

T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

SIRALI (ORDINAL) VE MULTINOMIAL
LOGIT MODELLER
ÜZERİNE BİR UYGULAMA

N. Anıl BARAK

Biyoistatistik Programı
YÜKSEK LİSANS TEZİ

ANKARA

2005

T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

SIRALI (ORDINAL) VE MULTINOMIAL
LOGIT MODELLER
ÜZERİNE BİR UYGULAMA

N. Anıl BARAK

Biyoistatistik Programı
YÜKSEK LİSANS TEZİ

TEZ DANIŞMANI
Doç. Dr. Osman SARAÇBAŞI

ANKARA
2005

TEŞEKKÜR

Çalışmalarım boyunca değerli yardım ve katkılarıyla beni yönlendiren, her aşamasında özellikle de yaz aylarının sıcağında internet üzerinden desteğini esirgemeyen danışman hocam, Sayın Doç. Dr. Osman Saraçbaşı'na teşekkürlerimi sunarım. Çalışmasına ait verileri kullanma olanağı tanıdığı için Uzm. Dr. Meltem Halil ve Prof. Dr. Servet Arıoğlu'na, tez çalışmalarım sırasında göstermiş oldukları anlayış, sabır ve desteklerinden dolayı Biyoistatistik Anabilim Dalındaki değerli tüm hocalarıma ve arkadaşlarıma içtenlikle teşekkür ederim. Son olarak annem, babam ve ablama, sevgili köpeğim Matilda'ya, hayata şu sıralarda gözünü açacak olan sevgili yeğenim Can'a, sevgili dostlarım Özge Kalyoncu ve Özgür Dolgun'a sonsuz teşekkürler.

ÖZET

Barak, N. A. Sıralı ve *Multinomial Logit* Modeller Üzerine Bir Uygulama. Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Programı Yüksek Lisans Tezi, Ankara, 2005. Sıralı *Logit* modelleri, çok kategorili ve sıralı yapıdaki bağımlı değişken olasılıklarını bağımsız değişkenler yardımı ile tahmin etmek için kullanılmaktadır. Sıralı *Logit* modellerinde, bağımlı değişken kategorileri karşılaştırmaları için kullanılan farklı *Logit* oluşturma biçimleri bulunmaktadır. Bunlar içerisinde Yığılımlı *Logit* modelleri, Orantısız Risk varsayımının sağlandığı ve sağlanmadığı durumlar için iki ana başlık altında incelenmektedir. Bunlar Orantısız Risk modeli ve Orantısız Olmayan Risk modelidir. Orantısız Risk Modelinde, *Multinomial Logit* modelinden farklı olarak, oluşturulan Yığılımlı *Logit*'lerin paralellik varsayımı bulunmaktadır. Orantısız Risk varsayımı adı verilen bu varsayım çoğu zaman sağlanmamaktadır. Bu varsayım sağlanmadığında yapılabilecek bir diğer çözümleme de *Multinomial Logit* çözümlemesidir. Ancak, *Multinomial Logit* çözümlemede bağımlı değişkenin sıralı yapısı göz ardı edilmekte ve nominal olarak modele katılmaktadır. Dolayısı ile hem sıralı yapıyı dikkate alan hem de oldukça katı olan Orantısız Risk varsayımını rahatlatan bir model arayışına girilmiştir. Son yıllarda Genelleştirilmiş Sıralı *Logit* model veya diğer adıyla Orantısız Olmayan Risk Modeli, paralellik varsayımının sağlanmadığı ve Orantısız Risk Modelinin yetersiz kaldığı durumlarda kullanılmaktadır. Bu çalışmanın amacı, Orantısız Risk varsayımının sağlanmadığı durumlarda Orantısız Olmayan Risk modeli ile *Multinomial Logit* modelini karşılaştırmak ve bu modeller arasından verilerin yapısına ve varsayımlarına göre en uygun modeli belirlemektir. Bu amaç doğrultusunda hipotez; bağımlı değişkenin sıralı yapıda olduğu ve Orantısız Risk varsayımının sağlanmadığı durumlarda Orantısız Olmayan Risk modelinin, *Multinomial Logit* modelinden daha iyi bir seçenek olduğudur. Bu hipotezi test etmek için Hacettepe Üniversitesi İç Hastalıkları Anabilim Dalı Geriatri Ünitesine ait veriler kullanılmış ve incelenen modellerin En Çok Olabilirlik tahmin edicileri ile *Odds* oranları elde edilmiştir. Modellerin geçerliliği Olabilirlik Oranı test istatistiği ile test edilmiş ve modellerin uyum iyiliği göstergeleri hesaplanmıştır. Çalışmada Orantısız Risk varsayımının sağlanmadığı durumlarda Orantısız Olmayan Risk modelinin kullanımının, bağımlı değişkenin yapısındaki tüm bilgiyi kullanması açısından avantaj sağladığı; dolayısı ile Orantısız Risk varsayımının sağlanmadığı durumlarda *Multinomial Logit* modeline tercih edilmesi gerektiği sonucuna varılmıştır.

Anahtar Kelimeler: Yığılımlı *Logit* Model, Orantısız Risk Modeli, Orantısız Olmayan Risk Modeli, *Multinomial Logit* Model.

ABSTRACT

Barak, N. A. An Application of Ordinal and Multinomial Logit Models. Hacettepe University Institute of Health Sciences, Master Thesis in Biostatistics, Ankara, 2005. Ordinal Logit models are used to estimate the probability of multicategory ordinal dependent variable, when the set of independent variables are given. There are several ways of constructing the Logits to compare the dependent categories in Ordinal Logit models. One of the most widely used Logit is called Cumulative Logit model which is defined as Proportional Odds model and Non-Proportional Odds model. In Proportional Odds model, there is an assumption of parallel Cumulative Logits, known as the Parallel Slopes assumption or Proportional Odds assumption. But in some cases this assumption does not hold. In such cases employing the Multinomial Logit model ignores the ordinal nature of dependent variable. Recently, Non-Proportional Odds model, in other words Generalized Ordinal Logit model that both takes into account ordinal structure and relaxes the strict assumption of proportional odds has been developed. The aim of this study is to compare Non-Proportional Odds model with Multinomial Logit model when the assumption of proportional odds is violated, also to determine the best model regarding the structure and assumptions of the data. The hypothesis that has been investigated in this study is; Non-Proportional Odds model is preferable to Multinomial Logit model when the dependent variable is ordinal and the assumption of proportional odds does not hold. The data was taken from Hacettepe University Faculty of Medicine Internal Medicine Department to test this hypothesis. Model parameter estimates, odds ratios, likelihood ratio test statistics and goodness of fit statistics were calculated. At the end of this study it has been concluded that Non-Proportional Odds model should be preferred to Multinomial Logit model when the assumption of proportional odds does not hold as Non-Proportional Odds model is more advantageous than the Multinomial Logit model in utilizing the information provided by the ordinal dependent variable.

Key Words: Cumulative Logit Model, Proportional Odds Model, Non-Proportional Odds Model, Multinomial Logit Model.

İÇİNDEKİLER

	Sayfa
ONAY SAYFASI	iii
TEŞEKKÜR.....	iv
ÖZET.....	v
ABSTRACT.....	vi
İÇİNDEKİLER	vii
SİMGELER ve KISALTMALAR DİZİNİ	ix
ŞEKİLLER DİZİNİ.....	x
TABLOLAR DİZİNİ	xi
1. GİRİŞ	1
2. GENEL BİLGİLER	4
2.1. Genelleştirilmiş Doğrusal Modeller	4
2.1.1. Genelleştirilmiş Doğrusal Modellerin Bileşenleri	4
2.1.2. Genelleştirilmiş Doğrusal Modellerde Bağ Fonksiyonları	6
3. LOGIT MODELLER	9
3.1. İki Durumlu <i>Logit</i> Model	9
3.1.1. Gizli Değişken Yaklaşımı ile İki Durumlu <i>Logit</i> Modelin Türetilmesi	10
3.1.2. Katsayıların ve <i>Odds</i> Oranlarının Yorumlanması.....	13
3.2. <i>Multinomial Logit</i> Model	14
3.2.1. İlgisiz Alternatiflerin Bağımsızlığı Varsayımı.....	16
3.2.2. Katsayıların ve <i>Odds</i> Oranlarının Yorumlanması.....	17
3.3. Sıralı <i>Logit</i> Modelleri.....	18
3.3.1. Gizli Değişken Yaklaşımı ile Sıralı <i>Logit</i> Modellerinin Türetilmesi	18
3.3.2. Yığılımlı <i>Logit</i> Modelleri.....	21
3.3.2.1. Orantısal Risk Modeli ve Orantısal Risk Varsayımı.....	22
3.3.2.2. Orantısal Olmayan Risk Modeli	26
3.3.2.3. Orantısal Risk Varsayımının Geçerliliğinin Testi.....	30

İÇİNDEKİLER (DEVAM)

	Sayfa
3.4. Model Geçerlilik Testi ve Uyum İyiliği Göstergeleri	31
3.4.1. Olabilirlik Oranı Test İstatistiği	31
3.4.2. Sapma Ölçüsü	33
3.4.3. Akaike Bilgi Kriteri.....	34
3.4.4. Bayesci Bilgi Kriteri	34
3.4.5. McFadden’ın Düzeltilmiş Olabilirlik Oranı Göstergesi.....	35
4. GEREÇ ve YÖNTEM.....	36
5. BULGULAR.....	39
5.1. Tanımlayıcı İstatistikler	39
5.2. <i>Multinomial Logit</i> Model	42
5.3. Orantısal Risk Modeli	47
5.4. Orantısal Olmayan Risk Modeli.....	49
6. TARTIŞMA ve SONUÇ	54
KAYNAKLAR	62

SİMGELER VE KISALTMALAR

AIC	Akaike Bilgi Kriteri
BIC	Bayesci Bilgi Kriteri
CHD	Koroner Kalp Hastalığı
Crp	C-reaktif Protein
D	Sapma Ölçüsü
EÇOK	En Çok Olabilirlik Kestirim Yöntemi
EKK	En Küçük Kareler Kestirim Yöntemi
ε	Hata Terimi
F	F test İstatistiği
F(.)	Dağılım Fonksiyonu
Fr	Ferritin
FRS	Framingham Risk Skoru
g(.)	Bağ Fonksiyonu
G^2	Olabilirlik Oranı Test İstatistiği
G_G^2	Genelleştirilmiş Olabilirlik Oranı Test İstatistiği
GDM	Genelleştirilmiş Doğrusal Modeller
Hs	Homosistein
IIA	İlgisiz Alternatiflerin Bağımsızlığı Varsayımı
Kr	Kreatinin
L	Log-Olabilirlik
Ldl	<i>Low-density lipoprotein</i>
ξ	Gizli Değişken
OR	<i>Odds</i> Oranı
π	Olgunun Ortaya Çıkma Olasılığı
$\bar{R}_{McFadden}^2$	McFadden'ın Düzeltilmiş Olabilirlik Oranı Göstergesi
sd	Serbestlik Derecesi
Tg	Trigliserit
τ	Eşik Değeri
Ür	Ürikasit
χ^2	Ki- Kare Test İstatistiği

ŞEKİLLER

Sayfa

3.1. J kategorili sıralı bağımlı değişken için gizli değişken ξ ve gözlenen y değişkeninin gösterimi.	19
3.2. Üç düzeyli sıralı bağımlı değişken için oluşturulan Yığılımlı <i>Logit</i> 'lerin paralelliğinin grafiksel olarak gösterimi.	24
3.3. Üç düzeyli sıralı bağımlı değişken için oluşturulan Yığılımlı <i>Logit</i> 'lerin paralelliğinin bozulunun grafiksel olarak gösterimi.	29
5.1. Düşük Riske karşı Orta Riski karşılaştıran <i>Logit</i> için elde edilen <i>Odds</i> Oranları ve % 95 güven aralıkları.	44
5.2. Düşük Riske karşı Yüksek Riski karşılaştıran <i>Logit</i> için elde edilen <i>Odds</i> Oranları ve % 95 güven aralıkları.	46
5.3. Düşük Riske karşı Orta ve Yüksek Riski karşılaştıran <i>Logit</i> için elde edilen <i>Odds</i> Oranları ve % 95 güven aralıkları.	51
5.4. Düşük ve Orta Riske karşı Yüksek Riski karşılaştıran <i>Logit</i> için elde edilen <i>Odds</i> Oranları ve % 95 güven aralıkları.	53
6.1. <i>Multinomial Logit</i> ve Orantısız/Orantısız Olmayan Risk modelinde referans ve karşılaştırılan kategori/kategorilerin gösterimi.	60

TABLOLAR

	Sayfa
4.1. Çalışmada Kullanılan Bağımsız Değişkenler ve Düzeyleri.	37
5.1. Yaş, <i>Low-density lipoprotein</i> (Ldl), C-reaktif protein (Crp), Homosistein (Hs), Ferritin (Fr), Trigliserit (Tg), Ürikasit (Ür) ve Kreatinin (Kr) Değerlerinin Tanımlayıcı İstatistikleri.	39
5.2. Framingham Risk Skoru Frekans Dağılımı.	40
5.3. Framingham Risk Skoru gruplarında Yaş, Ldl, Crp, Hs, Fr, Tg, Ür ve Kr Değerlerinin Tanımlayıcı İstatistikleri.	40
5.4. Ldl, Crp, Hs, Fr, Tg, Ür ve Kr Düzeylerinin Frekans Dağılımları.	41
5.5. <i>Multinomial Logit</i> Modelinin Katsayı, Standart Hata ve <i>Odds</i> Oranı Kestirimleri (n=996).	42
5.6. <i>Multinomial Logit</i> Modelinin Uyum İyiliği Göstergeleri.	43
5.7. Orantısal Risk Modelinin Katsayı, Standart Hata ve <i>Odds</i> Oranı Kestirimleri.	47
5.8. Orantısal Risk Modelinin Uyum İyiliği Göstergeleri.	48
5.9. Orantısal Risk Varsayımının Geçerliliğinin Testi (Brant Testi).	48
5.10. Orantısal Olmayan Risk Modelinin Katsayı, Standart Hata ve <i>Odds</i> Oranı Kestirimleri.	49
5.11. Orantısal Olmayan Risk Modelinin Uyum İyiliği Göstergeleri.	50

TABLÖLAR (DEVAM)**Sayfa**

6.1. Orantısız Risk ve Orantısız Olmayan Risk Modelindeki <i>Odds</i> Oranı Kestirimlerinin Karşılaştırılması.	55
6.2. Orantısız Risk ve Orantısız Olmayan Risk Modelinin Uyum İyiliği Göstergeleri Açısından Karşılaştırılması.	56
6.3. Orantısız Olmayan Risk ve <i>Multinomial Logit</i> Modelindeki <i>Odds</i> Oranı Kestirimlerinin Karşılaştırılması.	57
6.4. <i>Multinomial Logit</i> ve Orantısız Olmayan Risk Modelinin Uyum İyiliği Göstergeleri Açısından Karşılaştırılması.	59

BİRİNCİ BÖLÜM

1. GİRİŞ

İstatistiksel yöntemler içerisinde kategorik verilerin çözümlenmesi geniş bir yer tutar. Bu amaçla geliştirilmiş birçok yöntem vardır. Bu yöntemlerden *Logit* modeller, bir bağımlı ve bir ya da birden çok bağımsız değişkenin bulunduğu çok gözlü çapraz tabloların çözümlenmesinde kullanılır. *Logit* modellerde, kategorik bağımlı değişkene ilişkin *odds*'un doğal logaritması; bağımsız değişkenlerin doğrusal bir fonksiyonu olarak ifade edilir. Bu yönüyle *Logit* modeller Genelleştirilmiş Doğrusal Modeller ailesinin bir parçasıdır ve bağ (link) fonksiyonu olarak *Logit* dönüşümü, diğer bir ifade ile; kategorik bağımlı değişken *odds*'unun doğal logaritması kullanılır. *Logit* çözümler içerisinde İki durumlu (Binary) *Logit* çözümleme iki durumlu kategorik bağımlı değişkeni, açıklayıcı değişkenler kümesinin doğrusal bir fonksiyonu olarak modellemede kullanılır. *Multinomial Logit* çözümleme ise, en az 3 kategoriden oluşan sınıflanabilir (nominal) ve sıralama içermeyen bağımlı değişkeni, açıklayıcı değişkenler kümesinin doğrusal bir fonksiyonu olarak modellemede kullanılır. Bağımlı değişkenin sıralı (ordinal) olduğu durumlarda ise, Sıralı *Logit* çözümlemesi kullanılmaktadır.

Sağlık alanında yapılan çalışmaların çoğunda incelenen bağımlı değişken kategorik ve sıralı durumundadır. Hastalık şiddetini (az-orta-şiddetli) veya bir tümör derecesini (Grade I-II-III), incelenen bağımsız değişkenler yardımıyla kestirmek isteyen bir çalışma bu duruma örnek olarak verilebilir. Bu gibi çalışmalarda araştırmacıların sıklıkla yaptığı hatalardan biri, kategorik bağımlı değişkenin sıralı yapıda olduğunu unutarak *Multinomial Logit* çözümleme yardımı ile, kategorik bağımlı değişken olasılıklarını hesaplamasıdır. Bu gibi durumlarda *Multinomial Logit* çözümleme yerine, sıralı yapıyı dikkate alan Sıralı *Logit* çözümlemesi kullanılmaktadır (1).

Sıralı *Logit* çözümlemesinde ise bağımlı değişken karşılaştırmalarında kullanılan farklı *Logit* oluşturma biçimleri vardır. Bunlardan yorumlanması en kolay

olan ve en sık kullanılan Yığılımlı (Cumulative) *Logit*'lerdir. Yığılımlı *Logit* modelleri, Orantısal Risk Modeli (Proportional Odds Model) ve Orantısal Olmayan Risk Modeli (Non- Proportional Odds Model) olmak üzere ikiye ayrılmaktadır. Orantısal Risk Modelinde, *Multinomial Logit* modelinden farklı olarak, oluşturulan Yığılımlı *Logit*'lerin paralellik varsayımı (Proportional Odds Assumption) bulunmaktadır. Orantısal Risk varsayımı adı verilen bu varsayım çoğu zaman sağlanmamaktadır. Bu varsayımın sağlanmadığı durumlarda Orantısal Risk Modelinin kullanılması ise hatalı olmaktadır (2). Diğer yandan; bu varsayım sağlanmadığında yapılabilecek bir diğer çözümleme de *Multinomial Logit* çözümlemesidir. Ancak, *Multinomial Logit* çözümlemede bağımlı değişkenin sıralı yapısı göz ardı edilmekte ve nominal olarak modele katılmaktadır (2). Dolayısı ile hem sıralı yapıyı dikkate alan hem de oldukça katı olan orantısal risk varsayımını rahatlatan bir model arayışına girilmiştir. Son yıllarda Genelleştirilmiş Sıralı *Logit* model veya diğer adıyla Orantısal Olmayan Risk Modeli, paralellik varsayımının sağlanmadığı ve Orantısal Risk Modelinin yetersiz kaldığı durumlarda kullanılmaktadır (3).

Yukarıda belirtilen nedenlerden dolayı bu çalışmanın amacı, bağımlı değişken kategorisinin ikiden fazla olması durumunda geliştirilmiş olan Multinomial ve Sıralı Logit modellerini ayrıntılı olarak ele almak, Orantısal Risk varsayımının sağlanmadığı durumlarda Orantısal Olmayan Risk modeli ile Multinomial Logit modelini karşılaştırmak ve bu modeller arasından verilerin yapısına ve varsayımlarına göre en uygun modeli belirlemektir.

