

**T.C.
İSTANBUL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI
SAYISAL YÖNTEMLER BİLİM DALI**

DOKTORA TEZİ

**SOSYAL BİLİMLERDE KALİTATİF BAĞIMLI DEĞİŞKENLERE
YAKLAŞIMDA İSTATİSTİK YÖNTEMLER**

**Serap ŞAHİN
2502050314**

**Tez Danışmanı
Prof. Dr. Neyran ORHUNBİLGE**

İstanbul, 2011

D O K T O R A

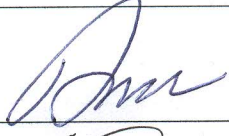


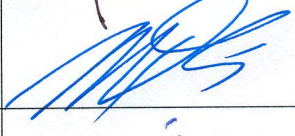
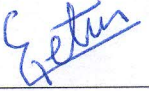
T E Z O N A Y I

ÖĞRENCİNİN

Adı ve Soyadı : SERAP ŞAHİN Numarası : 2502050314
Anabilim/Bilim Dalı : SAYISAL YÖNTEMLER Tez Savunma Tarihi : 31.01.2012
Danışman : PROF.DR.NEYRAN ORHUNBİLGE Tez Savunma Saati : 16.00

Tez Başlığı : SOSYAL BİLİMLERDE KALİTATİF BAĞIMLI DEĞİŞKENLERE
YAKLAŞIMDA İSTATİSTİK YÖNTEMLER

TEZ SAVUNMA SINAVI, Lisansüstü Öğretim Yönetmeliği'nin 35. Maddesi uyarınca yapılmış, sorulan sorulara alınan cevaplar sonunda adayın tezinin KABULÜ'NE OYBİRLİĞİ / ~~OYÇOKLUĞUYLA~~ karar verilmiştir.

JÜRİ ÜYESİ	İMZA	KANAATI (KABUL / RED / DÜZELTME)
1-PROF.DR.NEYRAN ORHUNBİLGE		Kabul
2- PROF.DR.ERHAN ÖZDEMİR		Kabul
3-PROF.DR.BESİM AKIN		Kabul
4-DOÇ.DR.MEHPARE TİMOR		Kabul
5-DOÇ.DR.EYÜP ÇETİN		Kabul.

ÖZ

Bir işletmenin sürekliliğini sağlaması ve büyümesi piyasadaki rekabet gücüne bağlıdır ve bu gücün ölçülmesi işletme performansının düzenli aralıklarla analiz edilmesi ile mümkündür. Bu nedenle, gelişmiş ve gelişmekte olan ülkelerdeki işletmeler üzerine performansı etkileyen faktörlerin belirlenmesine yönelik olarak birçok çalışma yapılmıştır. Bu faktörlerin saptanması işletmelerin finansal ve faaliyetlerine ilişkin hedeflerine ulaşmaları için kaynaklarını en etkin şekilde kullanmalarının bir ölçüsü olması bakımından önemlidir.

İşletmelerin ortak hedeflerinden olan karlılık, verimlilik, büyüme ve piyasa performans oranları performans değerlendirmesi amacıyla literatürde kullanılan başlıca göstergelerdir. İstatistik model kurularak gerçekleştirilen çalışmaların çoğunda bu kriterlerden sadece biri bağımlı değişken olarak seçilmiştir. Bu çalışmada ise söz konusu göstergelerden herhangi birinin tek başına işletme performansını belirlemede yeterli olmadığı hipoteziyle, birden çok göstergenin birlikte bağımlı değişkeni oluşturduğu mümkün tüm çoklu kombinasyonlarla alternatif modeller oluşturulmuştur.

Kurulan tüm modellerde açıklayıcı değişken olarak finansal oranlarla (likidite, finansal yapı ve faaliyet) finansal olmayan değişkenlerin (sermaye yoğunluğu, yabancı sermaye, yenilik, ihracat, pazar payı ve büyüklük) birlikte ele alınmasına karar verilmiştir. Bu aşamada işletmelerin performanslarını etkileyen faktörlerin ortak bir zeminde incelenebilmesi için performans göstergelerinden oluşan kombinasyonlar kategorik hale dönüştürülmüştür. Oluşturulan kategorik bağımlı değişken modellerinin analizinde çok değişkenli istatistik yöntemlerden Diskriminant ve Lojistik Regresyon Analizi kullanılmıştır.

ABSTRACT

The sustainability and the growth of a business depend on its competitiveness in the market. The measurement of this capacity is possible by regularly analyzing the firm's performance. Therefore, a significant amount of studies have been conducted on the firms in developed and emerging countries with an intention to determine the factors affecting the business performance. The determination of these factors is important for firms with respect to be a measure of using their resources in the most efficient way.

Profitability, productivity, growth, and market performance rates which are the common goals of businesses are the main indicators used in the literature for performance evaluation. In most of the studies done by building statistical models only one of these criteria is chosen as the dependent variable. In this dissertation on the other hand, alternative models with all possible multiple combinations are conducted in which more than one indicator generate the dependent variable under the hypothesis that one of the above mentioned indicators is not adequate to determine the performance of a firm.

