246
177	0	0	19	0	0	0	0	-1,1866	0,233868	0,766132
114	1	1	26	2	0	1	0	-1,1784	0,23534	0,76466
96	0	0	26	0	0	1	0	-1,1747	0,236006	0,763994
270	0	1	27	2	0	0	1	-1,168	0,237217	0,762783
160	0	0	27	0	0	0	1	-1,1643	0,237887	0,762113
71	1	0	19	1	0	1	0	-1,0697	0,25546	0,74454
145	1	0	19	1	0	1	0	-1,0697	0,25546	0,74454
22	1	0	20	1	0	0	1	-1,0593	0,257443	0,742557
244	0	0	20	1	0	0	1	-1,0593	0,257443	0,742557
39	1	1	32	0	0	0	0	-1,0491	0,259398	0,740602
173	0	0	18	0	0	0	0	-1,0108	0,266823	0,733177
181	0	0	18	0	0	0	0	-1,0108	0,266823	0,733177
251	0	0	18	0	0	0	0	-1,0108	0,266823	0,733177
155	1	0	25	0	0	1	0	-0,9989	0,269158	0,730842
186	1	0	25	0	0	1	0	-0,9989	0,269158	0,730842
268	0	0	25	0	0	1	0	-0,9989	0,269158	0,730842
103	0	1	32	1	0	1	0	-0,9322	0,282479	0,717521
14	0	0	17	0	0	0	0	-0,835	0,302589	0,697411
236	0	0	17	0	0	0	0	-0,835	0,302589	0,697411
219	0	0	24	0	0	1	0	-0,8231	0,305106	0,694894
63	1	1	25	2	0	0	1	-0,8164	0,306528	0,693472
30	1	0	25	0	0	0	1	-0,8127	0,307315	0,692685
52	1	1	24	1	0	0	0	-0,7683	0,316847	0,683153
203	0	1	24	1	0	0	0	-0,7683	0,316847	0,683153
126	0	1	31	1	0	1	0	-0,7564	0,319428	0,680572
223	1	0	19	0	1	0	0	-0,7118	0,329201	0,670799

n=29 $\Sigma P(x_i) = 6.470$
 $\Sigma(1-P(x_i)) = 22.526$

n=33 $\Sigma P(x_i) = 7.705$
 $\Sigma(1-P(x_i)) = 25.295$

n=28 $\Sigma P(x_i) = 8.995$
 $\Sigma(1-P(x_i)) = 19.00$

Ek-14. Sabit Denek Sayılı ve Sabit $P(x_i)$ Aralıklı Onlu Risk Grupları İçin Beklenen Frekansların Bulunması.