Bu amaç doğrultusunda hipotez; bağımlı değişkenin sıralı yapıda olduğu ve Orantısal Risk Varsayımının sağlanmadığı durumlarda Orantısal Olmayan Risk modelinin, Multinomial Logit modelinden daha iyi bir seçenek olduğudur. Bu hipotezi test etmek için sağlık alanında bir veri kullanılacak ve incelenen modellerin en çok olabilirlik kestiricileri ve *Odds* oranları elde edilecektir. Modellerin geçerliliği Olabilirlik Oranı test istatistiği ile test edilirken, Uyum İyiliği (Goodness of Fit) göstergesi olarak ise modellerin Akaike Bilgi Kriteri (AIC), Bayesci Bilgi Kriteri (BIC) ve McFadden'ın Olabilirlik Oranı göstergesi kullanılacaktır. Böylelikle bu

alışmanın sonunda uygun *Logit* modelinin seiminde gz nne alınması gereken ltler belirlenmiř olacaktır.

alışmanın ikinci blmnde Genelleřtirilmiř Doėrusal Modeller (Generalized Linear Models) ile ilgili bilgiler yer almaktadır. nc blmde kategorik baėımlı deėiřkenin iki durumlu olması durumunda kullanılan İki durumlu *Logit* model, ok kategorili ve nominal olması durumunda kullanılan *Multinomial Logit* model, ok kategorili ve sıralı olması durumunda kullanılan Sıralı *Logit* modeller ayrıntılı olarak incelenmektedir. Drdnc blmde *Multinomial* ve Sıralı *Logit* modelleri ile ilgili uygulamanın bulgularına ve beřinci blmde ise sonu ve tartiřmalara yer verilmektedir.

İKİNCİ BÖLÜM

2. GENEL BİLGİLER

2.1. Genelleştirilmiş Doğrusal Modeller (Generalized Linear Models)

Sağlık alanında yapılan çalışmaların çoğunda bağımlı değişken; yapılan bir testin sonucu (pozitif-negatif), hastalık tipi (diyabet-hipertansiyon-koroner hastalıklar) veya hastalık şiddeti (az-orta-şiddetli) gibi normal dağılım göstermeyen, iki düzeyli (binary), sınıflandırılmış (nominal) veya sıralı (ordinal) ölçek ile ölçülmüş olabilir. Bu tür verilerin çözümlenmesinde birçok varsayım bozulduğundan dolayı Klasik Doğrusal Modelleri (Ordinary Linear Models) yetersiz kalmaktadır. Genelleştirilmiş Doğrusal Modeller (GDM) klasik doğrusal modellerinin genişlemiş bir durumudur. Normal dağılım göstermeyen bağımlı değişkeni ve ortalama fonksiyonlarını doğrusal olarak modellemeyi de içine almaktadır (4).

2.1.1. Genelleştirilmiş Doğrusal Modellerin Bileşenleri

Genelleştirilmiş Doğrusal Modeller üç bileşenden oluşmaktadır. Rasgele bileşen (random component), bağımlı değişkeni ve bağımlı değişkenin olasılık dağılımını tanımlamaktadır. Sistemik bileşen (systematic component), bağımsız değişkenlerin doğrusal fonksiyonu ile belirlenir. Bağ (Link) ise rasgele bileşenin beklenen değeri ile sistemik bileşen arasındaki fonksiyonel ilişki olarak tanımlanmaktadır (4).

N genişliğinde bir örneklem için $y_1, \dots, y_n$ üstel aileden gelen bir dağılıma (Normal, Poisson...vb.) sahip birbirinden bağımsız gözlemler olduğunda Genelleştirilmiş Doğrusal Modelinin rasgele bileşeni, Y bağımlı değişken için aşağıdaki gibi bir olasılık yoğunluk fonksiyonu olarak tanımlanır (4).

$$f(y_i; \theta_i) = a(\theta_i)b(y_i) \exp[y_i Q(\theta_i)] \quad (2.1)$$

Poisson, Binom dağılımı gibi birçok önemli dağılım Eşitlik 2.1'deki formda yazılabilmektedir. Burada θ_i , bağımsız değişkenin ($i = 1, \dots, N$) değerlerine göre değerler alırken, $Q(\theta)$ terimi doğal parametre (natural parameter) olarak adlandırılır (4).

Genelleştirilmiş Doğrusal Modellerin sistematik bileşeni ($\eta_1, \dots, \eta_N$); modeldeki bağımsız değişkenlerin doğrusal bir fonksiyonu olarak ifade edilir. x_{ij} j 'nci ($j = 1, \dots, p$) bağımsız değişkenin i 'nci ($i = 1, \dots, N$) değerini göstermek üzere sistematik bileşen aşağıdaki gibi tanımlanır:

$$\eta_i = \sum_j \beta_j x_{ij} \quad (2.2)$$

Bağımsız değişkenlerin bu doğrusal kombinasyonu, doğrusal kestirici (linear predictor) olarak da adlandırılmaktadır. Genellikle tüm i 'ler için tek bir $x_{ij} = 1$ değeri modelde sabit terim (çoğu zaman α ile gösterilir) olarak bulunur (4).

Genelleştirilmiş doğrusal modellerin üçüncü bileşeni, rasgele bileşen ile sistematik bileşen arasındaki bağıdır. Bağ fonksiyonu (link function); bağımlı değişken ortalamasını ($\mu_i = E(Y_i)$), bağımsız değişkenlerin doğrusal kombinasyonuna (η_i) bağlayan fonksiyondur. En genel ifade ile genelleştirilmiş doğrusal modeller aşağıdaki gibi gösterilmektedir:

$$g(\mu_i) = \eta_i = \sum_j \beta_j x_{ij}, \quad i = 1, \dots, N \quad (2.3)$$

Eşitlikteki $g(\cdot)$ bağ fonksiyonu, monoton diferansiyelenebilir bir fonksiyon olmaktadır (4, 5). Bağ fonksiyonu, genelleştirilmiş doğrusal modeller ailesinin bir üyesini diğer üyesinden ayırt etmektedir. Genelleştirilmiş doğrusal modellerde kullanılan birçok bağ fonksiyonu bulunmaktadır (6). Bu çalışmada, bağ fonksiyonlarından yalnızca birkaç tanesinden söz edilecektir.

Özet olarak genelleştirilmiş doğrusal modeller aşağıdaki özelliklere sahiptir:

- i) Bağımlı değişken üstel aileden olmak üzere aynı dağılıma sahip birbirinden bağımsız raslantı değişkenleridir.
- ii) Bağımlı değişkeni, bağımsız değişkenlere doğrusal bir yapı ile bağlayan, monoton ve diferansiyelenebilir bir bağ fonksiyonu vardır (5).

2.1.2. Genelleştirilmiş Doğrusal Modellerde Bağ Fonksiyonları

Genelleştirilmiş doğrusal modellerde bağ fonksiyonu rasgele bileşenin dağılımı tarafından belirlenmektedir. GDM’de rasgele bileşen dağılımı, üstel aileden (Normal, Binom, Poisson,...vb.) gelen bir dağılımdır. Dolayısı ile her bir dağılım için farklı bağ fonksiyonları bulunmaktadır (6).

2.1.2.1. Özdeş Bağ (Identity Link) Fonksiyonu

Rasgele bileşenin (bağımlı değişkenin) normal dağılım göstermesi durumunda klasik doğrusal modellerin bağ olarak kullandığı fonksiyondur. Özdeş bağ fonksiyonu $g(\mu) = \mu$ şeklindedir ve bu fonksiyon için $\eta_i = \mu_i$ ’dir. Diğer bir ifade ile; özdeş bağ fonksiyonu, bağımlı değişken ortalamasının kendisi için doğrusal model tayin eder. Bağımlı değişken normal dağılım göstermesi durumunda klasik doğrusal modeller bu yapı ile aşağıdaki gibi oluşturulur (4).

$$\mu_i = \sum_j \beta_j x_{ij} \quad i = 1, \dots, N \quad (2.4)$$

Bu modelde parametre kestirimleri En Küçük Kareler Kestirim Yöntemi (EKK) (Ordinary Least Square Estimation) ile elde edilmektedir. Klasik doğrusal modellerde, bağımlı değişkenin normal dağılım göstermediği durumlarda ise EKK kestirim yöntemi yetersiz kalmaktadır.

2.1.2.2. *Logit* Bağ (Logit Link) Fonksiyonu

Bağımlı değişken her zaman normal dağılım göstermeyebilir. Genel olarak, kategorik bağımlı değişken bir olgunun ortaya çıkması ($y = 1$) ve çıkmaması ($y = 0$) gibi iki durumlu (binary) olduğunda bağımlı değişken Binom dağılımının $n = 1$ için özel bir durumu olan Bernoulli dağılımı göstermektedir (4). Bernoulli dağılımında olgunun ortaya çıkma olasılığı $P(y = 1) = \pi$, ve çıkmama olasılığı $P(y = 0) = 1 - \pi$ ile ifade edilirken; Bernoulli dağılımında beklenen değer $E(Y) = \pi$ 'dir. Rasgele bileşenin Binom dağılımı gösterdiği durumlarda kullanılan bağ fonksiyonu *Logit* bağ fonksiyonudur ve aşağıdaki gibi tanımlanır:

$$g(\mu) = \eta = \log[\mu / (1 - \mu)] \quad (2.5)$$

Eşitlik 2.5'de beklenen değerler olasılık cinsinden yazılırsa *Logit* bağ için aşağıdaki ifadeye ulaşılır:

$$g(\mu) = \eta = \log[\pi / (1 - \pi)] \quad (2.6)$$

Eşitlik 2.6'da $\pi / (1 - \pi)$ ifadesi bağımlı değişken *odds*'u olarak tanımlanır. *Logit* bağ, kategorik bağımlı değişkene ilişkin *odds*'un doğal logaritması olarak da ifade edilir. *Logit* bağ fonksiyonunu kullanan GDM'e ise *Logit* modeller denilmektedir (4).

2.1.2.3. *Multinomial Logit* Bağ (Multinomial Logit Link) Fonksiyonu

Bağımlı kategorik değişkenin J tane düzeyinin olduğu durumda bağımlı değişkenin dağılımı *Multinomial* dağılımı göstermektedir. Rasgele bileşenin *Multinomial* dağılımı gösterdiği durumlarda kullanılan bağ fonksiyonu *Multinomial Logit* bağ fonksiyonudur ve aşağıdaki gibi tanımlanır (6):

$$\eta_j = \log(\mu_j / \mu_J) \quad j = 1, \dots, J \quad (2.7)$$

Multinomial Logit bağ fonksiyonu, iki düzeyli bağımlı değişken için kullanılan *Logit* bağ fonksiyonunun genişletilmiş bir durumudur. Bağımlı değişken kategorilerinden bir tanesi referans kategori (genellikle J ile gösterilir) olarak alınır.

Eşitlik 2.7’de beklenen değerler olasılık cinsinden yazılırsa *Multinomial Logit* bağ için aşağıdaki eşitliğe ulaşılır:

$$\eta_j = \log(\pi_j / \pi_J) \quad j = 1, \dots, J \quad (2.8)$$

Eşitlik 2.8’de π_j / π_J ifadesi bağımlı değişkenin referans kategori için *odds*’u olarak tanımlanır. *Multinomial Logit* bağ, kategorik bağımlı değişken kategorisinin referans kategoriye karşı *odds*’unun doğal logaritması olarak da ifade edilir. *Multinomial Logit* bağ fonksiyonunu kullanan genelleştirilmiş doğrusal modellere ise *Multinomial Logit* veya Referans Kategori (Baseline Category) *Logit* modeller adı verilmektedir (4).

ÜÇÜNCÜ BÖLÜM

3. LOGIT MODELLER (LOGIT MODELS)

Logit modeller, bir bağımlı ve bir ya da birden çok bağımsız değişkenin bulunduğu çok gözlü çapraz tabloların çözümlenmesinde kullanılmaktadır. *Logit* modellerde amaç; bağımlı değişken kategori olasılıklarını, modeldeki bağımsız değişkenlerce açıklamaktır. Genelleştirilmiş Doğrusal Modeller ailesinin bir üyesi olan *Logit* modellerde, kategorik bağımlı değişkene ilişkin *odds*'un doğal logaritması; bağımsız değişkenlerin doğrusal bir fonksiyonu olarak ifade edilmektedir.

Logit modeller bağımlı değişken kategorisinin yapısına göre çeşitli adlar almaktadır. Kategorik bağımlı değişken iki durumlu (Örn: Test sonucu pozitif/negatif) olduğunda İki durumlu (binary) *Logit* model kullanılır. Kategorik bağımlı değişken en az 3 kategoriden oluşan sınıflanabilir (nominal) ve sıralama içermeyen bir yapıda ise *Multinomial Logit* model ve en az 3 kategorili ve sıralı (ordinal) bir yapıda ise Sıralı *Logit* modelleri kullanılmaktadır.

3.1. İki Durumlu *Logit* Model (Binary Logit Model)

İki durumlu *Logit* modelde kategorik bağımlı değişken, bir olgunun ortaya çıkması ($y = 1$) ve çıkmaması ($y = 0$) gibi iki durumludur ve Binom dağılımının $n = 1$ için özel bir durumu olan Bernoulli dağılımı göstermektedir. Bernoulli dağılımında olgunun ortaya çıkma olasılığı $P(y = 1) = \pi$, ve çıkmama olasılığı $P(y = 0) = 1 - \pi$ ile ifade edilir. Bölüm 2'de Eşitlik 2.6 ile verilen *Logit* bağ fonksiyonunu kullanarak İki durumlu *Logit* model aşağıdaki gibi yazılmaktadır.

$$\log\left[\frac{\pi}{(1 - \pi)}\right] = \log\left[\frac{P(y = 1)}{P(y = 0)}\right] = \sum_{k=1}^K \beta_k x_k \quad (3.1)$$

Eşitlik 3.1’de $\pi/(1-\pi)$ ifadesi bağımlı değişken *odds*’u olarak tanımlanır. İki durumlu *Logit* modelde, kategorik bağımlı değişkene ilişkin *odds*’un doğal logaritması bağımsız değişkenler tarafından doğrusal olarak açıklanmaktadır.

Eşitlik 3.1 ile verilen İki durumlu *Logit* modelin türetilmesinde alternatif bir yaklaşım; iki durumlu kategorik bağımlı değişkenin etkisi altında olduğu, sürekli fakat gözlenemeyen bir bağımlı değişkenin var olduğunu varsaymak ve bu şekilde çözümlemeye gitmektir. Konu ile ilgili yayınlarda İki durumlu *Logit* model çoğu zaman, bu sürekli gözlenmeyen bağımlı değişken diğer adıyla Gizli Değişken (Latent Variable) yaklaşımı ile açıklanmaktadır (6, 7, 8). Gizli değişken analizi başlı başına ayrı bir konu olmakla birlikte Bölüm 3.1.1’de bu yaklaşım ile İki durumlu *Logit* modelin türetilmesi üzerinde durulacaktır.

3.1.1. Gizli Değişken Yaklaşımı ile İki Durumlu *Logit* Modelin Türetilmesi

Bazı kategorik iki durumlu bağımlı değişkenler için, bağımlı değişkenin aslında sürekli olan bir değişkenin etkisi altında olduğu tartışılabilir. Bu sürekli değişkenlere gizli değişken (ξ) denilmektedir ve bağımsız değişkenler tarafından doğrusal olarak aşağıdaki şekilde yazılmaktadır.

$$\xi = \sum_{k=1}^K \beta_k x_k + \varepsilon \quad (3.2)$$

Eşitlikte ε , sıfır ortalama ile simetrik dağılım gösteren ve dağılım fonksiyonu $F(\varepsilon)$ ile tanımlanan hata terimidir. ξ ise gözlenemeyen sürekli bir değişken olup, bağımsız değişkenler tarafından doğrusal olarak açıklanmaktadır. Gözlenen y değişkeni ise, bir olgunun ortaya çıkması ($y = 1$) ve çıkmaması ($y = 0$) şeklindeki iki durumlu göstermelik (dummy) bir değişkendir ve gizli değişken ξ cinsinden aşağıdaki gibi ifade edilir:

$$y = \begin{cases} 1 & \text{eğer } \xi > 0, \\ 0 & \text{eğer } \xi \leq 0 \end{cases} \quad (3.3)$$

Eşitlik 3.2’de β_k ’lar ve x_k ’ların toplamsal ifadesi ξ gizli değişkenin beklenen değeri $E(\xi/x_1, \dots, x_k)$ için doğrusal bir formdadır. Bu özellikten yararlanılarak gözlenen y değişkeni için olgunun ortaya çıkma olasılığı $P(y=1)$, gizli değişken ξ yardımıyla aşağıdaki gibi yazılır (6, 7, 8):

$$P(y=1) = P(\xi > 0) \quad (3.4)$$

Eşitlik 3.2 ve 3.3’ü kullanarak olgunun ortaya çıkma olasılığı $\pi = P(y=1)$ için aşağıdaki eşitliğe ulaşılır:

$$P(y=1) = P\left(\sum_{k=1}^K \beta_k x_k + \varepsilon > 0\right)$$

$$P(y=1) = P\left(\varepsilon > -\sum_{k=1}^K \beta_k x_k\right) = 1 - F\left(-\sum_{k=1}^K \beta_k x_k\right) \quad (3.5)$$

Eşitlik 3.5’de $F(\cdot)$ fonksiyonu hata teriminin dağılım fonksiyonunu tanımlamaktadır. Gizli değişken ξ için tanımlanan ε hata teriminin dağılımı Lojistik dağılım göstermektedir (6, 7). Dolayısı ile Eşitlik 3.5 aşağıdaki gibi düzenlenir:

$$P(y=1) = 1 - F\left(-\sum_{k=1}^K \beta_k x_k\right)$$

$$P(y=1) = 1 - \frac{1}{1 + \exp\left(\sum_{k=1}^K \beta_k x_k\right)} = \frac{\exp\left(\sum_{k=1}^K \beta_k x_k\right)}{1 + \exp\left(\sum_{k=1}^K \beta_k x_k\right)} \quad (3.6)$$

Olgunun ortaya çıkmama olasılığı $P(y=0) = 1 - P(y=1)$ olacağından Eşitlik 3.6’dan yararlanılarak $P(y=0)$ aşağıdaki şekilde yazılır:

$$P(y = 0) = 1 - P(y = 1) = \frac{1}{1 + \exp\left(\sum_{k=1}^K \beta_k x_k\right)} \quad (3.7)$$

Eşitlik 3.6 ve 3.7'yi kullanarak *Logit* dönüşümü uygulandığında iki durumlu *Logit* model için aşağıdaki eşitliğe ulaşılmaktadır:

$$\begin{aligned} \log\left[\frac{\pi}{(1-\pi)}\right] &= \log\left[\frac{P(y=1)}{P(y=0)}\right] = \log\left[\exp\left(\sum_{k=1}^K \beta_k x_k\right)\right] \\ \log\left[\frac{\pi}{(1-\pi)}\right] &= \log\left[\frac{P(y=1)}{P(y=0)}\right] = \sum_{k=1}^K \beta_k x_k \end{aligned} \quad (3.8)$$

Gizli Değişken yaklaşımı ile elde edilen İki durumlu *Logit* modelinin Eşitlik 3.8 ile ifade edilen formunun aslında Bölüm 3.1'de verilen Eşitlik 3.1 olduğu da görülmektedir.