In terms of the factors affecting the performance indicators, financial ratios (liquidity, financial structure, and operation) and nonfinancial variables (capital intensity, foreign capital, innovation, export, market share, and growth) are decided to be considered together in all models developed. At this stage, the combinations consisting of the performance indicators are transformed into being categorical for normalization. Discriminant and Logistic Regression Analysis which are multivariate statistic techniques are used for the analysis of the models formed with categorical dependent variables.

ÖNSÖZ

Bu çalışmada, Sosyal Bilimlerde sıklıkla karşılaşılan nitel bağımlı değişkenlere çözüm önerilerinde bulunmakla birlikte, özellikle işletmelerin performans göstergelerinden birkaçının birlikte ele alınarak performansı etkileyen açıklayıcı değişkenlerin en iyi şekilde belirlenmesi amaçlanmıştır.

Çalışmada öncelikle kategorik bağımlı değişkenleri modellemeye uygun olan çok değişkenli istatistik yöntemlerden Lojistik Regresyon ve Diskriminant Analizi sırası ile birinci ve ikinci bölümde teorik olarak incelenmiştir. Üçüncü bölümde işletme performans göstergeleri ve bu göstergeleri etkileyebilecek faktörler hakkında bilgilere yer verilmiştir. Son bölümde ise İstanbul Menkul Kıymetler Borsası'nda imalat sanayi sektöründe faaliyet gösteren işletmelerin performans göstergelerini etkileyen faktörlerin saptanması amacıyla kurulan modellerin analizleri yapıp sonuçları yorumlanmıştır.

Çalışmanın her aşamasında göstermiş olduğu destek ve yapmış olduğu katkılardan ötürü Danışman Hocam Prof. Dr. Neyran Orhunbilge'ye teşekkürü bir borç bilirim.

Ayrıca çalışmanın oluşması sırasında hiç bir yardımını esirgemeyen meslektaşım Dr. Seda Tolun'a, işletme finansı konusunda desteğini esirgemeyen meslektaşım Dr. Ebru Demirci'ye ve İ.Ü. İşletme Fakültesi Sayısal Yöntemler Anabilim Dalı kürsüsün'deki tüm hocalarıma ve arkadaşlarıma çok teşekkür ederim.

Son olarak benim için çok değerli olan, bugüne gelmemde çok emekleri ve destekleri olan aileme en içten teşekkürlerimi sunarım.

İÇİNDEKİLER

ÖZ.....	iii
ABSTRACT.....	iv
ÖNSÖZ.....	v
İÇİNDEKİLER.....	vi
TABLolar LİSTESİ.....	xi
ŞEKİLLER LİSTESİ.....	xiii
KISALTMALAR LİSTESİ.....	xiv
GİRİŞ.....	1
BİRİNCİ BÖLÜM: LOJİSTİK REGRESYON ANALİZİ.....	3
1.1 Genelleştirilmiş Doğrusal Modeller.....	4
1.1.1 Doğrusal Olasılık Modeli.....	11
1.1.2 Lojit Model.....	12
1.1.3 Probit Model.....	13
1.2 İki Kategorili Lojistik Regresyon Modelleri.....	14
1.2.1 Lojistik Regresyon Modellerine İlişkin Varsayımlar.....	17
1.2.2 Lojistik Regresyon Analizinde Model Kurma Stratejileri.....	18
1.2.3 Lojistik Modellerde Parametre Tahmin Yöntemi.....	20
1.2.4 Parametrelerin Anlamlılık Testleri.....	24
1.2.5 Uyum İyiliği Testleri.....	29

1.2.6	Model Parametrelerinin Yorumu	39
1.3	İki Kategorili Probit Analizi.....	43
1.3.1	Probit Modelinin Kurulması	43
1.3.2	Tahmin Parametrelerinin ve Modelin Uyum İyiliği Testleri	46
1.3.3	Model Parametrelerinin Yorumu	47
1.4	Çok Kategorili Lojistik Regresyon Modelleri	49
1.4.1	Çok Terimli Lojistik Regresyon Modelleri.....	49
1.4.1.1	Model Parametrelerinin Tahmin Yöntemleri.....	50
1.4.1.2	Çok Terimli Lojistik Regresyon Modellerine İlişkin Varsayım	52
1.4.1.3	Uyum İyiliği Testleri	57
1.4.1.4	Model Parametrelerinin Yorumu	58
1.4.2	Çok Terimli Probit Modeller.....	61
1.4.3	Sıralı Lojistik Regresyon Modelleri.....	61
1.4.3.1	Sıralı Lojistik Regresyon Analizinde Model Kurma Stratejileri	65
1.4.3.2	Model Parametrelerinin Tahmin Yöntemleri.....	67
1.4.3.3	Sıralı Lojistik Regresyon Modellerine İlişkin Varsayım	68
1.4.3.4	Uyum İyiliği Testleri	70
1.4.3.5	Model Parametrelerinin Yorumu	71
1.4.4	Sıralı Probit Modeller.....	75
1.4.5	Çok Kategorili Bağımlı Değişken için Diğer Modeller	78