245	1	0	19	0	1	0	0	-0,7118	0,329201	0,670799	$n=27$ $\Sigma P(x_i) = 9,317$ $\Sigma(1-P(x_i)) = 17,68$
45	0	0	20	0	1	0	0	-0,6751	0,337356	0,662644	
89	0	0	20	0	1	0	0	-0,6751	0,337356	0,662644	
105	1	0	20	0	1	0	0	-0,6751	0,337356	0,662644	
152	0	0	23	0	0	1	0	-0,6473	0,343598	0,656402	
277	1	0	23	0	0	1	0	-0,6473	0,343598	0,656402	
46	0	0	21	0	1	0	0	-0,6384	0,345608	0,654392	
97	0	0	21	0	1	0	0	-0,6384	0,345608	0,654392	
1	0	1	14	0	1	0	0	-0,5979	0,354824	0,645176	
274	0	1	30	1	0	1	0	-0,5806	0,358795	0,641205	
69	0	0	24	0	1	0	0	-0,5283	0,370913	0,629087	
79	0	0	25	0	1	0	0	-0,4916	0,379517	0,620483	
2	0	0	15	0	0	0	0	-0,4834	0,38145	0,61855	
27	1	0	15	0	0	0	0	-0,4834	0,38145	0,61855	
167	1	0	22	0	0	1	0	-0,4715	0,384261	0,615739	
267	1	0	22	0	0	1	0	-0,4715	0,384261	0,615739	
62	1	1	22	1	0	0	0	-0,4167	0,397307	0,602693	
226	0	0	24	1	1	1	0	-0,4114	0,398576	0,601424	
239	1	1	29	1	0	1	0	-0,4048	0,40016	0,59984	
64	0	0	28	0	1	0	0	-0,3815	0,405765	0,594235	
73	0	0	28	0	1	0	0	-0,3815	0,405765	0,594235	
59	0	1	20	0	1	0	0	-0,3777	0,406682	0,593318	
95	1	1	20	0	1	0	0	-0,3777	0,406682	0,593318	
204	1	1	20	0	1	0	0	-0,3777	0,406682	0,593318	
169	1	0	27	1	1	1	0	-0,3013	0,42524	0,57476	
66	0	0	21	0	0	1	0	-0,2957	0,426609	0,573391	
195	1	1	20	1	1	1	0	-0,2608	0,435167	0,564833	
193	0	0	29	1	1	1	0	-0,2279	0,44327	0,55673	
272	1	0	24	1	1	0	1	-0,2252	0,443937	0,556063	
278	0	0	30	1	1	1	0	-0,1912	0,452345	0,547655	
75	0	1	26	0	1	0	0	-0,1575	0,460706	0,539294	
110	1	0	20	0	0	1	0	-0,1199	0,470061	0,529939	$n=19$ $\Sigma P(x_i) = 8,368$ $\Sigma(1-P(x_i)) = 10,63$
191	0	0	20	0	0	1	0	-0,1199	0,470061	0,529939	
259	1	0	20	0	0	1	0	-0,1199	0,470061	0,529939	
265	0	0	20	0	0	1	0	-0,1199	0,470061	0,529939	
282	1	0	20	0	0	1	0	-0,1199	0,470061	0,529939	
54	1	1	27	1	1	1	0	-3,90E-03	0,499025	0,500975	
109	1	1	26	0	0	0	0	5,70E-03	0,501425	0,498575	
212	0	0	21	1	0	1	1	7,40E-03	0,50185	0,49815	
283	1	0	39	0	1	0	0	2,22E-02	0,50555	0,49445	
120	1	0	19	0	0	1	0	5,59E-02	0,513971	0,486029	
127	1	0	19	0	0	1	0	5,59E-02	0,513971	0,486029	
231	1	0	19	0	0	1	0	5,59E-02	0,513971	0,486029	$n=11$ $\Sigma P(x_i) = 5,797$ $\Sigma(1-P(x_i)) = 5,202$
269	1	0	19	0	0	1	0	5,59E-02	0,513971	0,486029	
284	1	0	40	0	1	0	0	5,89E-02	0,514721	0,485279	
237	0	0	18	0	0	1	0	0,2317	0,557667	0,442333	
68	0	1	25	1	0	1	0	0,2984	0,574051	0,425949	

Ek-14. Sabit Denek Sayılı ve Sabit P(xi) Aralıklı Onlu Risk Grupları İçin Beklenen Frekansların Bulunması.