İki durumlu *Logit* model iki ayrı formda yazılabilmektedir. Bunlardan birincisi Lojistik Regresyon modeli olarak da bilinen, Eşitlik 3.6 ile gösterilen, olgunun ortaya çıkma olasılığı $\pi = P(y = 1)$ cinsinden yazılmış şeklidir. İkincisi ise; Eşitlik 3.8 ile gösterilen ve *Logit* dönüşümünün bulunması nedeniyle *Logit* model olarak adlandırılan biçimdir. Literatürde kimi zaman iki durumlu *Logit* model ile Lojistik Regresyon modeli arasında bağımsız değişkenlerin yapısına göre ayırım yapılmaktadır. Bağımsız değişken kümesi kategorik olduğu durumda *Logit* modeller adını alırken, bağımsız değişken kümesinde kategorik ve sürekli değişkenler birlikte (karışık) olduğu durumda Lojistik Regresyon modeli adını almaktadır (6).

3.1.2. Katsayıların ve *Odds* Oranlarının Yorumlanması

İki durumlu *Logit* modellerde katsayı kestirimleri diğer *Logit* modellerinde olduğu gibi En Çok Olabilirlik Kestirim Yöntemi (EÇOK) ile elde edilmektedir. Bu çalışmada EÇOK yönteminden söz edilmeyecektir; ancak bu yöntemle hesaplanan katsayıların nasıl yorumlandığı hakkında bilgi verilecektir.

İki durumlu *Logit* modelde katsayı kestirimlerinin yorumlanabilmesi için öncelikle diğer GDM’de olduğu gibi verinin modele uyumunun incelenmesi gerekmektedir. Klasik doğrusal modellerde modelin anlamlılığı F test istatistiği ile incelenirken, *Logit* modellerde Olabilirlik Oranı (Likelihood Ratio) test istatistiği ile incelenebilmektedir. Bu test istatistiğinden Bölüm 3.4.1’de söz edilecektir.

İki durumlu *Logit* modelde β_k katsayılarının doğrudan yorumlanması bir anlam ifade etmemektedir. Hesaplanan katsayıların *Odds* ve *Odds* Oranı (Odds Ratio:OR) cinsinden yorumlanması daha kolay ve anlaşılır olmaktadır. Eşitlik 3.8’de her iki tarafın doğal logaritma üssü alınırsa; iki durumlu *Logit* model *odds* cinsinden aşağıdaki şekilde yazılır:

$$\exp\left(\log\left[\frac{\pi}{(1-\pi)}\right]\right) = \exp\left(\log\left[\frac{P(y=1)}{1-P(y=1)}\right]\right) = \exp\left(\sum_{k=1}^K \beta_k x_k\right)$$

$$\frac{\pi}{(1-\pi)} = \frac{P(y=1)}{1-P(y=1)} = \exp\left(\sum_{k=1}^K \beta_k x_k\right) \quad (3.9)$$

Eşitlik 3.9’un sol tarafındaki $\pi/(1-\pi)$ ifadesi bağımlı değişken *odds*’u olarak tanımlanır. Eşitliğin sağ kısmında ise k ’inci bağımsız değişkenin, bağımlı değişken *odds*’una marjinal etkisi $\exp(\beta_k)$ ile ifade edilmektedir (6). *Logit* model *odds*’un doğal logaritması için toplamsal ve doğrusal olarak yazılabilirken; *odds* cinsinden yazıldığında çarpımsal olarak aşağıdaki şekilde ifade edilmektedir:

$$\frac{\pi}{(1-\pi)} = \exp\left(\sum_{k=1}^K \beta_k x_k\right) = \prod_{k=1}^K \exp(\beta_k x_k) \quad (3.10)$$

Dolayısı ile; $\exp(\beta_k)$ ifadesi k 'inci bağımsız değişkendeki bir birimlik artışın (diğer değişkenler sabit iken) *odds*'a çarpımsal etkisi olarak da yorumlanabilmektedir (7). Buradan hareketle bir *odds*'tan bir diğerine çarpımsal olarak geçişin *odds* oranı olacağı açıktır. Birden büyük bir *odds* oranı ($OR > 1$), bağımsız değişkendeki bir birimlik artışın, ilgilenilen olgunun ortaya çıkma şansını arttırdığını, birden küçük bir *odds* oranı ($OR < 1$) ise bu şansı azalttığını ifade etmektedir (6).

Logit modellerde bağımsız değişken kümesinin kategorik olduğundan ve bu yönüyle Lojistik Regresyon modelinden ayrıldığından söz edilmişti. Bağımsız değişkendeki bir birimlik artış sürekli ve sıralı yapıdaki bağımsız değişkenler için gerçek sayısal artışı belirtmektedir. Ancak nominal yapıdaki bağımsız değişken için bir birimlik artış artık sayısal değerinden çok o kategoriyi tanımlayıcı özellik içermekte ve kategori değişimini ifade etmektedir. Bu nedenle *Logit* modele dahil edilecek bağımsız değişkenlerin düzeylerinin kategori değişimini yansıtacak şekilde göstermelik değişken (dummy variable) yardımıyla kodlanması ve bu şekilde çözümlemeye gidilmesi gerekmektedir (1).

3.2. *Multinomial Logit Model (Multinomial Logit Model)*

Multinomial Logit model, iki durumlu *Logit* modelin J kategori için genişlemiş bir durumudur. Bu modelde J tane bağımlı değişken kategorisi birbirinden ayrık, nominal, sıralama içermeyen bir yapıdadır ve bağımlı değişken dağılımı *Multinomial* dağılım göstermektedir. *Multinomial Logit* modelde, bağımsız değişkenlerin sıralama içermeyen bağımlı değişken kategorisi üzerindeki etkisi bir bütün olarak aynı anda incelenmektedir. Genel olarak bağımlı değişkenin j 'inci kategoriye düşme olasılığı $\pi_j = P(y = j)$ ile aşağıdaki şekilde tanımlanmaktadır (6):

$$\pi_j = P(y = j) = \frac{\exp\left(\sum_{k=1}^K \beta_{jk} x_k\right)}{1 + \sum_{j=1}^{J-1} \exp\left(\sum_{k=1}^K \beta_{jk} x_k\right)} \quad j = 1, 2, \dots, J-1 \quad (3.11)$$

Eşitlik 3.11’de β_{jk} katsayısının iki alt indisi bulunmaktadır. Bunlardan k indisi bağımsız değişkeni belirtirken; j indisi bağımlı değişken kategorisini belirtmek için kullanılmaktadır.

Bağımlı değişken kategorilerinden bir tanesi referans kategori olarak alındığından daha önce söz edilmişti. Çoğunlukla son bağımlı değişken kategorisi J , referans kategorisi olarak seçilir ve bağımlı değişkenin referans kategoriye düşme olasılığı $\pi_J = P(y = J)$ aşağıdaki şekilde tanımlanır (6):

$$\pi_J = P(y = J) = \frac{1}{1 + \sum_{j=1}^{J-1} \exp\left(\sum_{k=1}^K \beta_{jk} x_k\right)} \quad j = 1, 2, \dots, J-1 \quad (3.12)$$

Ayrıca; referans kategorisine düşme olasılığı diğer olasılıklar yardımıyla aşağıdaki şekilde de hesaplanabilmektedir:

$$P(y = J) = 1 - [P(y = 1) + P(y = 2) + \dots + P(y = J-1)] \quad (3.13)$$

Bölüm 2’de kısaca *Multinomial Logit* bağ fonksiyonunu kullanan Genelleştirilmiş Doğrusal modellere *Multinomial Logit* veya Referans Kategori (Baseline Category) *Logit* modeller adı verildiğinden söz edilmişti. Eşitlik 3.11 ve 3.12 *Multinomial Logit* bağ fonksiyonunda yerine konulduğunda, *Multinomial Logit* model için aşağıdaki eşitliğe ulaşılır:

$$\log(\pi_j / \pi_J) = \log \left[\frac{P(y=j)}{P(y=J)} \right] = \log \left[\frac{\frac{\exp\left(\sum_{k=1}^K \beta_{jk} x_k\right)}{1 + \sum_{j=1}^{J-1} \exp\left(\sum_{k=1}^K \beta_{jk} x_k\right)}}{\frac{1}{1 + \sum_{j=1}^{J-1} \exp\left(\sum_{k=1}^K \beta_{jk} x_k\right)}} \right]$$

$$\log(\pi_j / \pi_J) = \log \left[\frac{P(y=j)}{P(y=J)} \right] = \log \left[\exp\left(\sum_{k=1}^K \beta_{jk} x_k\right) \right]$$

$$\log(\pi_j / \pi_J) = \log \left[\frac{P(y=j)}{P(y=J)} \right] = \sum_{k=1}^K \beta_{jk} x_k \quad j = 1, \dots, J-1 \quad (3.14)$$

Eşitlik 3.14’de görüldüğü gibi *Multinomial Logit* model iki durumlu *Logit* model ile benzerlik göstermektedir. *Multinomial Logit* model $J = 2$ için iki durumlu *Logit* modele dönüşmektedir. Modelde katsayı kestirimleri, iki durumlu *Logit* modelde olduğu gibi En Çok Olabilirlik Kestirim yöntemi ile yapılmaktadır. Ancak iki durumlu *Logit* model için yapılan çözümlemeden farklı olarak; bağımlı değişken için tek bir *Logit* ($\log[\pi / 1 - \pi]$) oluşturmak yerine, $J - 1$ tane j ’inci kategoriye referans kategori ile karşılaştıran *Logit* oluşturulur ve eşanlı olarak çözüme gidilir (6).

3.2.1. İlgisiz Alternatiflerin Bağımsızlığı Varsayımı

Multinomial Logit modelde katsayı kestirimlerinin yorumlanabilmesi için iki durumlu *Logit* modelde olduğu gibi verinin modele uyumunun incelenmesi gerekmektedir. *Multinomial Logit* modelinde de verinin modele uyumu Olabilirlik Oranı test istatistiği ile incelenebilmektedir. Ancak *Multinomial Logit* modelinde İlgisiz Alternatiflerin Bağımsızlığı (Independence from Irrelevant Alternatives: IIA) olarak adlandırılan önemli bir varsayım bulunmaktadır. Bu varsayım kısaca; bağımlı değişken kategorisindeki iki alternatif kategorinin seçim olasılıklarının oranının,

diğer alternatifler tarafından sistematik olarak etkilenmediği varsayımı olarak belirtilebilmektedir (6). Dolayısı ile *Multinomial Logit* çözümlemesi yapılmadan önce bu varsayımın sağlanıp sağlanmadığının kontrol edilmesi gerekmektedir. Bu çalışma içerisinde İlgisiz Alternatiflerin Bağımsızlığı varsayımının sağlandığı varsayılmaktadır.

3.2.2. Katsayıların ve Odds Oranlarının Yorumlanması

Multinomial Logit modelde, iki durumlu *Logit* modelde olduğu gibi β_{jk} katsayılarının doğrudan yorumlanması bir anlam ifade etmemektedir. Aynı şekilde, hesaplanan katsayıların *odds* ve *odds* oranı cinsinden yorumlanması daha kolay ve anlaşılır olmaktadır. Eşitlik 3.14’de her iki tarafın doğal logaritma üssü alındığında, *Multinomial Logit* model *odds* cinsinden aşağıdaki şekilde yazılır:

$$\exp\left(\log\left[\frac{\pi_j}{\pi_J}\right]\right) = \exp\left(\log\left[\frac{P(y=j)}{P(y=J)}\right]\right) = \exp\left(\sum_{k=1}^K \beta_{jk} x_k\right)$$

$$\frac{\pi_j}{\pi_J} = \frac{P(y=j)}{P(y=J)} = \exp\left(\sum_{k=1}^K \beta_{jk} x_k\right) \quad j = 1, \dots, J-1 \quad (3.15)$$

Eşitlik 3.15’de (π_j / π_J) ifadesi, j ’inci bağımlı değişken kategorisinin seçilen referans kategorisine göre *odds*’u olarak tanımlanırken; k ’inci bağımsız değişkenin, j ’inci bağımlı değişken kategorisinin referans kategorisine göre *odds*’una marjinal etkisi, $\exp(\beta_{jk})$ ile ifade edilmektedir (6).

Multinomial Logit modelinde *odds* yorumlamaları iki durumlu *Logit* modeline göre biraz daha karmaşıktır. Birden büyük bir *odds* oranı ($OR > 1$), bağımsız değişkendeki bir birimlik artışın, referans kategorisi yerine j ’inci bağımlı değişken kategorisinde olma şansını arttırdığına; birden küçük bir *odds* oranı ($OR < 1$) bu şans azalttığına işaret etmektedir (6).

3.3. Sıralı *Logit* Modelleri (Ordinal Logit Models)

Bağımlı değişken kategorisinin ikiden fazla olması durumunda geliştirilmiş bir diğer yöntem, Sıralı *Logit* modelleridir. Bir önceki bölümde bağımlı değişken düzeyinin ikiden fazla olduğu durum için *Multinomial Logit* modeli incelenmişti. Ancak bu modelde bağımlı değişken kategorisi belirli bir sıralama içermeyen nominal bir yapıdaydı. Bazı durumlarda ise çok kategorili bağımlı değişken sıralı bir yapıda olabilir. Bu tip yapılara en çok anket çalışmalarında likert tipi ölçeklerde ya da sağlık alanında hastalık şiddetini ölçmek için yapılan çalışmalarda rastlanmaktadır. Bağımlı değişken kategorileri arasında açık sıralanabilir bir yapı vardır. Ancak ardışık kategoriler arasındaki uzaklıklar birbirlerine eşit değildir.

Bağımlı değişken kategorileri aralık ölçek ile ölçülmediği ve ardışık kategoriler arasındaki uzaklıklar eşit olmadığından bu tip veriler klasik regresyon ile kolayca modellenememektedir. Diğer yandan; bu tip verilere Bölüm 3.1’de verilen *Multinomial Logit* modelini kullanmak ise sıralı yapıdaki bağımlı değişkenin bu yapısını gözardı etmekte ve bağımlı değişkendeki tüm bilgileri kullanmakta yetersiz kalmaktadır (1). Bu sebeplerden dolayı Sıralı *Logit* modelleri bu tip verilerin çözümlenmesinde geniş bir kullanım alanı bulmuştur (6).

3.3.1. Gizli Değişken Yaklaşımı ile Sıralı *Logit* Modellerinin Türetilmesi

Sıralı *Logit* modelleri, gizli değişken yaklaşımı ile türetilen iki durumlu *Logit* modelin doğal bir genişlemesidir. Gizli değişken modeli daha önce Bölüm 3.1’de tanımlandığı gibi aşağıda verilmektedir:

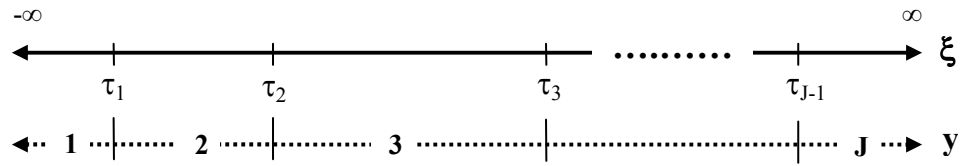
$$\xi = \sum_{k=1}^K \beta_k x_k + \varepsilon \quad (3.16)$$

İki durumlu *Logit* modelde olduğu gibi bu modelde de, gözlenen y değişkeninin etkisi altında olduğu, sürekli fakat gözlenmeyen ξ gizli değişkeni bulunmaktadır. ε ise, sıfır ortalama ile belirli simetrik bir dağılım gösteren ve

dağılım fonksiyonu $F(\varepsilon)$ ile tanımlanan hata terimidir. Bağımlı değişkenin J tane sıralı kategorisinin olduğu durumda gözlenen y değişkeni, gizli değişken ξ yardımıyla aşağıdaki gibi ifade edilir (8):

$$y = \begin{cases} 1 & \text{eğer } -\infty = \tau_0 \leq \xi \leq \tau_1 \\ 2 & \text{eğer } \tau_1 < \xi \leq \tau_2 \\ 3 & \text{eğer } \tau_2 < \xi \leq \tau_3 \\ \vdots & \\ J & \text{eğer } \tau_{J-1} < \xi = \infty \end{cases} \quad (3.17)$$

Eşitlik 3.18'deki τ 'lar, ardışık kategorileri ayıran, bilinmeyen eşik (threshold) değerleridir ve Şekil 3.1'deki gibi gösterilmektedir.



Şekil 3.1. J kategorili sıralı bağımlı değişken için gizli değişken ξ ve gözlenen y değişkeninin gösterimi.

Şekil 3.1'de koyu ile belirtilen doğru, gizli değişkeni ifade etmektedir. Eşik değerleri (τ 'lar) bu doğru üzerinde yer almaktadır. Gözlenen y bağımlı değişkeni ise kesikli çizgi ile belirtilen doğru üzerinde değer almaktadır. Eşitlik 3.17'den yararlanarak gözlenen bağımlı değişkenin birinci kategoriye düşme olasılığı $P(y = 1)$ aşağıdaki şekilde yazılır:

$$P(y = 1) = P(\tau_0 \leq \xi < \tau_1) \quad (3.18)$$

Eşitlik 3.16'yı Eşitlik 3.18'de yerine koyarak daha da açılırsa aşağıdaki eşitliğe ulaşılır:

$$\begin{aligned}
P(y=1) &= P(\tau_0 \leq \sum_{k=1}^K \beta_k x_k + \varepsilon < \tau_1) = P(\tau_0 - \sum_{k=1}^K \beta_k x_k \leq \varepsilon < \tau_1 - \sum_{k=1}^K \beta_k x_k) \\
P(y=1) &= P\left(\varepsilon < \tau_1 - \sum_{k=1}^K \beta_k x_k\right) - P\left(\varepsilon \leq \tau_0 - \sum_{k=1}^K \beta_k x_k\right) \\
P(y=1) &= F\left(\tau_1 - \sum_{k=1}^K \beta_k x_k\right) - F\left(\tau_0 - \sum_{k=1}^K \beta_k x_k\right) \\
P(y=1) &= F\left(\tau_1 - \sum_{k=1}^K \beta_k x_k\right) \tag{3.19}
\end{aligned}$$

Eşitlik 3.19'da $F(.)$ fonksiyonu hata teriminin dağılım fonksiyonunu tanımlamaktadır. Gizli değişken ξ için tanımlanan ε hata teriminin dağılımı Lojistik dağılım göstermektedir. Gözlenen bağımlı değişkenin birinci kategoriye düşme olasılığı $P(y=1)$ hesaplanırken eşitlikteki ikinci terim $F(\tau_0 - \sum_{k=1}^K \beta_k x_k) = F(-\infty - \sum_{k=1}^K \beta_k x_k) = 0$ olarak alınmıştır. Diğer kategori olasılıkları ise aşağıdaki gibi ifade edilir (8):

$$P(y=2) = F\left(\tau_2 - \sum_{k=1}^K \beta_k x_k\right) - F\left(\tau_1 - \sum_{k=1}^K \beta_k x_k\right) \tag{3.20}$$

$$P(y=3) = F\left(\tau_3 - \sum_{k=1}^K \beta_k x_k\right) - F\left(\tau_2 - \sum_{k=1}^K \beta_k x_k\right) \tag{3.21}$$

$\vdots$
 $\vdots$

Genel olarak gözlenen bağımlı değişkenin j 'inci kategoriye düşme olasılığı $P(y=j)$ aşağıdaki gibi yazılır:

$$P(y = j) = F\left(\tau_j - \sum_{k=1}^K \beta_k x_k\right) - F\left(\tau_{j-1} - \sum_{k=1}^K \beta_k x_k\right) \quad (3.22)$$

Gözlenen bağımlı değişkenin son kategoriye düşme olasılığı $P(y = J)$ hesaplanırken ilk terim $F(\tau_J - \sum_{k=1}^K \beta_k x_k) = F(\infty - \sum_{k=1}^K \beta_k x_k) = 1$ olarak alınır (8).