1.5	Lojistik Regresyon Modellerinin Örnek Büyüklüğü.....	83
	İKİNCİ BÖLÜM: DİSKRİMİNANT ANALİZİ	87
2.1	İki Kategorili Diskriminant Analizi	88
2.1.1	Ayrıcı Değişkenlerin Seçimi.....	89
2.1.2	Diskriminant Analizinin Varsayımları.....	91
2.1.2.1	Çoklu Normallik Varsayımı	92
2.1.2.2	Eşit Varyanslılık Varsayımı.....	95
2.1.2.3	Çoklu Doğrusal Bağlantı Problemi Olmaması Varsayımı.....	96
2.1.2.4	Doğrusallık Varsayımı.....	98
2.1.3	Diskriminant Fonksiyonunun Elde Edilmesi	98
2.1.4	Diskriminant Fonksiyonunun Test Edilmesi ve Yorumlanması	102
2.1.5	Gözlemlerin Gruplarının Belirlenmesi.....	104
2.1.6	Diskriminant Analizinin Ayrıcı Gücünün Değerlendirilmesi.....	105
2.2	Çok Kategorili Diskriminant Analizi	106
2.2.1	Diskriminant Fonksiyonunun Elde Edilmesi	107
2.2.2	Diskriminant Fonksiyonunun Anlamlılık Testleri	107
2.2.3	Gözlemlerin Gruplarının Belirlenmesi.....	108
2.3	Diskriminant Fonksiyonunun Dışsal Geçerliliği	109
	ÜÇÜNCÜ BÖLÜM: İŞLETME PERFORMANSINI ETKİLEYEBİLECEK FAKTÖRLERİN ARAŞTIRILMASI	112
3.1	İşletme Performans Kriterleri.....	112

3.1.1	Karlılık Oranları	113	
3.1.2	Verimlilik	116	
3.1.3	Büyüme Oranları	117	
3.1.4	Piyasa Performans Oranları.....	121	
3.2	İşletme Performansını Etkileyebilecek Faktörler	122	
3.2.1	Finansal Oranlar	122	
3.2.1.1	Likidite Oranları	123	
3.2.1.2	Finansal Yapı Analizinde Kullanılan Oranlar	125	
3.2.1.3	Faaliyet Oranları	127	
3.2.2	Finansal Olmayan Değişkenler	131	
DÖRDÜNCÜ BÖLÜM: İŞLETME PERFORMANSININ TAHMİNİNE			
YÖNELİK İMALAT SANAYİ SEKTÖRÜ UYGULAMASI			137
4.1	Araştırmanın Amacı ve Konusu	137	
4.2	Araştırmanın Veri Seti.....	137	
4.3	Araştırma Değişkenlerinin Seçimi	138	
4.4	Verilerin Analize Hazırlanması	141	
4.5	Finansal Performans Tahmin Modellerinin Uygulanması	142	
4.5.1	Model 1: Performans Göstergesi Olarak Karlılık Verimlilik Büyüme .	142	
4.5.1.1	Diskriminant Analizi Sonuçları	143	
4.5.1.2	Lojistik Regresyon Analizi Sonuçları.....	146	
4.5.2	Model 2: Performans Göstergesi Olarak Karlılık ve Verimlilik	149	

4.5.2.1	Diskriminant Analizi Sonuçları	149
4.5.2.2	Lojistik Regresyon Analizi Sonuçları.....	151
4.5.3	Model 3: Performans Göstergesi Olarak Karlılık ve Büyüme	153
4.5.3.1	Diskriminant Analizi Sonuçları	153
4.5.3.2	Lojistik Regresyon Analizi Sonuçları.....	155
4.5.4	Model 4: Performans Göstergesi Olarak Karlılık ve Piyasa	158
4.5.4.1	Diskriminant Analizi Sonuçları	158
4.5.4.2	Lojistik Regresyon Analizi Sonuçları.....	159
4.5.5	Model 5: Performans Göstergesi Olarak Verimlilik ve Büyüme.....	161
4.5.5.1	Diskriminant Analizi Sonuçları	161
4.5.5.2	Lojistik Regresyon Analizi Sonuçları.....	163
SONUÇ		166
KAYNAKÇA		175
EKLER		194
ÖZGEÇMİŞ		212