53	0	0	14	0	1	1	0	0,3472	0,585938	0,414062	$n=18$ $\Sigma P(x_i) = 11.602$ $\Sigma(1-P(x_i)) = 6.3977$
190	1	0	18	0	0	0	1	0,4179	0,602981	0,397019	
214	0	0	18	0	0	0	1	0,4179	0,602981	0,397019	
147	0	1	24	1	0	1	0	0,4742	0,616377	0,383623	
275	1	0	19	0	1	1	0	0,5307	0,629646	0,370354	
279	0	0	19	0	1	1	0	0,5307	0,629646	0,370354	
35	1	0	20	0	1	1	0	0,5674	0,638163	0,361837	
37	1	0	20	0	1	1	0	0,5674	0,638163	0,361837	
60	1	0	20	0	1	1	0	0,5674	0,638163	0,361837	
128	1	0	20	0	1	1	0	0,5674	0,638163	0,361837	
133	1	0	20	0	1	1	0	0,5674	0,638163	0,361837	
227	1	0	20	0	1	1	0	0,5674	0,638163	0,361837	
257	1	0	20	0	1	1	0	0,5674	0,638163	0,361837	
168	0	0	22	0	1	1	0	0,6408	0,654934	0,345066	
154	0	1	22	0	0	0	0	0,7089	0,670158	0,329842	
136	0	0	24	0	1	1	0	0,7142	0,671329	0,328671	
222	1	0	20	0	1	0	1	0,7536	0,679963	0,320037	
178	0	0	15	0	0	1	0	0,7591	0,681158	0,318842	$n=10$ $\Sigma P(x_i) = 7.319$ $\Sigma(1-P(x_i)) = 2.688$
241	1	1	19	0	1	1	0	0,8281	0,695953	0,304047	
65	1	1	20	0	1	1	0	0,8648	0,703663	0,296337	
94	1	1	20	0	1	1	0	0,8648	0,703663	0,296337	
196	1	1	20	0	1	1	0	0,8648	0,703663	0,296337	
176	1	0	22	0	0	1	1	0,9572	0,722561	0,277439	
129	0	0	23	1	1	1	1	0,9806	0,727227	0,272773	
157	1	0	24	1	1	1	1	1,0173	0,734446	0,265554	
56	0	1	25	0	1	1	0	1,0483	0,740448	0,259552	
158	1	1	20	0	0	0	0	1,0605	0,742786	0,257214	
15	0	0	21	0	0	1	1	1,133	0,756392	0,243608	
146	0	1	26	2	0	1	0	1,2482	0,776988	0,223012	
33	1	1	25	0	0	1	0	1,424	0,805965	0,194035	$n=16$ $\Sigma P(x_i) = 14.029$ $\Sigma(1-P(x_i)) = 1.971$
255	1	1	23	0	0	1	0	1,7756	0,855153	0,144847	
24	1	0	14	0	1	1	1	1,7759	0,85519	0,14481	
5	1	0	18	0	1	1	1	1,9227	0,872439	0,127561	
50	1	1	15	0	0	0	0	1,9395	0,874297	0,125703	
23	1	0	20	0	1	1	1	1,9961	0,880387	0,119613	
84	0	0	20	0	1	1	1	1,9961	0,880387	0,119613	
112	1	0	20	0	1	1	1	1,9961	0,880387	0,119613	
137	1	0	20	0	1	1	1	1,9961	0,880387	0,119613	
148	1	0	20	0	1	1	1	1,9961	0,880387	0,119613	
258	0	0	20	0	1	1	1	1,9961	0,880387	0,119613	
121	1	0	24	0	1	1	1	2,1429	0,895003	0,104997	
122	1	0	24	0	1	1	1	2,1429	0,895003	0,104997	
130	1	0	24	0	1	1	1	2,1429	0,895003	0,104997	
12	1	0	15	0	0	1	1	2,1878	0,899149	0,100851	
29	1	0	15	0	0	1	1	2,1878	0,899149	0,100851	

Ek-14. Sabit Denek Sayılı ve Sabit P(xi) Aralıklı Onlu Risk Grupları İçin Beklenen Frekansların Bulunması.

77	0	1	19	0	1	1	1	2,2568	0,905235	0,094765
19	1	1	20	0	1	1	1	2,2935	0,908337	0,091663
202	1	1	19	0	0	0	1	2,665	0,93493	0,06507
164	0	1	17	0	0	1	0	2,8304	0,944297	0,055703
9	1	1	19	0	0	1	1	3,9075	0,980305	0,019695
113	1	1	19	0	0	1	1	3,9075	0,980305	0,019695

$N=6$ $\sum P(x_i) = 5,653$
 $\sum (1-P(x_i)) = 0,347$

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

85372