3.3.2. Yığılımlı *Logit* Modelleri (Cumulative *Logit* Models)

Sıralı *Logit* modellerinde, bağımlı değişken kategori karşılaştırmalarında kullanılan farklı *Logit* oluşturma biçimleri bulunmaktadır. Bunlardan yorumlanması en kolay olan ve en sık kullanılan Yığılımlı *Logit*'lerdir. Sıralı *Logit* modellerinde Yığılımlı *Logit* modelleri dışında kullanılan diğer *Logit* modelleri ise Ardışık Kategori *Logit* (Adjacent Category *Logit*) ve Sürekli Oran *Logit* (Continuation Ratio *Logit*) modelleridir (1, 4). Ancak çalışma içerisinde yukarıda söz edilen modellere değinilmeyecektir.

Yığılımlı *Logit* Modelleri ise Orantısal Risk varsayımının (Proportional Odds Assumption) sağlandığı ve sağlanmadığı durumlar için iki ana başlık altında incelenecektir. Bunlar Orantısal Risk modeli (Proportional Odds Model: McCullagh 1980) ve Orantısal Olmayan Risk modelidir (Non- Proportional Odds Model: Fu 1998).

Bağımlı değişken kategorisinin sıralı olduğu durumda, kategori olasılıkları çoğu zaman yığılımlı olasılıklar cinsinden ifade edilmektedir. Yığılımlı olasılık; belirlenen kategori ve daha düşük kategorilere düşme olasılığını birlikte ele almakta ve dolayısı ile sıralı bağımlı değişkenin yapısındaki tüm bilgileri kullanmaktadır (5). Genel olarak j kategorisi için yığılımlı olasılık aşağıdaki şekilde tanımlanır:

$$P(y \leq j) = P(y = 1) + P(y = 2) + \dots + P(y = j) \quad (3.23)$$

Eşitlik 3.22, Eşitlik 3.23’de yerine konulduğunda ve sadeleştirmeler yapıldığında, yığılımlı olasılık için aşağıdaki eşitliğe ulaşılır (8).

$$P(y \leq j) = F\left(\tau_j - \sum_{k=1}^K \beta_k x_k\right)$$

$$P(y \leq j) = \frac{\exp\left(\tau_j - \sum_{k=1}^K \beta_k x_k\right)}{1 + \exp\left(\tau_j - \sum_{k=1}^K \beta_k x_k\right)} \quad j = 1, 2, \dots, J-1 \quad (3.24)$$

Yığılımlı *Logit* modeli, *Logit* dönüşümünde kategori olasılığı ($P(y = j)$) yerine yığılımlı olasılığın ($P(y \leq j)$) kullanılmasıyla oluşturulmaktadır *Logit* bağ fonksiyonunun tanımından Yığılımlı *Logit* model aşağıdaki gibi yazılmaktadır (4):

$$\text{logit}[P(y \leq j)] = \log\left[\frac{P(y \leq j)}{1 - P(y \leq j)}\right] \quad j = 1, 2, \dots, J-1 \quad (3.25)$$

Eşitlik 3.25’deki ($P(y \leq j)/1 - P(y \leq j)$) ifadesi, j ’inci bağımlı değişken kategorisi için Yığılımlı *Odds* (Cumulative Odds) olarak tanımlanmaktadır. Dolayısı ile Yığılımlı *Logit* modellerinde *Odds*; İki durumlu ve *Multinomial Logit* modellerinden farklı olarak sıralı yapıyı kullanır ve j ’inci bağımlı değişken kategorisi için daha düşük kategorilere düşme olasılığının, daha yüksek kategoriye düşme olasılığı ile oranlanmasıyla elde edilir (4).

3.3.2.1. Orantısal Risk Modeli (Proportional Odds Model) ve Orantısal Risk Varsayımı (Proportional Odds Assumption)

Yığılımlı *Logit* modelde, yığılımlı olasılık değerleri Lojistik dağılımında yerlerine konulduğunda, Orantısal Risk Modeli için aşağıdaki eşitliklere ulaşılır (6):

$$\text{logit}[P(y \leq j)] = \log \left[\frac{P(y \leq j)}{1 - P(y \leq j)} \right] = \log \left[\frac{F(\tau_j - \sum_{k=1}^K \beta_k x_k)}{1 - F(\tau_j - \sum_{k=1}^K \beta_k x_k)} \right]$$

$$\text{logit}[P(y \leq j)] = \log \left[\exp \left(\tau_j - \sum_{k=1}^K \beta_k x_k \right) \right]$$

$$\text{logit}[P(y \leq j)] = \tau_j - \sum_{k=1}^K \beta_k x_k \quad j = 1, 2, \dots, J-1 \quad (3.26)$$

Eşitlik 3.26 ile verilen Orantısal Risk Modelinde; her bir yığılımlı *Logit*'in τ_j ile gösterilen kendine ait bir eşik (threshold) değeri bulunmaktadır. Bağımlı değişken kategorileri $j = 1, 2, \dots, J-1$ ile gösterildiğinde eşitlikteki β_k katsayılarının bağımlı değişken kategorilerinden bağımsız olduğu görülmektedir. Diğer bir ifade ile; bu modelde bağımsız değişkenlerin bağımlı değişken *odds*'una (yığılımlı *odds*), tüm bağımlı değişken kategorilerinde eşit bir etkisi vardır (5).

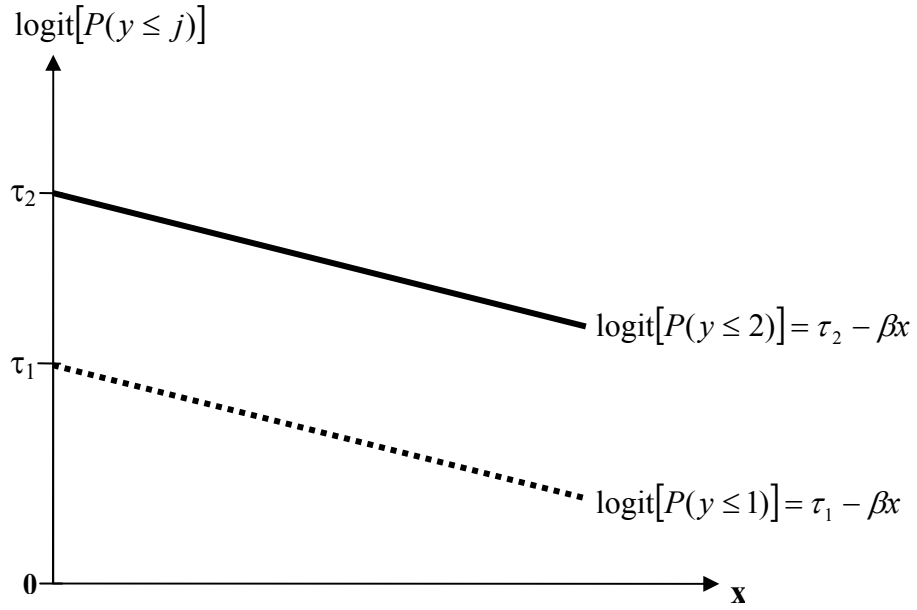
Üç kategorili sıralı yapıdaki bağımlı değişkeni açıklayan tek bir bağımsız değişken olduğu durum için oluşturulan iki adet yığılımlı *Logit* aşağıdaki gibidir:

$$\text{logit}[P(y \leq 1)] = \log \left[\frac{P(y = 1)}{P(y = 2) + P(y = 3)} \right] = \tau_1 - \beta x \quad (3.27)$$

$$\text{logit}[P(y \leq 2)] = \log \left[\frac{P(y = 1) + P(y = 2)}{P(y = 3)} \right] = \tau_2 - \beta x \quad (3.28)$$

Oluşturulan her bir yığılımlı *Logit* için ayrı eşik değerleri bulunmaktadır. x bağımsız değişkeninin bağımlı değişkene etkisi (β) ise, her bir *Logit*'de eşittir. Birden fazla bağımsız değişken olması halinde durum aynıdır. Tüm yığılımlı *Logit*'lerde k 'ıncı bağımsız değişkene ait β_k katsayıları birbirine eşittir. Bu da sağlanması zor olan ve öncelikle doğruluğunun test edilmesi gerekli bir varsayımdır.

Orantısal Risk Modellerinde oluşturulan Yığılımlı *Logit*'lerin paralellliği varsayımı, Paralel Eğimler Varsayımı (Parallel Slopes Assumption) ya da Orantısal Risk Varsayımı (Proportional Odds Assumption) olarak da bilinmektedir (5). Bu varsayımın testinden Bölüm 3.3.2.3'de söz edilecektir. Eşitlik 3.27 ve 3.28 için oluşturulan Yığılımlı *Logit*'ler ve bu *Logit*'lerin paralellliği Şekil 3.2'de gösterilmiştir.



Şekil 3.2. Üç düzeyli sıralı bağımlı değişken için oluşturulan Yığılımlı *Logit*'lerin paralellığının grafiksel olarak gösterimi.

Orantısal Risk Modelinde (Eşitlik 3.26) β_k katsayılarının önünde (-) işareti bulunmaktadır. Bu negatif işaretin anlamı; pozitif bir β_k katsayısı ile daha düşük kategoriye düşme olasılığının ($P(y \leq j)$) ters orantılı olduğudur. Diğer bir ifade ile; pozitif bir β_k katsayısı, daha düşük kategoriye düşme olasılığının azaldığını, daha yüksek bir kategoriye düşme olasılığının ($P(y \geq j)$) ise arttığını ifade etmektedir (1, 4-6). Tam tersi bir durum için; negatif bir β_k katsayısı, daha düşük kategoriye düşme olasılığının ($P(y \leq j)$) arttığını, daha yüksek kategoriye düşme olasılığının ($P(y \geq j)$) ise azaldığını ifade etmektedir.

$\text{logit}[P(y \leq j)] = \tau_j - \sum_{k=1}^K \beta_k x_k$ eşitliğinde eşik değerleri, diğer doğrusal modellerde olduğu gibi; x_k bağımsız değişkenlerin yokluğunda var olan etkiyi (riski) tanımlamaktadır. Dolayısı ile; Eşitlik 3.27 ve 3.28'de $x = 0$ yerine konulduğunda $\beta x = 0$ olacağından aşağıdaki sonuçlara ulaşılır:

$$\text{logit}[P(y \leq 1)] = \log \left[\frac{P(y = 1)}{P(y = 2) + P(y = 3)} \right] = \tau_1 - \beta(0) = \tau_1$$

$$\text{logit}[P(y \leq 2)] = \log \left[\frac{P(y = 1) + P(y = 2)}{P(y = 3)} \right] = \tau_2 - \beta(0) = \tau_2$$

Orantısal Risk Modelinde β_k katsayısı, diğer doğrusal modellerde olduğu gibi k 'ıncı bağımsız değişkendeki bir birimlik artışın etkisini vermektedir. Bu durum Eşitlik 3.27 ve 3.28 için aşağıdaki şekilde ifade edilir:

$$\text{logit}[P(y \leq 1)/x = 1] - \text{logit}[P(y \leq 1)/x = 0] = (\tau_1 - \beta) - (\tau_1) = -\beta$$

$$\text{logit}[P(y \leq 2)/x = 1] - \text{logit}[P(y \leq 2)/x = 0] = (\tau_2 - \beta) - (\tau_2) = -\beta$$

Orantısal Risk modelinde k 'ıncı bağımsız değişkendeki bir birimlik artışın, j 'inci bağımlı değişken kategorisi yığılımlı *odds*'una etkisi, diğer bir ifade ile *Odds* Oranı aşağıdaki gibi hesaplanmaktadır (8).

$$OR = \frac{\frac{P(y \leq j/x_k + 1)}{1 - P(y \leq j/x_k + 1)}}{\frac{P(y \leq j/x_k)}{1 - P(y \leq j/x_k)}} = \exp(-\beta_k) \quad (3.29)$$

Orantısal Risk modelinde tek değişkenli durum için hesaplamalar basit iken, çok değişkenli çözümlemelerde, En Çok Olabilirlik Kestirim yöntemi ile elde

edilecek kestirimler için iterasyonlara gerek duyulmaktadır. Çok değişkenli durumlarda istatistiksel paket programları bu çözümlemeleri yapabilmektedir. Ancak burada dikkat edilmesi gereken nokta, istatistiksel paket programlarında Orantısız Risk modeli için farklı gösterimler kullanılmaktadır. Örneğin STATA paket programı Eşitlik 3.26 ile ifade edilen formu kullanırken, SAS paket programı bu eşitliğin (+) işaretlisini kullanmaktadır. Paket programlarında yapılacak çözümlemelerde dikkat edilmesi gereken diğer bir nokta ise; bağımlı değişken kategorilerini artan mı yoksa azalan bir yapı ile oluşturacağımızdır. Çünkü yığılımlı *Logit* oluşturma esnasında, küçük kategorilerin mi yoksa büyük kategorilerin mi referans alınacağı, hesaplanacak β_k katsayısının işaretini etkilemesi açısından önemlidir (6).

3.3.2.2. Orantısız Olmayan Risk Modeli (Non-Proportional Odds Model)

Orantısız Risk varsayımının kimi durumlarda sağlanması zor olabilir. Bu varsayımın sağlanmadığı durumlarda yapılabilecek bir diğer çözümleme ise, bağımlı değişken sıralı yapıda olmasına rağmen Bölüm 3.2’de verilen *Multinomial Logit* çözümlemesidir (9). Ancak *Multinomial Logit* çözümlemede bağımlı değişkenin sıralı yapısının göz ardı edildiği ve nominal olarak modele katıldığı unutulmamalıdır (2). Dolayısı ile hem sıralı yapıyı dikkate alan hem de oldukça katı olan orantısız risk varsayımını rahatlatan bir model arayışına girilmiştir.

Fu (1998) tarafından önerilen Orantısız Olmayan Risk modelinde; *Multinomial Logit* modelinin aksine *Logit* oluşturma esnasında yığılımlı *Logit*’ler kullanılmaktadır ancak orantısız risk varsayımı sağlanmamaktadır. Diğer bir ifade ile bu modelde; bağımsız değişkenlerin bağımlı değişken *odds*’una (yığılımlı *odds*) etkisi artık eşit değildir ve bağımlı değişken kategorisi $j = 1, 2, \dots, J - 1$ ile gösterildiğinde β_{jk} katsayıları her bir bağımlı değişken kategorisi için farklıdır (3,10).

Orantısal Olmayan Risk modelinde (diğer adıyla Genelleştirilmiş Sıralı *Logit* model), j 'inci kategori için yığılımlı olasılık değeri aşağıdaki şekilde tanımlanır (3,10):

$$P(y \leq j) = F\left(\tau_j - \sum_{k=1}^K \beta_{jk} x_k\right) \quad j = 1, 2, \dots, J-1 \quad (3.30)$$

Eşitlik 3.30'da $F(.)$ fonksiyonu hata teriminin dağılım fonksiyonunu tanımlamaktadır. Orantısal Risk modelinde olduğu gibi bu modelde de hatalar Lojistik dağılım göstermektedir. Dolayısı ile Eşitlik 3.30 aşağıdaki şekilde yazılır (3):

$$P(y \leq j) = \frac{\exp\left(\tau_j - \sum_{k=1}^K \beta_{jk} x_k\right)}{1 + \exp\left(\tau_j - \sum_{k=1}^K \beta_{jk} x_k\right)} \quad j = 1, 2, \dots, J-1 \quad (3.31)$$

Yığılımlı olasılık için *Logit* dönüşümü yapıldığında ve Orantısal Risk modelindeki adımlar izlendiğinde, Orantısal Olmayan Risk modeli için aşağıdaki eşitliklere ulaşılır:

$$\text{logit}[P(y \leq j)] = \log\left[\frac{P(y \leq j)}{1 - P(y \leq j)}\right] = \log\left[\frac{F\left(\tau_j - \sum_{k=1}^K \beta_{jk} x_k\right)}{1 - F\left(\tau_j - \sum_{k=1}^K \beta_{jk} x_k\right)}\right]$$

$\vdots$
 $\vdots$

$$\text{logit}[P(y \leq j)] = \log\left[\exp\left(\tau_j - \sum_{k=1}^K \beta_{jk} x_k\right)\right]$$

$$\text{logit}[P(y \leq j)] = \tau_j - \sum_{k=1}^K \beta_{jk} x_k \quad j = 1, 2, \dots, J-1 \quad (3.32)$$

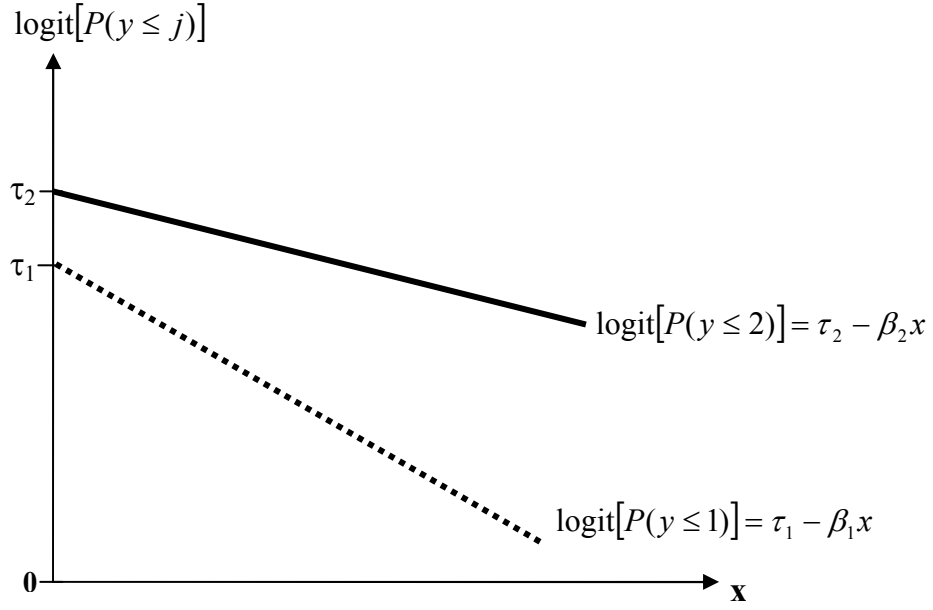
Eşitlik 3.32 ile verilen Orantısız Olmayan Risk Modelinde; her bir yığılımlı *Logit*'in τ_j ile gösterilen kendine ait bir eşik değeri bulunmaktadır. Bağımlı değişken kategorileri $j = 1, 2, \dots, J - 1$ ile gösterildiğinde eşitlikteki β_{jk} katsayıları her bir bağımlı değişken kategorisinde farklı değerler almaktadır (3,10).

Üç kategorili sıralı yapıdaki bağımlı değişkeni açıklayan tek bir bağımsız değişken olduğu durumda, Orantısız Olmayan Risk modeli için oluşturulan iki adet yığılımlı *Logit* aşağıdaki gibidir:

$$\text{logit}[P(y \leq 1)] = \log \left[\frac{P(y = 1)}{P(y = 2) + P(y = 3)} \right] = \tau_1 - \beta_1 x \quad (3.33)$$

$$\text{logit}[P(y \leq 2)] = \log \left[\frac{P(y = 1) + P(y = 2)}{P(y = 3)} \right] = \tau_2 - \beta_2 x \quad (3.34)$$

Oluşturulan her bir yığılımlı *Logit* için ayrı eşik değerleri bulunmaktadır. x bağımsız değişkeninin bağımlı değişkene etkisi ise, her bir *Logit*'de birbirinden farklı olmaktadır. Birinci bağımlı değişken kategorisini daha büyük kategoriler ile karşılaştıran *Logit*'de bağımsız değişkenin etkisi β_1 ile ifade edilirken; birinci ve ikinci bağımlı değişken kategorisini daha büyük kategori ile karşılaştıran *Logit*'de bağımsız değişkenin etkisi β_2 ile ifade edilir. Dolayısı ile oluşturulan yığılımlı *Logit*'lerin eğimleri birbirinden farklı olmaktadır. Eşitlik 3.33 ve 3.34 için oluşturulan Yığılımlı *Logit*'lerin paralelliğinin bozulumu Şekil 3.3'de gösterilmiştir.