TABLolar LİSTESİ

Tablo 1.1	Normalleştirilmiş Sabitler ve Varyans Fonksiyonları	11
Tablo 3.1	Yıllara göre İhracatın İthalatı Karşılama ve İhracat Değişim Oranları ...	120
Tablo 3.2	Yabancı Sermaye ve Ekonomik Büyüme	133
Tablo 4.1	Bağımsız Değişken Olarak Kullanılan Finansal Oranlar.....	140
Tablo 4.2	Bağımsız Değişken Olarak Kullanılan Finansal Olmayan Oranlar	141
Tablo 4.3	Model 1 için Diskriminant Analizi Testleri ve Sonuçları.....	143
Tablo 4.4	Model 1 için Diskriminant Modeli ve Doğru Sınıflandırma Sonuçları ...	145
Tablo 4.5	Model 1 için Test ve Doğru Sınıflandırma Sonuçları.....	146
Tablo 4.6	Model 1'deki Açıklayıcı Değişkenler ve Katsayıları	147
Tablo 4.7	Model 2 için Diskriminant Analizi Testleri ve Sonuçları.....	149
Tablo 4.8	Model 2 için Diskriminant Modeli ve Doğru Sınıflandırma Sonuçları ...	150
Tablo 4.9	Model 2 için Test ve Doğru Sınıflandırma Sonuçları.....	152
Tablo 4.10	Model 2'deki Açıklayıcı Değişkenler ve Katsayıları	152
Tablo 4.11	Model 3 için Diskriminant Analizi Testleri ve Sonuçları.....	154
Tablo 4.12	Model 3 için Diskriminant Modeli ve Doğru Sınıflandırma Sonuçları ..	155
Tablo 4.13	Model 3 için Test ve Doğru Sınıflandırma Sonuçları.....	156
Tablo 4.14	Model 3'deki Açıklayıcı Değişkenler ve Katsayıları	157
Tablo 4.15	Model 4 için Diskriminant Analizi Testleri ve Sonuçları.....	158
Tablo 4.16	Model 4 için Diskriminant Modeli ve Doğru Sınıflandırma Sonuçları ..	159

Tablo 4.17	Model 4 için Test ve Doğru Sınıflandırma Sonuçları.....	160
Tablo 4.18	Model 4'teki Açıklayıcı Değişkenler ve Katsayıları	160
Tablo 4.19	Model 5 için Diskriminant Analizi Testleri ve Sonuçları.....	162
Tablo 4.20	Model 5 için Diskriminant Modeli ve Doğru Sınıflandırma Sonuçları ..	163
Tablo 4.21	Model 5 için Test ve Doğru Sınıflandırma Sonuçları.....	164
Tablo 4.22	Model 5'deki Açıklayıcı Değişkenler ve Katsayıları	164

ŞEKİLLER LİSTESİ

Şekil 1.1	Modeldeki “sabit” in olasılık üzerindeki etkisi.....	39
Şekil 1.2	Modeldeki “eğim” katsayısının olasılık eğrisi üzerindeki etkisi.	40
Şekil 4.1	Model 1 için Serpilme Diyagramı	144
Şekil 4.2	Model 1 için ROC Eğrisi	148
Şekil 4.3	Model 2 için Serpilme Diyagramı	150
Şekil 4.4	Model 2 için ROC Eğrisi	153
Şekil 4.5	Model 3 için Serpilme Diyagramı	154
Şekil 4.6	Model 3 için ROC Eğrisi	157
Şekil 4.7	Model 4 için Serpilme Diyagramı	158
Şekil 4.8	Model 4 için ROC Eğrisi	161
Şekil 4.9	Model 5 için Serpilme Diyagramı	162
Şekil 4.10	Model 5 için ROC Eğrisi	165

KISALTMALAR LİSTESİ

AIC	Akaike Information Criteria – Akaike Bilgi Kriteri
AR-GE	Araştırma Geliştirme
BIC	Bayesian Information Criteria – Bayes Bilgi Kriteri
CI	Conditional Index – Koşul İndeksi
D	Deviance - Sapma
DD	Defter Değeri
EKK	En Küçük Kareler
GLM	Generalized Linear Models – Genelleştirilmiş Doğrusal Modeller
IIA	Independence of Irrelavent Alternatives – İlgisiz Alternatiflerin Bağımsızlığı
İMKB	İstanbul Menkul Kıymetler Borsası
İSO	İstanbul Sanayi Odası
K.Vd.	Kısa Vadeli
LL	Log-Likelihood – Log-olabilirlik
LR	Likelihood Ratio – Olabilirlik Oranı
LRM	Linear Regression Model – Doğrusal Regresyon Modeli
MLE	Maksimum Likelihood Estimation – En çok Olabilirlik Tahmini
MNL	Multinomial Logit – Çok Kategorili Lojit
MTT test	McFadden, Train ve Tye tarafından sunulan test

OR	Odds Ratio – Odds Oranı
PD	Piyasa Deęeri
ROC Eęrisi	Receiver Operating Characteristic Curves
SH testi	Small ve Hsiao testi
U.Vd.	Uzun Vadeli
VIF	Variance Inflation Factor – Varyans Artıř Faktörü

GİRİŞ

İşletmelerin rekabet güçlerine bağlı olarak performanslarının sürekliliğinin sağlanmasının önemi özellikle yatırımcılar, ortaklar ve işletme yöneticileri açısından her geçen gün artmaktadır. Bu nedenle hem yatırımcı ve ortakların kararlarında, hem de ülke ekonomilerinin verimliliğinin ve etkinliğinin belirlenmesinde işletme performanslarını etkileyen faktörlerin en iyi şekilde belirlenmesi amacıyla araştırmacılar tarafından birçok çalışma yapılmıştır. Bu konu ile ilgili yapılan çalışmalarda performans göstergelerinin ayrı ayrı inelenmesine karşın işletmelerin performanslarının birden çok ve farklı kriterler ile ölçülmesi ve değerlendirmesinin daha uygun olacağı düşünülmektedir.

Çok değişkenli istatistik yöntemlerden olan Lojistik Regresyon ile Diskriminant Analizi'nin ele alındığı ve bu yöntemler sonucunda işletme performansını etkileyen faktörlerin belirlenmesi amacıyla kurulan modellerin incelendiği bu çalışma dört bölümden oluşmaktadır.

Çalışmada öncelikle işletmelerin performanslarını çok boyutlu değerlendirebilmek için performans göstergelerinin farklı kombinasyonlarından meydana gelen çeşitli göstergelere göre sınıflandırılmış olan kategorik bağımlı değişkenleri modellemeye en uygun olan çok değişkenli istatistik yöntemlerden Lojistik Regresyon ve Diskriminant Analizi teorik olarak incelenmiştir.

Birinci bölümde, bağımlı değişkenin nitel olduğu durumlarda kullanılan modeller hakkında genel bilgi verilmesi açısından öncelikle Genelleştirilmiş Doğrusal Modeller başlığı altında Doğrusal Olasılık Modeli, Lojit ve Probit Modelleri'ne kısaca değinilmiştir. Çalışmanın uygulama bölümünde de kullanılan Lojistik Regresyon Analizi'nin iki ve çok kategorili olduğu durumlarından ise ayrıntılı olarak bahsedilmiştir.

İkinci bölümde, bağımlı değişkenin nitel, bağımsız değişkenlerin genellikle nicel ve sayıca fazla olmamak koşuluyla nitel de olduğu durumlar için kullanılabilen çok

değişkenli istatistik yöntemlerden Diskriminant Analizi, Lojistik Regresyon Analizi'nde olduğu gibi yine hem iki hem de çok kategorili olarak incelenmiştir.

İşletmelerin performans göstergeleri ve bu göstergeleri etkileyebilecek faktörler hakkında bilgilere ise üçüncü bölümde yer verilmiştir. Son bölümde ise tezin ampirik çalışması yer almaktadır. İstanbul Menkul Kıymetler Borsası'nda imalat sanayi sektöründe faaliyet gösteren işletmeler performansları açısından çeşitli göstergelere göre sınıflandırıldıktan sonra hem Lojistik Regresyon hem de Diskriminant Analizi ile söz konusu göstergelerin hangi faktörlere bağımlı olduğu araştırılmıştır.

BİRİNCİ BÖLÜM: LOJİSTİK REGRESYON ANALİZİ

Çeşitli bilim dallarında bağımlı değişkenin nicel (nümerik) olduğu durumlarda kullanılan en temel istatistik yöntem regresyon analizidir. Ancak bağımlı değişkenin nitel olduğu durumlarda klasik regresyon modeli kullanılamamaktadır. Bunun nedeni, sosyal bilimlerdeki araştırmaların birçoğunda tutumların (tavırların), davranışların, karakteristik özelliklerin, kararların ve olayların, özünde sürekli ya da süreksiz olması, bir başka ifadeyle; kesikli, isimsel, sıralı ve sürekli olmayan yollarla ölçülebilir olmasıdır. Bu tür değişkenlerin analizi için birçok uygun istatistik model söz konusudur. Bu modellerden bazıları, “İki Kategorili Veri Analizi (Binary Data Analysis)”, “Sıralı Veri Analizi (Ordinal Data Analysis)”, Kategorik Veri Analizi (Categorical Data Analysis)”, “Kesikli Tercih Analizi (Discrete Choice Analysis)” ya da Lojit veya Probit gibi özel modellerdir. Birbiriyle ilişkili bu istatistik yöntemlerin ortak özelliği, olayların olasılıklarını yani olayın ne kadar sıklıkla ortaya çıktığını modellemeleridir. Bu nedenle, olay olasılığını analiz eden tüm istatistik yöntemlere genel olarak olasılık modelleri adı verilmektedir¹. Çalışmada olasılık modellerinden, İki Kategorili Lojit (Binary Logit), Çok Kategorili Lojit (Multinomial Logit), Sıralı Lojit (Ordinal Logit) ve Probit ve Diskriminant modelleri ele alınacaktır.