Şekil 3.3. Üç düzeyli sıralı bağımlı değişken için oluşturulan Yığılımlı *Logit*'lerin paralelliğinin bozululumunun grafiksel olarak gösterimi.

$$\text{logit}[P(y \leq j)] = \tau_j - \sum_{k=1}^K \beta_{jk} x_k \quad \text{eşitliğinde eşik değerleri, Orantısız Risk}$$

modelinde de olduğu gibi; x_k bağımsız değişkenlerin yokluğunda var olan etkiyi tanımlamaktadır. Dolayısı ile; Eşitlik 3.33 ve 3.34'de $x = 0$ yerine konulduğunda $\beta_j x = 0$ olacağından aşağıdaki sonuçlara ulaşılır:

$$\text{logit}[P(y \leq 1)] = \log \left[\frac{P(y = 1)}{P(y = 2) + P(y = 3)} \right] = \tau_1 - \beta_1(0) = \tau_1$$

$$\text{logit}[P(y \leq 2)] = \log \left[\frac{P(y = 1) + P(y = 2)}{P(y = 3)} \right] = \tau_2 - \beta_2(0) = \tau_2$$

Orantısız Olmayan Risk Modelinde β_{jk} katsayısı, k 'inci bağımsız değişkendeki bir birimlik artışın yığılımlı *Logit*'e etkisini vermektedir. Bu durum Eşitlik 3.33 ve 3.34 için aşağıdaki şekilde ifade edilir:

$$\text{logit}[P(y \leq 1)/x = 1] - \text{logit}[P(y \leq 1)/x = 0] = (\tau_1 - \beta_1) - (\tau_1) = -\beta_1$$

$$\text{logit}[P(y \leq 2)/x = 1] - \text{logit}[P(y \leq 2)/x = 0] = (\tau_2 - \beta_2) - (\tau_2) = -\beta_2$$

Orantısız Olmayan Risk modelinde k 'nci bağımsız değişkendeki bir birimlik artışın, j 'nci bağımlı değişken kategorisi yığılımlı *odds*'una etkisi, diğer bir ifade ile *Odds Oranı* aşağıdaki gibi hesaplanmaktadır.

$$OR = \frac{\frac{P(y \leq j/x_k + 1)}{1 - P(y \leq j/x_k + 1)}}{\frac{P(y \leq j/x_k)}{1 - P(y \leq j/x_k)}} = \exp(-\beta_{jk}) \quad (3.35)$$

3.3.2.3. Orantısız Risk Varsayımının Geçerliliğinin Testi (Test of Proportional Odds Assumption)

Orantısız Risk varsayımının sağlandığı durumda oluşturulan tüm Yığılımlı *Logit*'lerde k 'nci bağımsız değişkene ait β_k katsayıları birbirine eşit olacağından, bu varsayımın testinde k 'nci bağımsız değişkene ait katsayıların birbirine eşit olup olmadığı aşağıdaki hipotez ile test edilir (8):

$$H_0 : \beta_{1k} = \beta_{2k} = \dots = \beta_{(J-1)k} \quad j = 1, 2, \dots, J-1 \quad (3.36)$$

Orantısız Risk modeli, Orantısız Olmayan Risk modelinin kısıtlanmış bir durumudur. Orantısız Risk modelinden elde edilen Log- Olabilirlik L_1 , Orantısız Olmayan Risk modelinden elde edilen Log- Olabilirlik L_2 ile gösterildiğinde bu iki Log- Olabilirlik arasındaki fark bize bu kısıtın kaldırılması durumunda Log- Olabilirlikteki değişimi verir (8). Dolayısı ile Orantısız Risk varsayımının yaklaşık olarak geçerliliğinin testi, Olabilirlik Oranı test istatistiği ile aşağıdaki gibi yapılmaktadır (11, 12).

$$G^2 = -2(L_1 - L_2) \quad G^2 \sim \chi^2_{[K(J-2)]} \quad (3.37)$$

G^2 Olabilirlik Oranı test istatistiği $\chi^2_{[K(J-2)]}$ değeri ile karşılaştırılarak test edilir. Eğer H_0 red edilirse, en az bir katsayı diğerlerinden farklı olduğundan Orantısal Risk varsayımı sağlanmamaktadır. Ancak bu testin yaklaşık bir test olduğu unutulmamalıdır. Ayrıca bu test ile modeldeki hangi bağımsız değişken veya değişkenlerin orantısallığı bozduğu hakkında bilgi edinilememektedir (8).

Orantısal Risk varsayımının geçerliliğinin testi için kullanılan bir diğer test ise Brant (1990) tarafından önerilen Wald test istatistiğidir (8, 13, 14). Bu test hem genel olarak orantısallığı test etmekte, hem de orantısallığın hangi değişken veya değişkenler yüzünden bozulduğu hakkında bilgi vermektedir. Ancak test işlemleri yoğun matris işlemleri gerektirdiğinden hesaplamalar oldukça güçtür. İstatistiksel paket programları (STATA) Brant testi adı altında bu testi yapmaktadır (3, 13, 14).

3.4. Model Geçerlilik Testi ve Uyum İyiliği Göstergeleri (Goodness of Fit Statistics)

3.4.1. Olabilirlik Oranı Test İstatistiği (Likelihood Ratio Test Statistics)

Logit modellerde katsayı kestirimlerinin yorumlanabilmesi için öncelikle diğer Genelleştirilmiş Doğrusal Modellerde olduğu gibi modelin geçerliliğinin incelenmesi gerekmektedir. Diğer bir ifade ile modele katılan bağımsız değişkenlerin bağımlı değişkeni, sadece sabit terimin (β_0) bulunduğu modelden daha anlamlı bir biçimde açıklaması gerekmektedir. Klasik doğrusal modellerde modelin geçerliliği F test istatistiği ile incelenirken, *Logit* modellerde Olabilirlik Oranı (Likelihood Ratio) test istatistiği ile incelenmektedir. Sadece sabit terimin bulunduğu modelin Log-Olabilirliği L_0 ve serbestlik derecesi sd_0 , bir ya da birden fazla bağımsız değişkenin bulunduğu modelin Log- Olabilirliği L_1 ve serbestlik derecesi sd_1 ile gösterildiğinde Olabilirlik Oranı test istatistiği aşağıdaki şekilde elde edilir (7):

$$G^2 = -2(L_0 - L_1) \quad G^2 \sim \chi^2_{[sd_1 - sd_0]} \quad (3.38)$$

G^2 Olabilirlik Oranı test istatistiği $sd_1 - sd_0$ serbestlik derecesindeki Ki-Kare dağılımı göstermektedir. Eğer $G^2 \geq \chi^2_{[sd_1 - sd_0]}$ ise modele katılan bağımsız değişkenlerin bağımlı değişkeni sadece sabit terimin olduğu modelden daha anlamlı bir biçimde açıkladığı, diğer bir ifade ile oluşturulan modelin anlamlı olduğu söylenir (7).

Olabilirlik Oranı test istatistiği modele katılan bir çok bağımsız değişkenin anlamlılığının testi için genelleştirilebilir. k tane bağımsız değişkenin bulunduğu modelin Log- Olabilirliği L_k ve serbestlik derecesi sd_k , $k + p$ tane bağımsız değişkenin bulunduğu modelin Log- Olabilirliği L_{k+p} ve serbestlik derecesi sd_{k+p} ile gösterildiğinde Genelleştirilmiş Olabilirlik Oranı (Generalized Likelihood Ratio) test istatistiği aşağıdaki gibidir (7):

$$G_G^2 = -2(L_k - L_{k+p}) \quad G^2 \sim \chi_p^2 \quad (3.39)$$

G_G^2 Olabilirlik Oranı test istatistiği p serbestlik derecesindeki Ki- Kare dağılımı göstermektedir. Eğer $G_G^2 \geq \chi_p^2$ ise modele katılan p tane bağımsız değişkenin bağımlı değişkeni sadece k tane bağımsız değişken içeren modelden daha anlamlı bir biçimde açıkladığı söylenir. Ancak Genelleştirilmiş Olabilirlik Oranı test istatistiğinin yalnızca içiçe model karşılaştırmalarında kullanıldığına dikkat edilmesi gerekmektedir (7).

3.4.2. Sapma Ölçüsü (Deviance)

Genelleştirilmiş Doğrusal Modellerde Sapma ölçüsü (Deviance), gerçek değerlerin kestirilen değerlerden ne kadar ayrıldığıнын bir göstergesidir. Bu yönü ile Sapma ölçüsü klasik doğrusal modellerdeki Artık Kareler Toplamına eşdeğer bir ölçüdür (1).

Gözlem sayısı kadar parametre içeren ve kompleks olan modele tam doygun model (saturated model) denilmektedir (1). Doygun modelden elde edilen Log- Olabilirlik L_S (veri setinin Log- Olabilirliği) ve önemli olduğu düşünülen değişkenler tarafından oluşturulan doygun modelin herhangi bir alt modelinden elde edilen Log- Olabilirlik L_M (modelin Log- Olabilirliği) ile gösterildiğinde, Sapma ölçüsü aşağıdaki şekilde ifade edilmektedir:

$$D = -2(L_M - L_S) \quad (3.40)$$

Sapma ölçüsü gerçek değerlerden sapmanın büyüklüğünü vereceğinden dolayı, daha küçük bir sapma modelin veriye daha iyi uyum sağladığının bir göstergesidir.

Sapma ölçüsü Eşitlik 3.40'dan da anlaşılacağı gibi, tam doygun model ile kullanılan modeli karşılaştıran bir Olabilirlik Oranı test istatistiğidir. $D(M_k)$ k tane bağımsız değişken içeren modelin sapma ölçüsü ve $D(M_{k+p})$ $k + p$ tane bağımsız değişken içeren modelin sapma ölçüsünü göstermek üzere; Olabilirlik Oranı test istatistiği (Eşitlik 3.39) Sapma ölçüsü cinsinden aşağıdaki şekilde de yazılabilmektedir (1).

$$G_G^2 = D(M_k) - D(M_{k+p}) \quad G^2 \sim \chi_p^2 \quad (3.41)$$

3.4.3. Akaike Bilgi Kriteri (Akaike Information Criteria)

Akaike Bilgi Kriteri (AIC) içiçe olmayan (non- nested) veya farklı örneklemelerden hesaplanan model karşılaştırmalarında kullanılmaktadır. Akaike (1973) tarafından önerilen bu kriter ile yarışan iki model için, hangi modelin veriye daha iyi uyum sağladığı hakkında bilgi edinilebilmektedir (3). Modelin Log-Olabilirliği L_M , modeldeki parametre sayısı k ve gözlem sayısı n ile gösterildiğinde; Akaike Bilgi Kriteri aşağıdaki eşitlik ile hesaplanır:

$$AIC = \frac{-2L_M + 2k}{n} \quad (3.42)$$

AIC'nin daha küçük değerleri, o modelin veriye daha iyi uyum sağladığının bir göstergesidir. Dolayısı ile birkaç model içerisinde veriye en iyi uyum sağlayan model AIC'si küçük olan modeldir (3).

3.4.4. Bayesci Bilgi Kriteri (Bayesian Information Criteria)

Bayesci Bilgi Kriteri (BIC) modellerin Bayesci karşılaştırmasına dayanmaktadır. Bayesci Bilgi Kriteri, AIC'de olduğu gibi içiçe olmayan veya farklı örneklemelerden hesaplanan model karşılaştırmalarında kullanılmaktadır. Raftery (1996) tarafından önerilen bu kriterde tam doygun modelin değeri "0" olmaktadır (3). Modelin Sapma ölçüsü $D(M)$, Sapma ölçüsünün serbestlik derecesi $sd_{D(M)}$ ve gözlem sayısı n ile gösterildiğinde; Bayesci Bilgi Kriteri aşağıdaki eşitlik ile hesaplanır:

$$BIC = D(M) - sd_{D(M)} \times \ln(n) \quad (3.43)$$

BIC'nin daha küçük değerleri, o modelin veriye daha iyi uyum sağladığının bir göstergesidir. Dolayısı ile birkaç model içerisinde veriye en iyi uyum sağlayan model BIC'si küçük olan modeldir. Eğer BIC pozitif ise, tam doygun model tercih edilmelidir (3).

3.4.5. McFadden'ın Düzeltilmiş Olabilirlik Oranı Göstergesi (McFadden's Adjusted Likelihood Ratio Index)

Sadece sabit terimin bulunduğu modelin Log- Olabilirliği L_0 ve k tane bağımsız değişkenin bulunduğu modelin Log-Olabilirliği L_1 ile gösterildiğinde McFadden'ın düzeltilmiş Olabilirlik Oranı göstergesi aşağıdaki şekilde tanımlanır (3):

$$\bar{R}_{McFadden}^2 = 1 - \frac{L_1 - k}{L_0} \quad (3.44)$$

McFadden'ın düzeltilmiş Olabilirlik Oranı göstergesi çoğu istatistiksel paket programlarında Yapay (Pseudo) R^2 olarak geçmektedir. En Çok Olabilirlik Kestirim yöntemi ile hesaplanan herhangi bir model için bu gösterge hesaplanabilmektedir. Modeller arasında en büyük $\bar{R}^2$ 'ye sahip olan model veriye en iyi uyum sağlayan modeldir (3).

DÖRDÜNCÜ BÖLÜM

4. GEREÇ ve YÖNTEM

Bu çalışmanın amacı, bağımlı değişken kategorisinin ikiden fazla olması durumunda geliştirilmiş olan *Multinomial* ve Sıralı *Logit* modellerini ayrıntılı olarak ele almak, Orantısız Risk varsayımının sağlanmadığı durumlarda Orantısız Olmayan Risk modeli ile *Multinomial Logit* modelini karşılaştırmak ve bu modeller arasından verilerin yapısına ve varsayımlarına göre en uygun modeli belirlemektir.

Bu amaç doğrultusunda hipotez; bağımlı değişkenin sıralı yapıda olduğu ve Orantısız Risk Varsayımının sağlanmadığı durumlarda Orantısız Olmayan Risk modelinin, *Multinomial Logit* modelinden daha iyi bir seçenek olduğudur. Bu hipotezi test etmek için Hacettepe Üniversitesi İç Hastalıkları Anabilim Dalı Geriatri Ünitesine Şubat 2002 ile Haziran 2005 tarihleri arasında başvuran hastaların verileri kullanılmıştır.

Bilindiği gibi Koroner Kalp Hastalığı (Coronary Heart Disease: CHD) en önemli ölümcül hastalıklardan biridir (15, 16). Yapılan çalışmalarda hipertansiyon, diyabet, hiperlipidemi, yaş, aile öyküsü ve sigaranın koroner kalp hastalığı için risk faktörü olduğu saptanmıştır (16, 17). Son zamanlarda yapılan çalışmalarda ise ferritin, homosistein, ürik asit ve C-reaktif protein gibi yeni risk faktörlerinin (novel cardiac risk factors) koroner kalp hastalığı gelişimi üzerindeki etkisi araştırılmış ve en az bilinen diğer risk faktörleri kadar etkili olduğu görülmüştür (18-20).

Çalışmada sıralı yapıdaki bağımlı değişken Framingham Risk Skorudur (Framingham Risk Score: FRS). Framingham risk skoru, Wilson ve ark. tarafından 10 yıl içinde koroner kalp hastalığı gelişimini kestirmek için geliştirilen bir skordur (16). Bu skor hesaplanırken hastanın yaşı, toplam kolesterol, hdl, kan basıncı, sigara içme durumu ve diyabet hastası olup olmadığı dikkate alınmaktadır. Framingham risk skoru üç risk grubuna aşağıdaki şekilde ayrılmaktadır (16):

$$FRS = \begin{cases} 0 : \text{Düşük Risk} & \text{Eğer } 0 \leq FRS \leq 9 \\ 1 : \text{Orta Risk} & \text{Eğer } 10 \leq FRS \leq 19 \\ 2 : \text{Yüksek Risk} & \text{Eğer } 20 \leq FRS \end{cases}$$

Framingham risk skorunu etkilediği düşünülen bağımsız değişkenler ise yeni risk faktörleri olarak değerlendirilen, hastanın *Low-density lipoprotein*, C-reaktif protein, homosistein, ferritin, trigliserit, ürikasit ve kreatinin değerleridir. Bu sürekli yapıdaki bağımsız değişkenler Tablo 4.1’de verildiği gibi kategorilere ayrılmıştır.

Tablo 4.1. Çalışmada Kullanılan Bağımsız Değişkenler ve Düzeyleri.

Bağımsız Değişken (Kısaltması)	Bağımsız Değişken Düzeyleri
<i>Low-density lipoprotein</i> (Ldl)	0: Ldl < 130mg/dl Normal
	1: Ldl ≥ 130mg/dl Yüksek
C-reaktif protein (Crp)	0: Crp < 1mg/dl Normal
	1: Crp ≥ 1mg/dl Yüksek
Homosistein (Hs)	0: Hs < 15 mg/dl Normal
	1: Hs ≥ 15 mg/dl Yüksek
Ferritin (Fr)	0: Fr < 100 mg/dl Normal
	1: Fr ≥ 100 mg/dl Yüksek
Trigliserit (Tg)	0: Tg < 150 mg/dl Normal
	1: Tg ≥ 150 mg/dl Yüksek
Ürikasit (Ür)	0: Ür < 10 mg/dl Normal
	1: Ür ≥ 10 mg/dl Yüksek
Kreatinin (Kr)	0: Kr < 1,2 mg/dl Normal
	1: Kr ≥ 1,2 mg/dl Yüksek

Uygulama bölümünde yeni risk faktörleri olarak değerlendirilen, hastanın *Low-density lipoprotein*, C-reaktif protein, homosistein, ferritin, trigliserit, ürikasit ve kreatinin değişkenlerindeki kategori değişimlerinin FRS’una; dolayısı ile 10 yıl içinde koroner kalp hastalığı geçirme riskine olan etkileri, Bölüm 3’de verilen *Multinomial Logit* ve Sıralı *Logit* modelleri ile çok değişkenli (multivariate) olarak incelenecektir. Sıralı *Logit* modellerinde Orantısal Risk varsayımı test edilecek ve bu varsayım testinin sonucuna göre gerekirse Orantısal Olmayan Risk modeli uygulanacaktır.

İncelenen modellerin en çok olabilirlik kestiricileri ve *odds* oranları elde edilecek ve karşılaştırılacaktır. Modellerin geçerliliği Olabilirlik Oranı test istatistiği ile test edilirken, Sıralı *Logit* modellerinden Orantısız ve Orantısız Olmayan Risk modelleri arasında veriye en iyi uyum sağlayan model, uyum iyiliği göstergelerinden Akaike Bilgi Kriteri (AIC), Bayesci Bilgi Kriteri (BIC) ve McFadden'ın Olabilirlik Oranı göstergesi yardımıyla belirlenecektir. Böylelikle tez çalışmasının sonunda uygun *Logit* modelinin seçiminde göz önüne alınması gereken ölçütler belirlenmiş olacaktır.

BEŞİNCİ BÖLÜM

5. BULGULAR

5.1. Tanımlayıcı İstatistikler

Hacettepe Üniversitesi İç Hastalıkları Anabilim Dalı Geriatri Ünitesine Şubat 2002 ile Haziran 2005 tarihleri arasında başvuran 65 yaş üstü hastalara ait yaş, *low-density lipoprotein* (Ldl), C-reaktif protein (Crp), homosistein (Hs), ferritin (Fr), trigliserit (Tg), ürikasit (Ür) ve kreatinin (Kr) değerlerine ilişkin tanımlayıcı istatistikler Tablo 5.1’de verilmiştir.