Lojistik Regresyon Analizi, bağımlı değişkenlerinin* kategorik olarak ikili, üçlü ve çoklu kategorilerde gözlemlendiği durumlarda, açıklayıcı değişkenlerin (bağımsız değişken) bağımlı değişken ile sebep sonuç ilişkisini belirlemede yararlanılan bir yöntemdir. Ayrıca, açıklayıcı değişkenlere göre bağımlı değişkeninin beklenen değerinin olasılık olarak elde edildiği, sınıflama ve atama işlemini yapmaya yardımcı olan bir regresyon modelidir. Lojistik Regresyon Analizinin en önemli amacı, istatistikteki diğer model kurma yöntemlerinde olduğu gibi bağımsız değişkenlerle

¹Tim Futing Liao, **Interpreting Probability Models Logit Probit and Other Generalized Linear Models**, Sage University Paper, 1994, pp. 1–2.

* Doğrusal Regresyon Modeli’nde “bağımlı değişken” (dependent variable) olarak ifade edilen kavram Lojistik Regresyon Modeli içi literatürde “cevap değişkeni” (response variable) veya “sonuç değişkeni” (outcome variable) olarak ifade edilmektedir.

bağımlı deęişken arasındaki ilişkinin tanımlanmasında en uygun modelin belirlenmesidir².

Bu bölümde Lojistik Regresyon Analizine başlamadan önce Genelleştirilmiş Doğrusal Modeller başlığı altında Doğrusal Olasılık Modeli, Lojit ve Probit modelleri hakkında genel bilgi verilecektir. Bu modeller bağımlı deęişkenin nitel olduęu durumlarda kullanılan modellerdir.

1.1 Genelleştirilmiş Doğrusal Modeller

Regresyon ve Varyans Analizi'nde kullanılan modeller, "Genelleştirilmiş Doğrusal Modeller" (Generalized Linear Models – GLM) olarak bilinen geniş bir model grubunun üyeleridir. Genelleştirilmiş Doğrusal Modeller 1972 yılında ilk olarak Nelder ve Wedderburn tarafından önerilmiş ve 1989 yılında McCullagh ve Nelder tarafından daha kapsamlı şekilde ele alınmıştır. Lojistik regresyon ise GLM'in özel bir durumu olarak kabul edilmektedir³.

GLM'in temel yaklaşımı, fonksiyonları doğrusal modelin sistematik bileşenleri olarak ifade edecek şekilde dönüştürerek Standart Doğrusal Modele oldukça benzeyen modeller elde etmektir⁴. Farkı ise doğrusal ve aralık ölçekli olmayan bağımlı deęişkenlerin açıklayıcı deęişkenlerle ilişkilerini göstermek bakımından uygulanabilirliğidir.

GLM; verilerin sistematik bileşenleri, eşit varyanslılık ve asimptotik normallik varsayımlarını gerektirmeyen, bağımlı deęişkenle açıklayıcı deęişkenler arasındaki ilişkiyi tanımlamak için "bağ fonksiyonunun (link function)" kullanıldığı, gözlemlerin birbirinden bağımsız olması varsayımının ise sağlanması gereken modellerdir⁵.

² David W. Hosmer, Stanley Lemeshow, **Applied Logistic Regression**, 2nd Edition, USA, John Wiley & Sons, 2000, p.1.

³ Ronald Christensen, **Log-Linear Models and Logistic Regression**, 2nd Edition, Springer, 1997, p. 297.

⁴ Alan Agresti, **An Introduction to Categorical Data Analysis**, John Wiley & Sons, 1996, p. 72.

⁵ Jeff Gill, **Generalized Linear Models**, USA Sage University Paper, 2001, p. 2.

GLM hakkında tartışmaya, üstel dağılım grubuyla başlanır. Bunun nedeni, GLM'in yalnızca bu kategoriye uyan parametrik formlara uygulanabilirliğidir. Bernoulli, Binom, Geometrik, Negatif Binom, Poisson, Üstel, Normal, Beta ve Gamma dağılımlarının bazıları üstel gruba aittir⁶. Burada bahsi geçen Gamma dağılımının özel iki durumu olan, Gammanın $(\rho/2, \frac{1}{2})$ parametreleri ile ρ serbestlik dereceli Ki-kare dağılımı ve Gammanın $(1, \beta)$ parametreleriyle üstel dağılıma uyan formudur⁷. İki önemli olasılık yoğunluk fonksiyonu olan Student-t ve Uniform dağılım ise üstel dağılım grubunun üyesi değildir.⁸

GLM teorisi, değişkenin dağılımının genellikle tek parametreliliğiyle üstel gruptan olmasını gerektirmektedir. Bu modeller içinde y rastgele ya da tahmini kısımdır. Bu öge bazen “olasılık dağılımı” ya da “bağımlı değişken dağılımı” olarak adlandırılmakla birlikte bazı kaynaklarda “rassal öge”⁹ olarak adlandırılmaktadır. Bağımlı değişkenin olasılık dağılımı üstel dağılım grubundan olmakta ve genel ifade şekli ile aşağıdaki gibi yazılmaktadır.

$$f(y|\theta, \phi; w) = \exp\left\{\frac{w}{\phi}[y\theta - r(\theta)]\right\} h(\phi, y, w) \quad 1.1$$

θ , ϕ ve w sayısal değerleri ifade etmektedir. Varsayıma göre w , sabit bilinen bir değerdir. Bu fonksiyon içinde ϕ farklı bir rol üstlenmektedir. Bilinmeyen sabit gibi davranılmakla birlikte ϕ “sorunlu parametre (nuisance parameter)” veya “dağılım parametresi (dispersion parameter)” olarak adlandırılmaktadır. Aslında ϕ , çeşitli özel durumların tanımlanması için uygun pozitif bir değerdir.¹⁰ Bazı kaynaklarda $w/\phi = \alpha(\phi)$ ölçek faktörü olarak ifade edilmektedir. Normal dağılım ve Gamma dağılımı için durum biraz farklıdır. Bu dağılımların sırası ile $\theta = (\mu, \sigma^2)$ ve $\theta = (\alpha, \beta)$ olmak üzere iki parametresi bulunmaktadır. Bu durumda, Normal dağılım için

⁶ Ronald Christensen, **a.g.e.**, pp. 299–304.

⁷ John E. Freund, **Mathematical Statistics**, Fifth Edition, Prentice Hall International Editions, p. 224

⁸ Jeff Gill, **a.g.e.**, p.10.

⁹ Joseph F. Hair, William C. Black, Barry J. Babin, Rolph E. Anderson and Ronald L. Tatham, **Multivariate Data Analysis**, Sixth Edition, Pearson, 2006, p. 411.

¹⁰ Ronald Christensen, **a.g.e.**, p. 300.

ϕ parametresi normal bir dağılımın varyansı (σ^2) ya da Gamma dağılımının p parametresi gibi özel bir tip sorunlu parametreye işaret edebilir¹¹.

Tek parametrelili üstel dağılım grubu için yukarıdaki eşitliğin en genel ifadesi $w/\phi = 1$ olacak şekilde,

$$f(y|\theta) = \exp[y\theta - r(\theta) + h(y)] \quad 1.2$$

biçiminde gösterilmektedir. Eşitlik 1.2’de $h(y)$ ile $\log[h(\phi, y, w)]$ ifade edilmektedir. Eşitliğin sağ tarafındaki y ve θ ’nin çarpımından oluşan ilk kısım etkileşim bileşeni, diğer kısım ise katkı bileşeni olarak adlandırılmaktadır. Üstel dağılım grubu ile kanonik formun* kullanılması verinin dağılımı hakkında doğrudan bilgi vermesi bakımından çalışmayı daha da kolaylaştırmaktadır¹².

1989 yılında McCullagh ve Nelder, katkı bileşeni olarak adlandırılan ikinci kısım içindeki $r(\theta)$ ’yi “kümülatif fonksiyon (cumulant function)”¹³ olarak adlandırmıştır. Bu değer ile ilgili en önemli olgu, $r(\theta)$ ’nin dağılım momentlerinin hesaplanmasında anahtar rol oynamasıdır, $r(\theta)$ ’nin θ ’ya göre birinci mertebeden türevi bağımlı değişkenin beklenen değerini ikinci mertebeden türevi de varyansını vermektedir. Ayrıca, θ ’nin bu formu, orijinal formu ile parametre değeri olarak ifade edilen formu arasında kanonik bir bağ oluşturmaktadır¹⁴. Başka bir ifade ile modellerde, bağımlı değişken açıklayıcı değişkenlere doğrusal bir yapı ile bağlıdır. Bu yapıya da *bağ fonksiyonu* adını almaktadır.

GLM’in gelişiminde, beklenen değer (ilk moment) teorik olarak önemli rol oynamaktadır. Eşitlik 1.2’deki dağılımlar için y değişkeninin beklenen değeri θ ’ya

¹¹ J. A. Nelder, R. W. M. Wedderburn, “Generalized Linear Models”, **Journal of the Royal Statistical Society Series A**, Vol:135, No:3, 1972, pp. 370–384.

* “Kanonik form” ile kastedilen; matematikte ilgilenilen olgunun belirli bir şekle göre gösterilmesidir. Başka bir ifadeyle ele alınan olgu için standart, genel kabul görmüş formlara kanonik form denilmektedir. Bu açıklamanın ışığında bu çalışmada kanonik form dağılımı, Eşitlik 1.2’de gösterilen ifadedir.

¹² Jeff Gill, **a.g.e.**, p. 11.

¹³ P. McCullagh, J. A. Nelder, **Generalized Linear Models**, 2nd Edition, Chapman & Hall, 1989, p. 30.

¹⁴ Jeff Gill, **a.g.e.**, p. 12.

bağlı ancak ϕ 'ye bağlı değildir. Herhangi bir fonksiyon için aşağıdaki eşitlik geçerlidir.¹⁵

$$1 = \int f(y|\theta, \phi; w) dy \quad 1.3$$

y değişkeninin kesikli olduğu durumda integral yerine toplam (sigma) kullanılır.