Tablo 5.1. Yaş, *Low-density lipoprotein* (Ldl), C-reaktif protein (Crp), Homosistein (Hs), Ferritin (Fr), Trigliserit (Tg), Ürikasit (Ür) ve Kreatinin (Kr) Değerlerinin Tanımlayıcı İstatistikleri¹.

Değişken	N	Ortalama	Standart Sapma	Minimum	Maksimum
Yaş	1174	72,008	6,062	65,00	103,00
Ldl	1172	128,560	43,058	11,00	657,00
Crp	1106	0,799	1,426	0,01	20,50
Hs	1099	17,451	8,735	2,00	101,00
Fr	1139	100,418	88,570	0,00	950,00
Tg	1172	146,603	77,157	40,00	591,00
Ür	1119	5,532	1,863	0,46	22,00
Kr	1166	0,914	0,292	0,00	3,09

Çalışmaya katılan 1174 kişinin % 37,9’u (n=445) erkek ve % 62,1’i (n=729) kadındır. Çalışmada sıralı yapıdaki bağımlı değişken, Framingham risk skorunun frekans dağılımı ise Tablo 5.2’de verilmiştir.

¹ Çalışmaya dahil edilen toplam kişi sayısı 1174’dir. Ancak bazı ölçümlerde eksik veriler bulunmaktadır. Tanımlayıcı istatistiklerin hesaplanmasında geçerli (valid) gözlem sayısı kullanılmıştır.

Tablo 5.2. Framingham Risk Skoru Frekans Dağılımı.

FRS	Sayı	%
Düşük Risk Grubu	408	34,75
Orta Risk Grubu	440	37,48
Yüksek Risk Grubu	326	27,77
Toplam	1174	100,00

Framingham risk skoru gruplarında yaş, *low-density lipoprotein*, C-reaktif protein, homosistein, ferritin, trigliserit, ürikasit ve kreatinin değerlerine ilişkin tanımlayıcı istatistikler Tablo 5.3’de verilmiştir.

Tablo 5.3. Framingham Risk Skoru Gruplarında Yaş, Ldl, Crp, Hs, Fr, Tg, Ür ve Kr Değerlerinin Tanımlayıcı İstatistikleri.

FRS	Değişken	N	Ortalama	Standart Sapma
Düşük Risk Grubu	Yaş	408	69,323	4,45
	Ldl	408	127,042	38,98
	Crp	385	0,745	1,40
	Hs	390	15,634	8,09
	Fr	406	97,905	73,31
	Tg	407	145,486	72,08
	Ür	391	5,118	1,96
	Kr	403	0,801	0,23
Orta Risk Grubu	Yaş	440	71,65	5,16
	Ldl	440	130,38	46,28
	Crp	414	0,717	1,10
	Hs	412	18,00	8,26
	Fr	426	98,78	105,47
	Tg	439	137,58	69,11
	Ür	420	5,622	1,79
	Kr	439	0,936	0,28
Yüksek Risk Grubu	Yaş	326	75,85	6,90
	Ldl	324	127,98	43,43
	Crp	307	0,980	1,78
	Hs	297	19,06	9,72
	Fr	307	106,00	81,03
	Tg	326	160,14	90,70
	Ür	308	5,93	1,72
	Kr	324	1,02	0,32

Low-density lipoprotein , C-reaktif protein, homosistein, ferritin, trigliserit, ürikasit ve kreatinin düzeylerinin (Bkz. Tablo 4.1) frekans dağılımları ise Tablo 5.4’de verilmiştir.

Tablo 5.4. Ldl, Crp, Hs, Fr, Tg, Ür ve Kr Düzeylerinin Frekans Dağılımları.

Değişken	Düzeyi	Sayı	%
Ldl	Normal	616	52,56
	Yüksek	556	47,44
	Toplam	1172	100,00
Crp	Normal	879	79,48
	Yüksek	227	20,52
	Toplam	1106	100,00
Hs	Normal	487	44,31
	Yüksek	612	55,69
	Toplam	1099	100,00
Fr	Normal	744	65,32
	Yüksek	395	34,68
	Toplam	1139	100,00
Tg	Normal	742	63,31
	Yüksek	430	36,69
	Toplam	1172	100,00
Ür	Normal	1105	98,75
	Yüksek	14	1,25
	Toplam	1119	100,00
Kr	Normal	1008	86,45
	Yüksek	158	13,55
	Toplam	1166	100,00

5.2. Multinomial Logit Model

Multinomial Logit çözümlemesinde düşük risk grubu (FRS=0) referans kategorisi olarak alındığında Tablo 5.5'deki sonuçlara ulaşılmıştır.

Tablo 5.5. *Multinomial Logit* Modelinin Katsayı, Standart Hata ve Odds Oranı Kestirimleri (n=996)².

FRS	Değişken	Katsayı (β)	Standart Hata	Odds Oranı (e^β)	p değeri
Düşük Risk'e karşı Orta Risk (Karşılaştırma 1)	Eşik 1	-0,441	0,160	---	---
	Ldl	0,016	0,157	1,016	0,941
	Crp	-0,052	0,197	0,949	0,790
	Hs	0,888	0,156	2,430	0,000*
	Fr	-0,172	0,163	0,842	0,289
	Tg	0,025	0,163	1,025	0,874
	Ür	-0,036	0,714	0,965	0,959
	Kr	0,873	0,291	2,394	0,003*
Düşük Risk'e karşı Yüksek Risk (Karşılaştırma 2)	Eşik 2	-1,162	0,186	---	---
	Ldl	-0,098	0,173	0,907	0,572
	Crp	0,273	0,205	1,314	0,183
	Hs	0,883	0,173	2,418	0,000*
	Fr	0,162	0,175	1,176	0,353
	Tg	0,475	0,175	1,608	0,007*
	Ür	-0,442	0,842	0,643	0,599
	Kr	1,464	0,291	4,323	0,000*

* p<0,05

Sadece sabit terimin bulunduğu *Multinomial Logit* modelin Log- Olabilirliği $L_0 = -1085,887$ ve Tablo 5.5'de verilen modelin Log-Olabilirliği $L_1 = -1034,148$ olarak hesaplanmıştır. *Multinomial Logit* modelin geçerliliği Olabilirlik Oranı testi ile incelendiğinde, model anlamlı bulunmuştur (p=0,000). *Multinomial Logit* modelin uyum iyiliği göstergeleri ise Tablo 5.6 ile verilmiştir.

² Bağımsız değişken düzeylerinde eksik veri bulunduğundan dolayı tüm *Logit* çözümlemelerinde geçerli gözlem sayısı 996'dır.

Tablo 5.6. *Multinomial Logit* Modelinin Uyum İyiliği Göstergeleri.

Uyum İyiliği Göstergesi	
Sapma Ölçüsü	2068,295
McFadden'ın R^2	0,03291
AIC	2,10873
BIC	-4697,377

Multinomial Logit model anlamlı olduğu için kestirilen katsayı ve *odds* oranları yorumlanabilmektedir.

Düşük Risk'e karşı Orta Risk (Karşılaştırma 1): Bu karşılaştırmada her bir bağımsız değişkendeki kategori değişiminin, orta risk grubunun ($10 \leq \text{FRS} \leq 19$) referans kategorisine ($0 \leq \text{FRS} \leq 9$) göre riskini (*odds*'unu) ne kadar attırdığı ya da azalttığı aşağıdaki şekilde yorumlanmaktadır:

Ldl: *Low-density lipoprotein* düzeyi yüksek olan bir hastanın düşük risk grubu yerine orta risk grubunda olma olasılığı, *Low-density lipoprotein* düzeyi normal olan bir hastaya göre 1,016 kat fazladır.

Crp: C-reaktif protein düzeyi yüksek olan bir hastanın düşük risk grubu yerine orta risk grubunda olma olasılığı, C-reaktif protein düzeyi normal olan bir hastaya göre 0,949 kat fazladır.

Hs: Homosistein düzeyi yüksek olan bir hastanın düşük risk grubu yerine orta risk grubunda olma olasılığı, Homosistein düzeyi normal olan bir hastaya göre 2,43 kat fazladır.

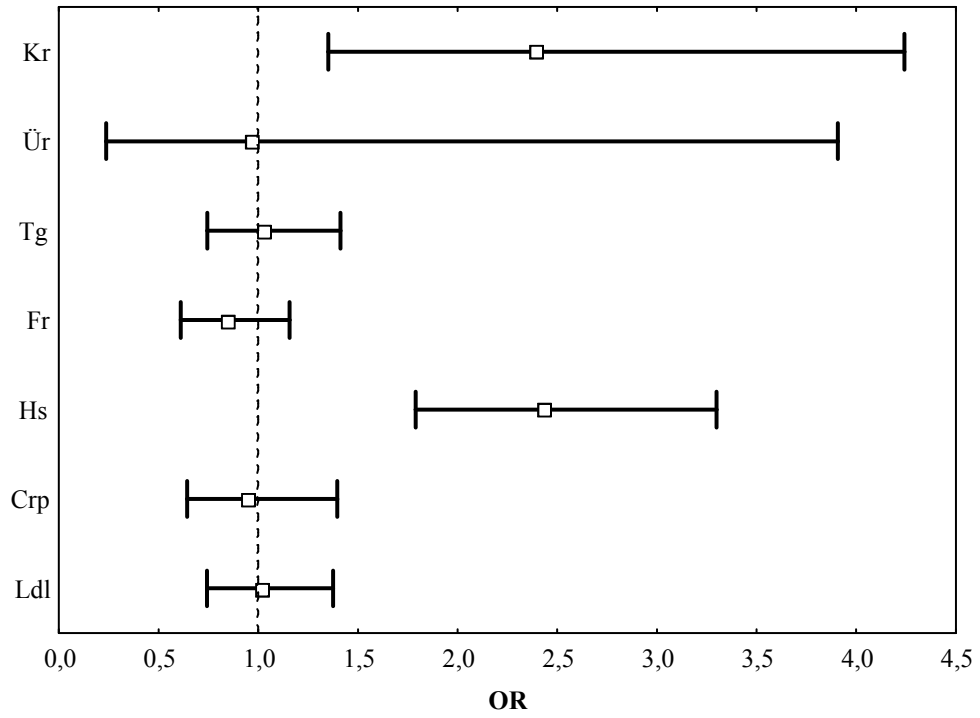
Fr: Ferritin düzeyi yüksek olan bir hastanın düşük risk grubu yerine orta risk grubunda olma olasılığı, Ferritin düzeyi normal olan bir hastaya göre 0,842 kat fazladır.

Tg: Trigliserit düzeyi yüksek olan bir hastanın düşük risk grubu yerine orta risk grubunda olma olasılığı, Trigliserit düzeyi normal olan bir hastaya göre 1,025 kat fazladır.

Ür: Ürikasit düzeyi yüksek olan bir hastanın düşük risk grubu yerine orta risk grubunda olma olasılığı, Ürikasit düzeyi normal olan bir hastaya göre 0,965 kat fazladır.

Kr: Kreatinin düzeyi yüksek olan bir hastanın düşük risk grubu yerine orta risk grubunda olma olasılığı, Kreatinin düzeyi normal olan bir hastaya göre 2,394 kat fazladır.

Tablo 5.5’de düşük riske karşı orta riski karşılaştıran *Logit* için elde edilen *Odds* Oranları ve % 95 güven aralıkları Şekil 5.1 ile verilmiştir.



Şekil 5.1. Düşük Riske karşı Orta Riski karşılaştıran *Logit* için elde edilen *Odds* Oranları ve % 95 güven aralıkları.

Şekil 5.1’de görüldüğü gibi *odds* oranlarından yalnızca Homosistein ve Kreatinin değerlerine ilişkin *odds* oranları istatistiksel açıdan önemlidir³.

³ Diğer bağımsız değişkenlere ilişkin *Odds* oranları istatistiksel açıdan önemsizdir. Ancak burada amaç diğer bağımsız değişkenlerin etkilerine göre düzeltilmiş katsayıları bulmak olduğundan tüm bağımsız değişkenler (yeni risk faktörleri) modelde tutulmuştur.

Düşük Risk'e karşı Yüksek Risk (Karşılaştırma 2): Bu karşılaştırmada her bir bağımsız değişkendeki kategori değişiminin, yüksek risk grubunun ($20 \leq \text{FRS}$) referans kategorisine ($0 \leq \text{FRS} \leq 9$) göre riskini ne kadar attırdığı ya da azalttığı aşağıdaki şekilde yorumlanmaktadır:

Ldl: *Low-density lipoprotein* düzeyi yüksek olan bir hastanın düşük risk grubu yerine yüksek risk grubunda olma olasılığı, *Low-density lipoprotein* düzeyi normal olan bir hastaya göre 0,907 kat fazladır.

Crp: C-reaktif protein düzeyi yüksek olan bir hastanın düşük risk grubu yerine yüksek risk grubunda olma olasılığı, C-reaktif protein düzeyi normal olan bir hastaya göre 1,314 kat fazladır.

Hs: Homosistein düzeyi yüksek olan bir hastanın düşük risk grubu yerine yüksek risk grubunda olma olasılığı, Homosistein düzeyi normal olan bir hastaya göre 2,418 kat fazladır.

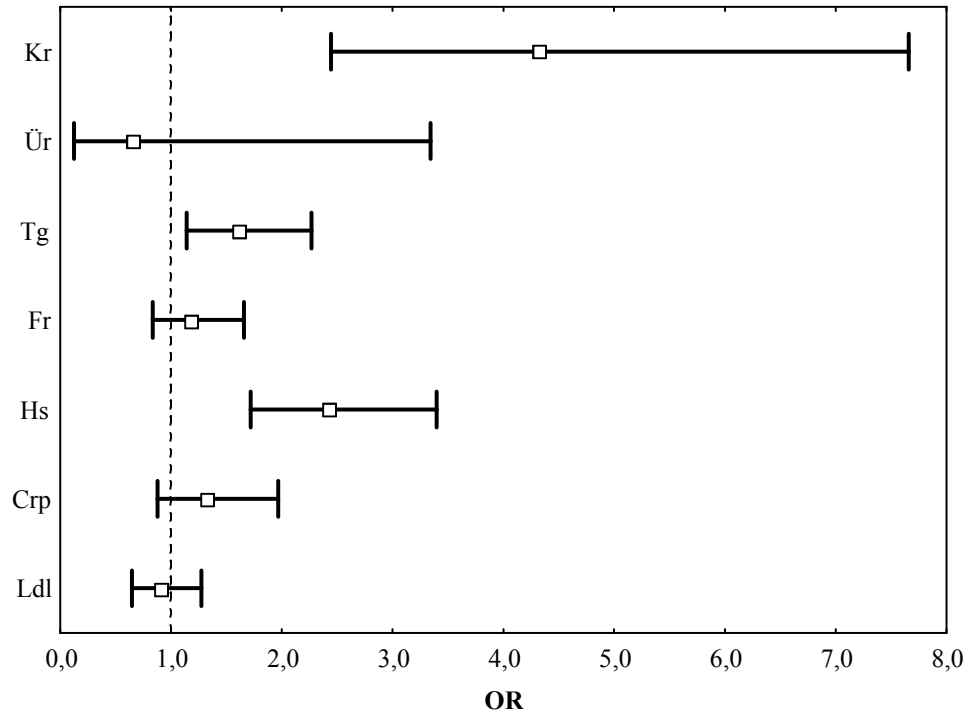
Fr: Ferritin düzeyi yüksek olan bir hastanın düşük risk grubu yerine yüksek risk grubunda olma olasılığı, Ferritin düzeyi normal olan bir hastaya göre 1,176 kat fazladır.

Tg: Trigliserit düzeyi yüksek olan bir hastanın düşük risk grubu yerine yüksek risk grubunda olma olasılığı, Trigliserit düzeyi normal olan bir hastaya göre 1,608 kat fazladır.

Ür: Ürikasit düzeyi yüksek olan bir hastanın düşük risk grubu yerine yüksek risk grubunda olma olasılığı, Ürikasit düzeyi normal olan bir hastaya göre 0,643 kat fazladır.

Kr: Kreatinin düzeyi yüksek olan bir hastanın düşük risk grubu yerine yüksek risk grubunda olma olasılığı, Kreatinin düzeyi normal olan bir hastaya göre 4,323 kat fazladır.

Tablo 5.5’de düşük riske karşı yüksek riski karşılaştıran *Logit* için elde edilen *Odds* Oranları ve % 95 güven aralıkları Şekil 5.2 ile verilmiştir.



Şekil 5.2. Düşük Riske karşı Yüksek Riski karşılaştıran *Logit* için elde edilen *Odds* Oranları ve % 95 güven aralıkları.

Şekil 5.2’de görüldüğü gibi *odds* oranlarından Homosistein, Trigliserit ve Kreatinin değerlerine ilişkin *odds* oranları istatistiksel açıdan önemlidir.

5.3. Orantısal Risk Modeli

Sıralı *Logit* modellerinden Orantısal Risk modeli veriye uygulandığında Tablo 5.7’deki sonuçlara ulaşılmıştır.

Tablo 5.7. Orantısal Risk Modelinin Katsayı, Standart Hata ve Odds Oranı Kestirimleri.

FRS	Değişken	Katsayı ⁴ (β)	Standart Hata	Odds Oranı (e^{β})	p değeri
Düşük Risk'e karşı Orta ve Yüksek Risk	Eşik 1	0,056	0,131	---	---
Düşük ve Orta Risk'e karşı Yüksek Risk	Eşik 2	1,743	0,143	---	---
	Ldl	-0,066	0,122	0,936	0,589
	Crp	0,194	0,149	1,214	0,192
	Hs	0,710	0,124	2,034	0,000*
	Fr	0,115	0,125	1,122	0,361
	Tg	0,339	0,125	1,404	0,007*
	Ür	-0,412	0,548	0,662	0,452
	Kr	0,944	0,183	2,570	0,000*

* p<0,05

Sadece sabit terimin bulunduğu Orantısal Risk modelinin Log- Olabilirliği $L_0 = -1085,887$ ve Tablo 5.7’de verilen modelin Log-Olabilirliği $L_1 = -1042,827$ olarak hesaplanmıştır. Orantısal Risk modelinin geçerliliği Olabilirlik Oranı testi ile incelendiğinde, model anlamlı bulunmuştur (p=0,000). Orantısal Risk modelinin uyum iyiliği göstergeleri ise Tablo 5.8 ile verilmiştir.

⁴ Orantısal Risk Modeli için “Daha düşük riskli grup/gruplar” referans olarak alındığından Eşitlik 3.26 ve 3.29’daki $-\beta$ katsayısının negatif işaretlisi (diğer bir ifade ile β) alınmıştır.

Tablo 5.8. Orantısal Risk Modelinin Uyum İyiliği Göstergeleri.

Uyum İyiliği Göstergesi	
Sapma Ölçüsü	2085,654
McFadden'ın R²	0,03136
AIC	2,112102
BIC	-4728,344

Orantısal Risk modeli anlamlı olmasına rağmen bu modelin kullanabilmesi için öncelikle Orantısal Risk varsayımının geçerliliğinin test edilmesi gereklidir. Orantısal Risk Modelinin yaklaşık olarak geçerliliğinin testi Bölüm 3.3.2.3'de verildiği gibi aşağıdaki şekilde yapılır:

$$G^2 = -2(L_1 - L_2) = -2[-1042,827 - (-1034,379)] = 16,896$$

G^2 Olabilirlik Oranı test istatistiği $\chi^2_{7,0,05} = 14,067$ değerinden büyük olduğu için Orantısal Risk varsayımı sağlanmamaktadır ($p = 0,018$). Ancak bu test yaklaşık bir test olduğundan Brant'ın Wald testi ile de varsayımın sağlanıp sağlanmadığı incelenmiş ve Tablo 5.9 ile verilen sonuçlara ulaşılmıştır.

Tablo 5.9. Orantısal Risk Varsayımının Geçerliliğinin Testi (Brant Testi).

Değişken	Ki-kare değeri	p değeri	sd
Tümü	18,25	0,011*	7
Ldl	0,20	0,651	1
Crp	1,34	0,248	1
Hs	8,55	0,003*	1
Fr	3,29	0,070	1
Tg	2,53	0,112	1
Ür	0,10	0,752	1
Kr	0,83	0,361	1

* $p < 0,05$

Tablo 5.9'a göre de Orantısal Risk varsayımı sağlanmamaktadır. Bu testin orantısallığın hangi değişken veya değişkenler yüzünden bozulduğu hakkında bilgi

verdiğinden daha önce söz edilmişti. Tablo 5.9'a göre orantısallığı bozan bağımsız değişken Homosistein'dir.

Orantısal Risk varsayımı sağlanmadığından ve Tablo 5.7'den elde edilen katsayılar ve *odds* oranları, gerçekte doğru olmayan bir varsayım üzerinden elde edildiklerinden dolayı, bu katsayıları yorumlamak hatalı olacaktır. Dolayısı ile Tablo 5.7'deki katsayı ve *odds* oranlarının yorumlanması üzerinde durulmamıştır.

5.4. Orantısal Olmayan Risk Modeli

Sıralı *Logit* modellerinden Orantısal Olmayan Risk modeli veriye uygulandığında Tablo 5.10'daki sonuçlara ulaşılmıştır.

Tablo 5.10. Orantısal Olmayan Risk Modelinin Katsayı, Standart Hata ve *Odds* Oranı Kestirimleri.

FRS	Değişken	Katsayı ⁵ (β)	Standart Hata	<i>Odds</i> Oranı (e^{β})	p değeri
Düşük Risk'e karşı Orta ve Yüksek Risk (Karşılaştırma 1)	Eşik 1	-0,064	0,143	---	---
	Ldl	-0,025	0,139	0,975	0,856
	Crp	0,099	0,173	1,104	0,564
	Hs	0,872	0,140	2,392	0,000*
	Fr	0,007	0,144	1,007	0,958
	Tg	0,223	0,144	1,250	0,121
	Ür	-0,351	0,677	0,704	0,603
	Kr	1,124	0,271	3,077	0,000*
Düşük ve Orta Risk'e karşı Yüksek Risk (Karşılaştırma 2)	Eşik 2	-1,676	0,170	---	---
	Ldl	-0,105	0,149	0,901	0,485
	Crp	0,296	0,174	1,344	0,088
	Hs	0,445	0,153	1,560	0,004*
	Fr	0,266	0,151	1,305	0,078
	Tg	0,471	0,150	1,602	0,002*
	Ür	-0,507	0,715	0,602	0,478
	Kr	0,925	0,202	2,522	0,000*

* p<0,05

⁵ Orantısal Olmayan Risk Modeli için “Daha düşük riskli grup/gruplar” referans olarak alındığından Eşitlik 3.32 ve 3.35'deki $-\beta$ katsayısının negatif işaretlisi (diğer bir ifade ile β) alınmıştır.

Sadece sabit terimin bulunduğu Orantısız Olmayan Risk modelinin Log-Olabilirliği $L_0 = -1085,887$ ve Tablo 5.10’da verilen modelin Log-Olabilirliği $L_1 = -1034,380$ olarak hesaplanmıştır. Orantısız Olmayan Risk modelinin geçerliliği Olabilirlik Oranı testi ile incelendiğinde, model anlamlı bulunmuştur ($p=0,000$). Orantısız Olmayan Risk modelinin uyum iyiliği göstergeleri ise Tablo 5.11 ile verilmiştir.

Tablo 5.11. Orantısız Olmayan Risk Modelinin Uyum İyiliği Göstergeleri.

Uyum İyiliği Göstergesi	
Sapma Ölçüsü	2068,76
McFadden’ın R^2	0,03269
AIC	2,109197
BIC	-4696,912

Orantısız Olmayan Risk modeli anlamlı olduğu için kestirilen katsayı ve *odds* oranları yorumlanabilmektedir.

Düşük Risk'e karşı Orta ve Yüksek Risk (Karşılaştırma 1): Bu karşılaştırmada her bir bağımsız değişkendeki kategori değişiminin, orta ve yüksek risk grubunun ($10 \leq \text{FRS}$) düşük risk grubuna ($0 \leq \text{FRS} \leq 9$) göre riskini (*odds*’unu) ne kadar attırdığı ya da azalttığı aşağıdaki şekilde yorumlanmaktadır:

Ldl: *Low-density lipoprotein* düzeyi yüksek olan bir hastanın düşük risk grubu yerine daha riskli (orta + yüksek risk grubu) grupta olma olasılığı, *Low-density lipoprotein* düzeyi normal olan bir hastaya göre 0,975 kat fazladır.

Crp: C-reaktif protein düzeyi yüksek olan bir hastanın düşük risk grubu yerine daha riskli grupta olma olasılığı, C-reaktif protein düzeyi normal olan bir hastaya göre 1,104 kat fazladır.

Hs: Homosistein düzeyi yüksek olan bir hastanın düşük risk grubu yerine daha riskli grupta olma olasılığı, Homosistein düzeyi normal olan bir hastaya göre 2,392 kat fazladır.

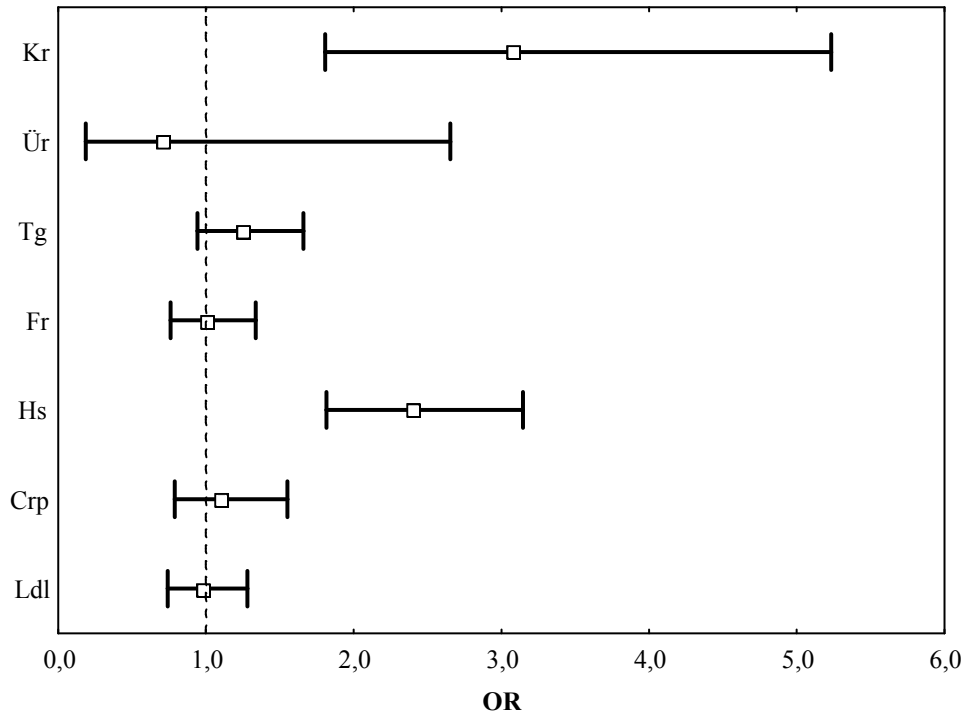
Fr: Ferritin düzeyi yüksek olan bir hastanın düşük risk grubu yerine daha riskli grupta olma olasılığı, Ferritin düzeyi normal olan bir hastaya göre 1,007 kat fazladır.

Tg: Trigliserit düzeyi yüksek olan bir hastanın düşük risk grubu yerine daha riskli grupta olma olasılığı, Trigliserit düzeyi normal olan bir hastaya göre 1,250 kat fazladır.

Ür: Ürikasit düzeyi yüksek olan bir hastanın düşük risk grubu yerine daha riskli grupta olma olasılığı, Ürikasit düzeyi normal olan bir hastaya göre 0,704 kat fazladır.

Kr: Kreatinin düzeyi yüksek olan bir hastanın düşük risk grubu yerine daha riskli grupta olma olasılığı, Kreatinin düzeyi normal olan bir hastaya göre 3,077 kat fazladır.

Tablo 5.10’da düşük riske karşı orta ve yüksek riski karşılaştıran *Logit* için elde edilen *Odds* Oranları ve % 95 güven aralıkları Şekil 5.3 ile verilmiştir.



Şekil 5.3. Düşük Riske karşı Orta ve Yüksek Riski karşılaştıran *Logit* için elde edilen *Odds* Oranları ve % 95 güven aralıkları.

Şekil 5.3’de görüldüğü gibi *odds* oranlarından yalnızca Homosistein ve Kreatinin değerlerine ilişkin *odds* oranları istatistiksel açıdan önemlidir.

Düşük ve Orta Risk'e karşı Yüksek Risk (Karşılaştırma 2): Bu karşılaştırmada her bir bağımsız değişkendeki kategori değişiminin, yüksek risk grubunun ($20 \leq \text{FRS}$) düşük ve orta risk grubuna ($0 \leq \text{FRS} \leq 19$) göre riskini (*odds*’unu) ne kadar attırdığı ya da azalttığı aşağıdaki şekilde yorumlanmaktadır:

Lld: *Low-density lipoprotein* düzeyi yüksek olan bir hastanın düşük ve orta risk grubu yerine daha riskli (yüksek risk grubu) grupta olma olasılığı, *Low-density lipoprotein* düzeyi normal olan bir hastaya göre 0,901 kat fazladır.

Crp: C-reaktif protein düzeyi yüksek olan bir hastanın düşük ve orta risk grubu yerine daha riskli grupta olma olasılığı, C-reaktif protein düzeyi normal olan bir hastaya göre 1,344 kat fazladır.

Hs: Homosistein düzeyi yüksek olan bir hastanın düşük ve orta risk grubu yerine daha riskli grupta olma olasılığı, Homosistein düzeyi normal olan bir hastaya göre 1,560 kat fazladır.

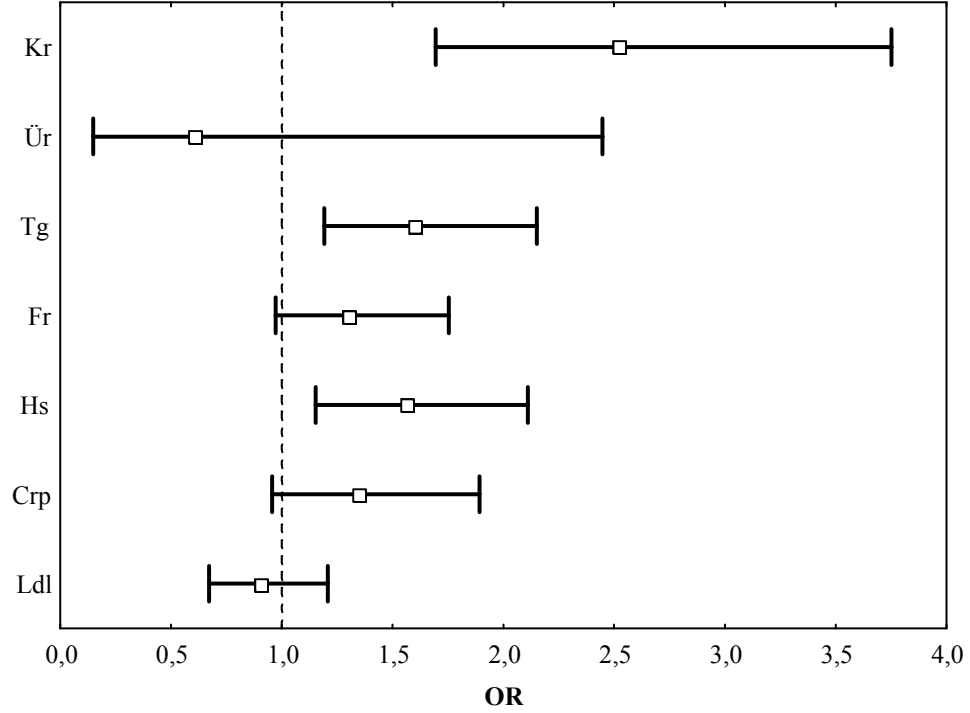
Fr: Ferritin düzeyi yüksek olan bir hastanın düşük ve orta risk grubu yerine daha riskli grupta olma olasılığı, Ferritin düzeyi normal olan bir hastaya göre 1,305 kat fazladır.

Tg: Trigliserit düzeyi yüksek olan bir hastanın düşük ve orta risk grubu yerine daha riskli grupta olma olasılığı, Trigliserit düzeyi normal olan bir hastaya göre 1,602 kat fazladır.

Ür: Ürikasit düzeyi yüksek olan bir hastanın düşük ve orta risk grubu yerine daha riskli grupta olma olasılığı, Ürikasit düzeyi normal olan bir hastaya göre 0,602 kat fazladır.

Kr: Kreatinin düzeyi yüksek olan bir hastanın düşük ve orta risk grubu yerine daha riskli grupta olma olasılığı, Kreatinin düzeyi normal olan bir hastaya göre 2,522 kat fazladır.

Tablo 5.10’da düşük ve orta riske karşı yüksek riski karşılaştıran *Logit* için elde edilen *Odds* Oranları ve % 95 güven aralıkları Şekil 5.4 ile verilmiştir.



Şekil 5.4. Düşük ve Orta Riske karşı Yüksek Riski karşılaştıran *Logit* için elde edilen *Odds* Oranları ve % 95 güven aralıkları.

Şekil 5.4’de görüldüğü gibi *odds* oranlarından yalnızca Homosistein, Trigliserit ve Kreatinin değerlerine ilişkin *Odds* oranları istatistiksel açıdan önemlidir.

ALTINCI BÖLÜM

6. TARTIŞMA ve SONUÇ

Bu çalışmada hipotez; bağımlı değişkenin sıralı yapıda olduğu ve Orantısız Risk Varsayımının sağlanmadığı durumlarda Orantısız Olmayan Risk modelinin, *Multinomial Logit* modelinden daha iyi bir seçenek olduğudur. Bu hipotezi test etmek için *Multinomial Logit* ve Sıralı *Logit* modellerinden Orantısız Risk ve Orantısız Olmayan Risk modelleri ile Framingham Risk skorunu etkileyen yeni risk faktörleri incelenmiştir.

Uygulamada kullanılan veriye Sıralı *Logit* modellerinden Orantısız Risk modeli ve Orantısız Olmayan Risk modeli uygulandığında her iki model de anlamlı bulunmuştur ($p=0,000$). Ancak Orantısız Risk modelinde, modelin temelini oluşturan “Orantısız Risk Varsayımı” sağlanmamıştır (Brant Testi; $p=0,011$). Dolayısı ile oluşturulan *Logit*’lerde bağımlı değişken için hesaplanan yığılımlı *odds*’lar birbirine eşit değildir. Tablo 5.7 ile verilen *odds* oranları ise doğru olmayan bir varsayım üzerinden elde edildiğinden verideki gerçek eğilimi yansıtmamaktadır.

Orantısız Olmayan Risk modelinde ise orantısızlık kısıtı olmadan yığılımlı *odds*’lar hesaplandığından, oluşturulan iki *Logit*’te de verinin gerçek eğilimini yansıtan *odds* oranları elde edilmiştir. Bu durum karşılaştırılmalı olarak Tablo 6.1 ile açıklanmaktadır.

Tablo 6.1. Orantısal Risk ve Orantısal Olmayan Risk Modelindeki *Odds* Oranı Kestirimlerinin Karşılaştırılması.

Orantısal Olmayan Risk Modeli			Orantısal Risk Modeli		
FRS	Değişkenler	OR	FRS	Değişkenler	OR
Düşük Risk'e karşı Orta ve Yüksek Risk (Karşılaştırma 1)	Ldl	0,975	Düşük Risk'e karşı Orta ve Yüksek Risk (Karşılaştırma 1)	Ldl	0,936
	Crp	1,104		Crp	1,214
	Hs	2,392*		Hs	2,034*
	Fr	1,007	Düşük ve Orta Risk'e karşı Yüksek Risk (Karşılaştırma 2)	Fr	1,122
	Tg	1,250		Tg	1,404*
	Ür	0,704		Ür	0,662
	Kr	3,077*		Kr	2,570*
Düşük ve Orta Risk'e karşı Yüksek Risk (Karşılaştırma 2)	Ldl	0,901			
	Crp	1,344			
	Hs	1,560*			
	Fr	1,305			
	Tg	1,602*			
	Ür	0,602			
	Kr	2,522*			

* p<0,05

Tablo 6.1’de görüldüğü gibi Orantısal Olmayan Risk modelinden elde edilen *odds* oranları Ldl değişkeni için sırasıyla 0,975 ve 0,901’dir. Orantısal Risk modeli ise Ldl değişkenine ait riskin bağımlı değişken kategorilerinde eşit ve 0,936 olduğunu (yaklaşık olarak 0,975 ve 0,901’in ortalaması) belirtmektedir. Ldl değişkeni için Orantısal Risk varsayımının testi, Orantısal Olmayan modelde oluşturulan Karşılaştırma 1 ve Karşılaştırma 2’deki *odds* oranlarının (yani 0,975 ve 0,901’in) birbirine eşit olup olmadığının testidir. Eğer bu iki *odds* oranı arasında istatistiksel açıdan fark yok ise tek bir *odds* oranı olarak belirtilmesinde bir sakınca yoktur; yani Orantısal Risk varsayımı sağlanmaktadır. Diğer yandan; Orantısal Risk varsayımının sağlanmadığı durumda Karşılaştırma 1 ve Karşılaştırma 2’deki *odds* oranları birbirine eşit olmadığından iki *odds* oranının tek bir *odds* oranı ile ifade edilmesi hatalı yorumlara neden olacaktır. Örneğin, Orantısal Risk varsayımını bozan Hs değişkeni için Orantısal Olmayan Risk modelinden elde edilen *odds* oranları sırasıyla 2,392 ve 1,560’dır. Ancak Orantısal Risk modeli Hs değişkenine ait yığılımlı *odds* oranının bağımlı değişken kategorilerinde eşit ve 2,034 olduğunu

belirtmektedir. Böyle bir durumdan sakınmak ve verideki gerçek eğilimin anlaşılması açısından Orantısal Risk varsayımının sağlanmadığı durumlarda Orantısal Olmayan Risk modelinin kullanılması gerekmektedir.

Orantısal Risk modeli ve Orantısal Olmayan Risk modeli uyum iyiliği açısından karşılaştırıldığında ise Tablo 6.2’deki sonuçlara ulaşılmıştır.

Tablo 6.2. Orantısal Risk ve Orantısal Olmayan Risk Modelinin Uyum İyiliği Göstergeleri Açısından Karşılaştırılması⁶.

Uyum İyiliği Göstergesi	Model		Fark
	Orantısal Risk	Orantısal Olmayan Risk	
Sapma Ölçüsü	2085,654	2068,76	16,894
McFadden’ın R²	0,03136	0,03269	-0,00133
AIC	2,112102	2,109197	0,002905
BIC	-4728,344	-4696,912	-31,432

Tablo 6.2’deki uyum iyiliği göstergelerine (Sapma ölçüsü, McFadden’ın R² ve AIC) göre Orantısal Olmayan Risk modeli veriye daha iyi uyum sağlamaktadır. BIC ise bunun tam tersini yani Orantısal Risk modelinin uyum iyiliğinin daha iyi olduğunu söylemektedir. Bu durumun sebebi Orantısal Risk modelinin Orantısal Olmayan Risk modeline göre 7 daha az parametre kestirimine sahip olması ve sapma ölçüsüne ilişkin serbestlik derecesinin (BIC bunu dikkate alıyor) daha yüksek olması olmaktadır. Diğer yandan Orantısal Risk modelinde Orantısal Risk varsayımı sağlanmamakta ve uyum iyiliği göstergelerinin çoğunluğu (Sapma ölçüsü, McFadden’ın R², AIC) Orantısal Olmayan Risk modelinin lehine sonuçlar vermektedir. Bu durum Orantısal Risk varsayımının sağlanmadığı durumlarda Orantısal Olmayan Risk modelinin kullanılması gerektiğini uyum iyiliği göstergeleri açısından da desteklemektedir.

⁶ Modellerin uyum iyiliğini karşılaştıran Tablo 6.2 ve 6.4’de “kalın ve italik” yazı tipi ile belirtilen rakamlar daha iyi uyum sağlayan modeli göstermektedir.

Uygulamada kullanılan veriye *Multinomial Logit* modeli ve Orantısız Olmayan Risk modeli uygulandığında her iki model de anlamlı bulunmuştur ($p=0,000$). Her iki modelden elde edilen *odds* oranları karşılaştırılmalı olarak Tablo 6.3 ile verilmiştir.

Tablo 6.3. Orantısız Olmayan Risk ve *Multinomial Logit* Modelindeki *Odds* Oranı Kestirimlerinin Karşılaştırılması.

Orantısız Olmayan Risk Modeli			<i>Multinomial Logit</i> Model		
FRS	Değişkenler	OR	FRS	Değişkenler	OR
Düşük Risk'e karşı Orta ve Yüksek Risk (Karşılaştırma 1)	Ldl	0,975	Düşük Risk'e karşı Orta Risk (Karşılaştırma 1)	Ldl	1,016
	Crp	1,104		Crp	0,949
	Hs	2,392*		Hs	2,430*
	Fr	1,007		Fr	0,842
	Tg	1,250		Tg	1,025
	Ür	0,704		Ür	0,965
	Kr	3,077*		Kr	2,394*
Düşük ve Orta Risk'e karşı Yüksek Risk (Karşılaştırma 2)	Ldl	0,901	Düşük Risk'e karşı Yüksek Risk (Karşılaştırma 2)	Ldl	0,907
	Crp	1,344		Crp	1,314
	Hs	1,560*		Hs	2,418*
	Fr	1,305		Fr	1,176
	Tg	1,602*		Tg	1,608*
	Ür	0,602		Ür	0,643
	Kr	2,522*		Kr	4,323*

* $p<0,05$

Multinomial Logit modeli sonuçlarına bakıldığında düşük riske karşı orta risk grubunun karşılaştırılmasında anlamlı bulunan değişkenler Homosistein ve Kreatinin'dir. Orantısız Olmayan Risk modelinde düşük riske karşı orta ve yüksek risk gruplarının karşılaştırılmasında anlamlı bulunan değişkenler Homosistein ve Kreatinin'dir. Aynı şekilde, *Multinomial Logit* modelinde düşük riske karşı yüksek risk grubunun karşılaştırılması ve Orantısız Olmayan Risk modelinde düşük ve orta riske karşı yüksek risk grubunun karşılaştırılmasında önemli bulunan değişkenler sırasıyla Homosistein, Triglicerit ve Kreatinin'dir. Ancak iki modelde de önemli bulunan değişkenler aynı olmasına karşın katsayıları birbirinden farklıdır.

Tablo 6.3’de görüldüğü üzere Orantısız Olmayan Risk modelinde Hs değişkeni için düşük riske karşı orta ve yüksek risk grubunun karşılaştırılmasıyla elde edilen *odds* oranı 2,392 iken; *Multinomial Logit* modelinde düşük riske karşı orta risk grubu karşılaştırmasından elde edilen 2,430’dur. *Odds* oranları arasındaki bu fark *Multinomial Logit* modelinde yapılan karşılaştırmanın yüksek risk grubunu hiç hesaba katmamasından kaynaklanmaktadır. Orantısız Olmayan Risk modelinde Kr değişkeni için düşük riske karşı orta ve yüksek risk grubunun karşılaştırılmasıyla elde edilen *odds* oranı 3,077 iken; *Multinomial Logit* modelinde düşük riske karşı orta risk grubunun karşılaştırmasından elde edilen 2,394’tür. Hs değişkeni için bu kadar belirgin olmayan bu fark Kr değişkeni için oldukça belirgindir. Aynı şekilde, Orantısız Olmayan Risk modelinde Tg değişkeni için düşük ve orta riske karşı yüksek risk grubunun karşılaştırılmasından elde edilen *odds* oranı 1,602 iken; *Multinomial Logit* modelinde düşük riske karşı yüksek risk grubu karşılaştırmasından elde edilen 1,608’dir. *Odds* oranları arasındaki bu fark *Multinomial Logit* modelinde yapılan karşılaştırmanın orta risk grubunu hiç hesaba katmamasından kaynaklanmaktadır. Orantısız Olmayan Risk modelinde Kr değişkeni için düşük ve orta riske karşı yüksek risk grubunun karşılaştırılmasıyla elde edilen *odds* oranı 2,522 iken; *Multinomial Logit* modelinde düşük riske karşı yüksek risk karşılaştırmasından elde edilen 4,323’tür. Tg değişkeni için bu kadar belirgin olmayan bu fark Kr değişkeni için oldukça belirgindir. Dolayısı ile *odds* oranlarındaki bu farklılık *Multinomial Logit* modelinin sıralı yapıyı kullanmamasından kaynaklanmaktadır. Bu durumda akla gelen bir soru da “Orantısız Risk varsayımının sağlanmadığı durumlarda *Multinomial Logit* modelinin mi yoksa Orantısız Olmayan Risk modelinin mi kullanılacağı” veya “hangi modelden elde edilen katsayının yorumlanacağı” olmaktadır.

Multinomial Logit ve Orantısız Olmayan Risk modeli uyum iyiliği açısından karşılaştırıldığında ise Tablo 6.4’deki sonuçlara ulaşılmıştır.

Tablo 6.4. *Multinomial Logit* ve Orantısız Olmayan Risk Modelinin Uyum İyiliği Göstergeleri Açısından Karşılaştırılması.

Uyum İyiliği Göstergesi	Model		Fark
	<i>Multinomial Logit</i>	Orantısız Olmayan Risk	
Sapma Ölçüsü	2068,295	2068,76	-0,465
McFadden'ın R^2	0,03291	0,03269	0,00022
AIC	2,10873	2,109197	-0,000467
BIC	-4697,377	-4696,912	-0,465

Tablo 6.3'deki uyum iyiliği göstergelerine (Sapma ölçüsü, McFadden'ın R^2 , AIC ve BIC) göre *Multinomial Logit* modeli veriye daha iyi uyum sağlamaktadır. Ancak modellerin uyum iyiliği göstergeleri arasındaki farkların oldukça küçük olduğu görülmektedir (Sapma Ölçüsü farkı: -0,465; McFadden'ın R^2 farkı: 0,00022; AIC farkı: -0,000467; BIC farkı: -0,465).

Fujimoto, “Japon kadın çalışanları” ile ilgili yaptığı çalışmada *Multinomial Logit*, Orantısız Risk ve Orantısız Olmayan Risk modelini uygulamış ve uyum iyiliği açısından bu modelleri karşılaştırmıştır. Bu çalışmada da bağımlı değişken sıralı yapıda olmasına rağmen *Multinomial Logit* modelinin Orantısız Olmayan Risk modelinden daha iyi uyum sağladığı, ancak uyum iyiliği göstergeleri arasındaki farkların oldukça küçük olduğu (Sapma Ölçüsü farkı: -0,404; McFadden'ın R^2 farkı: 0,000; AIC farkı: -0,001; BIC farkı: -0,404) belirtilmiştir (21). Bu sonuçlar elde edilen sonuçlar ile benzerlik göstermektedir. Ancak burada uyum iyiliği göstergelerine bakılarak Orantısız Risk varsayımının sağlanmadığı durumlarda *Multinomial Logit* modelinin Orantısız Olmayan Risk modelinden daha iyi bir seçenek olup olmadığına ilişkin kesin bir sonuç elde edilemez. Çünkü daha önce de belirtildiği gibi *Multinomial Logit* ve Orantısız Olmayan Risk modelinin risk yaklaşımları birbirlerinden farklıdır. Bu duruma aşağıda verilen Şekil 6.1 ile açıklık getirilmektedir.

Orantısız / Orantısız Olmayan Risk Modeli	Düşük Risk	Orta Risk	Yüksek Risk
Karşılaştırma 1	Referans		
Karşılaştırma 2	Referans		

<i>Multinomial Logit Model</i>	Düşük Risk	Orta Risk	Yüksek Risk
Karşılaştırma 1	Referans		
Karşılaştırma 2	Referans		

Şekil 6.1. *Multinomial Logit* ve Orantısız/Orantısız Olmayan Risk modelinde referans ve karşılaştırılan kategori/kategorilerin gösterimi⁷.

Şekil 6.1’de görüldüğü gibi Orantısız Olmayan Risk modelinde, Orantısız Risk modelinde olduğu gibi oluşturulan *Logit*’lerde referans ve karşılaştırılan kategoriler yığılımlı olarak ele alınmaktadır. Dolayısı ile Orantısız ve Orantısız Olmayan Risk modelinin risk yaklaşımı yığılımlıdır. *Multinomial Logit* modelinde ise sadece seçilen referans kategorisine göre diğer kategorilerin karşılaştırılması nominal olarak yapılmakta ve bağımlı değişkenin sıralı yapısı gözardı edilmektedir. Bu noktadan bakıldığında Orantısız Risk varsayımının sağlanmadığı durumlarda Orantısız Olmayan Risk modelinin kullanımı, bağımlı değişkenin yapısındaki tüm bilgiyi kullanması açısından avantaj sağlamakta ve bağımlı değişkenin sıralı yapısı nedeni ile yığılımlı risk yaklaşımına uyum sağlamaktadır.

Orantısız Risk varsayımının sağlanmadığı durumlarda, *Multinomial Logit* modeli ile karşılaştırılabilecek bir diğer model de Sıralı *Logit* modellerinden Kısmi Orantısız Risk modelidir (Partial Proportional Odds Model). Kısmi Orantısız Risk modeli Orantısız Risk varsayımının kimi değişkenler için sağlandığı kimisi için sağlanmadığı durumda kullanılan bir modeldir (22). Dolayısı ile hem Orantısız Risk modelinin hem de Orantısız Olmayan Risk modelinin özelliklerini taşımaktadır. Bulgular bölümünde Tablo 5.9’a göre veride orantısallığı bozan değişkenin

⁷ Şekil 6.1’de “Referans” ile belirtilen alanlar *Multinomial Logit*, Orantısız ve Orantısız Olmayan Risk modelinde oluşturulan *Logit*’lerin referans kategori/kategorilerini, taralı alanlar ise karşılaştırılan grup/grupları göstermektedir.

Homosistein olduğundan daha önce söz edilmişti. Homosistein değişkeni için orantısallığın sağlanmadığı, kalan diğer değişkenler için orantısallığın sağlandığı Kısmi Orantısal Risk modeli veriye uygulandığında, orantısallığın tüm değişkenler için sağlanmadığı Orantısal Olmayan Risk modelinden daha iyi bir sonuç vereceği düşünülmektedir. Dolayısı ile de daha sonraki çalışmalarda Orantısal Risk varsayımının sağlanmadığı durumlar için Kısmi Orantısal Risk modelinin *Multinomial Logit* modelinden daha iyi bir seçenek olup olmadığının araştırılması önerilmektedir.

KAYNAKLAR

1. Hosmer, D. W. ve Lemeshow, S. (2000) Applied Logistic Regression, Second Edition, Wiley, New York.
2. Bender, R. ve Grouven U. (1998) Using Binary Logistic Regression Models for Ordinal Data with Non-Proportional Odds. Journal of Clinical Epidemiology, Vol.51, No. 10, 809-816.
3. Hardin, J. ve Hilbe, J. (2001) Generalized Linear Models and Extensions, Stata Press, College Station, Texas.
4. Agresti, A. (2002) Categorical Data Analysis, Second Edition, Wiley, Hoboken, New Jersey.
5. McCullagh, P. ve Nelder, J. A. (1989) Generalized Linear Models, Second Edition, Chapman and Hall, London.
6. Liao, T. F. (1994) Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models, Sage Publications, Thousand Oaks, CA.
7. Fox, J. (1997) Applied Regression Analysis, Linear Models, and Related Methods, Sage Publications, Thousand Oaks, CA.
8. Long, S. J. (1997) Regression Models for Categorical and Limited Dependent Variables, Sage Publications, Thousand Oaks, CA
9. Ordinal and Multinomial Logit Models. [http:// www. fiu.edu. /~tardanic/lec7.pps#21](http://www.fiu.edu/~tardanic/lec7.pps#21) [29/07/2005].
10. Fu, V. K. Gologit: Stata Module to Estimate Generalized Ordered Logit Models, [http://fmwww. bc. edu./ repec / bcode / g / gologit. hlp](http://fmwww.bc.edu/repec/bcode/g/gologit.hlp). [02/08/2005]
11. POL 682: Some Notes on the Proportional Odds/ Parallel Slopes Assumption. [http://www.u.arizona.edu/ ~bsjones/gologit.doc](http://www.u.arizona.edu/~bsjones/gologit.doc). [02/08/2005]
12. Frequently Asked Questions about Gologit. [http:// www. soc. utah. edu/ ~vfu/gologit/gologitfaq. html](http://www.soc.utah.edu/~vfu/gologit/gologitfaq.html). [02/08/2005]

13. Brant, R. (1990) Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression. *Biometrics* 46:1771-1178.
14. Brant Test of Parallel Regression Assumption. <http://www.indiana.edu/~jslsoc/stata/spostado/brant.hlp>. [29/07/2005]
15. Lloyd, J. D. M. Wilson, P. W. Larson, M. G. Beiser, A. Leip, E. P. D'Agostino, R. B. Levy, D. (2004) Framingham risk score and prediction of lifetime risk for coronary heart disease. *Am J Cardiol.* Jul 1; 94 (1):20-24.
16. Wilson, P. W. D'Agostino, R. B. Levy, D. Belanger, A. M. Silbershatz, H. Kannel, W. B. (1998) Prediction of coronary heart disease using risk factor categories. *Circulation.* May 12; 97 (18):1837-1847.
17. Gordon, T. Kannel, W. B. (1982) Multiple risk functions for predicting coronary heart disease: the concept, accuracy, and application. *Am. Heart J.* Jun; 103 (6):1031-1039.
18. Salonen, J. T. Nyyssonen, K. Korpela, H. Tuomilehto, J. Seppanen, R. Salonen, R. (1992) High stored iron levels are associated with excess risk of myocardial infarction in eastern Finnish men. *Circulation.* Sep; 86 (3):803-811.
19. Ascherio, A. Willett, W. C. (1994) Are body iron stores related to the risk of coronary disease. *N. Engl. J. Med.* 330:1152–1154.
20. Kaltwasser, J. P. Werner, E. (1989) Diagnosis and clinical evaluation of iron overload. *Baillieres Clinical Haematology* 2:363–385.
21. Fujimoto, K. (2003) Application of Multinomial and Ordinal Regressions to the Data of Japanese Female Labor Market. <http://etd.library.pitt.edu/ETD/available/etd-12162003-153155/unrestricted/Fujimoto-12-16-03b.pdf>. [12/07/2005]
22. Peterson, B.ve F. Harrel, Jr. (1990) Partial Proportional Odds Models for Ordinal Response Variables. *Applied Statistics* 39: 205-217.

23. Williams, R. (2005) Gologit2: A Program for Generalized Logistic Regression/ Partial Proportional Odds Models for Ordinal Variables. Retrieved May 12, 2005 <http://www.nd.edu/~rwilliam/stata/gologit2.pdf>. [12/07/2005]
24. More on the Proportional Odds/ Parallel Regressions Assumption. <http://www.nd.edu/~rwilliam/gologit2> [10/07/2005]
25. Maximum-Likelihood ordered logit estimation. <http://www.stata.com/help.cgi?ologit> [18/08/2005]
26. Omodel: Stata modules to perform test on ordered probit and ordered logit models. <http://ideas.repec.org/c/boc/bcode/s320901.html>. [21/07/2005]
27. Statistical Computing Seminars: Beyond Binary Logistic Regression in Stata. http://www.ats.ucla.edu/stat/stata/seminars/stata_BeyondBinaryLogistic/default.htm. [15/08/2005]
28. Gameroff, M. J. Using the Proportional Odds Model for Health- Related Outcomes: Why, When, and How with various SAS[®] Procedures. New York State Psychiatric Institute New York, NY <http://www2.sas.com/proceedings/sugi30/205-30.pdf>. [15/08/2005]
29. Applied Categorical and Nonnormal Data Analysis, Ordered Logit and Probit Models. <http://www.gseis.ucla.edu/courses/ed231c/notes.htm>. [14/07/2005]
30. Ordered/Ordinal Logistic Regression with SAS and STATA. <http://staff.washington.edu/~glynn/olr.pdf>. [11/06/2005]
31. Bender, R. Benner, A. (2000) Calculating Ordinal Regression Models in SAS and S-Plus. *Biometrical Journal* 42:6; 677-699.
32. Pulkstenis, E. Robinson, T. J. (2004) Goodness of Fit Tests for Ordinal Response Regression Models. *Statistics in Medicine* 23:999-1014.

33. Stiger, T. R. Barnhart, H. X. Williamson, J. M. (1999) Testing Proportionality in the Proportional Odds Model fitted with GEE. *Statistics in Medicine* 18:1419-1433.
34. Glonek, G. F. V. McCullagh, P. (1995) Multivariate Logistic Models. *Journal of the Royal Statistical Society. Series B* Volume 57, Issue 3, 533-546.
35. McCullagh, P. (1980) Regression Models for Ordinal Data. *Journal of the Royal Statistical Society. Series B* Volume 42, Issue 2, 109-142.
36. Liu, I. Agresti, A. (2005) The Analysis of Ordered Categorical Data: An Overview and a Survey of Recent Developments. *Sociedad de Estadística e Investigación Operativa Test* Volume 14, No: 1, 1-73.
37. Ananth, C. V. Kleinbaum, D. G. (1997) Regression Models for Ordinal Responses: A Review of Methods and Applications. *International Journal of Epidemiology* Volume 26, No:6, 1323-1333.
38. Harrell, F. E. (2001) Regression Modeling Strategies with Application to Linear Models, Logistic Regression and Survival Analysis. Springer-Verlag, New York.
39. Agresti, A. (1984) Analysis of Ordinal Categorical Data. Wiley Series in Probability and Mathematical Statistics, Wiley, New York.
40. Clogg, C. C. Shiehadeh, E. S. (1994) Statistical Models for Ordinal Variables. Sage Publications, Thousand Oaks, CA.
41. Powers, D. A. Xie, Y. Statistical Methods for Categorical Data Analysis. Academic Press, New York.
42. DeMaris, A. (1992) Logit Modeling Practical Applications. Sage Publications, Thousand Oaks, CA.
43. Aldrich, J. H. Nelson, F. D. (1984) Linear Probability, Logit and Probit Models. Sage Publications, Thousand Oaks, CA.