Her iki tarafın θ 'ya göre türevi alındığında;

$$0 = \int f'(y|\theta, \phi; w) dy. \quad 1.4$$

f' ; f 'nin θ 'ya göre birinci mertebeden türevidir ve Eşitlik 1.1'deki fonksiyon Eşitlik 1.4'de yerine konulduğunda Eşitlik 1.4'ün aşağıdaki şekle dönüştüğü görülmektedir.

$$0 = \frac{w}{\phi} \int (y - r'(\theta)) f(y|\theta, \phi; w) dy \quad 1.5$$

Buradan basit matematik işlemler yapıldığında y 'nin beklenen değeri,

$$E(y) \equiv \mu = \frac{\partial}{\partial \theta} r(\theta) \quad 1.6$$

olarak elde edilmektedir¹⁶. r' ; r 'nin θ 'ya göre birinci mertebeden türevini göstermek üzere, $r'(\theta)$ tersi alınabilen fonksiyondur. Bu nedenle θ ayrıca beklenen değer de bir fonksiyonudur, $\theta = (r')^{-1}(\mu)$. Eşitlik 1.1'deki formun dağılımlar için doğrusal yapısı genellikle, $\theta = x'\beta$ tarafından β bilinmeyen parametreler vektörü ve x bilinen ve sabit değer olarak en doğal hali ile tanımlanmaktadır. Aslında, GLM'in analizi, doğrusal yapının beklenen değerinin daha genel bir fonksiyonu ile gösterildiğinde başarılı sonuca ulaşılabilir.

Genelleştirilmiş Doğrusal Model $i = 1, 2, \dots, n$ olacak şekilde birbirinden bağımsız gözlemlerden (y_i) meydana geliyorsa,

¹⁵ Jeff Gill, **a.g.e.**, p. 300.

¹⁶ Nicholas T. Longford, **Handbook of Statistical Modeling for the Social and Behavioral Sciences**, Plenum Press, Edited by Gerhard Arminger, Clifford C. Clogg and Michael E. Sobel, 1995, p. 551.

$$y_i \sim f(y_i | \theta_i, \phi, w_i)$$

$$E(y_i) \equiv \mu_i$$

$$g(\mu_i) = x_i' \beta$$

olarak ifade edilmektedir. Eğer $g(\mu_i) = \theta_i$ ise model, Kanonik (Standart) Genelleştirilmiş Doğrusal Modeldir. Başka bir deyişle kanonik bir model $g(.) \equiv (r')^{-1}(.)$ demektir¹⁷.

GLM'in çeşitli bileşenleri için bazı özel ifadeler kullanılmaktadır. GLM içinde açıklayıcı değişkenlerin tanımlandığı kısım "sistemik öge"¹⁸ olarak adlandırılmakla birlikte doğrusal yapı $x' \beta$, *doğrusal tahminleyen*¹⁹ olarak adlandırılan $g(.)$ fonksiyonu ile ortalama (beklenen değer) ve doğrusal tahminleyen arasındaki ilişkiyi tanımlayan bir başka öge de $g(\mu) = x' \beta$, *bağ fonksiyonu*²⁰ olarak bilinmektedir. Eğer $g(\mu) = \theta$ ise $g(.)$ fonksiyonu *kanonik (standart) bağ fonksiyonu* olarak anılır²¹.

Dağılım parametresi tahmini ve modelin uygunluk testleri için tek parametrelili üstel dağılım varyansının hesaplanması gereklidir. Eşitlik 1.1'in θ 'ya göre ikinci mertebeden türevi alındığında, ($0 = \int f''(y|\theta, \phi; w) dy$);

$$0 = \frac{w}{\phi} \int \frac{d[(y - r'(\theta))f(y|\theta, \phi; w)]}{d\theta} dy \quad 1.7$$

elde edilmekte ve gerekli işlemler yapıldığında Eşitlik 1.8'e ulaşılmaktadır.

$$0 = \frac{w^2}{\phi^2} \int (y - r'(\theta))^2 f(y|\theta, \phi; w) dy + \frac{w}{\phi} \int -r''(\theta) f(y|\theta, \phi; w) dy \quad 1.8$$

(Eşitliklerdeki "''" ikinci mertebeden türev için kullanılmaktadır.) Yukarıdaki eşitlik (1.8),

¹⁷ P. McCullagh, J. A. Nelder, **a.g.e.**, pp. 28–29.

¹⁸ **A.e.**, p. 27.

¹⁹ Alvin C. Rencher, G. Bruce Schaalje, **Linear Models in Statistics**, A John Wiley & Sons, 2008, p. 513.

²⁰ Alan Agresti, **a.g.e.**, p. 72.

²¹ Ronald Christensen, **a.g.e.**, p. 301.

$$0 = [Var(y)w^2/\phi^2] - [r''(\theta)w/\phi] \quad 1.9$$

şeklini almaktadır ve $Var(y) = r''(\theta)\phi/w$.²²

Bazı yaygın olasılık fonksiyonlarının normalleştirilmiş katsayıları ve varyans fonksiyonları Tablo 1.1'de verilmekle birlikte çalışmada incelenen iki kategorili bağımlı değişkenlerin dağılım fonksiyonu olan Binom dağılımın üstel dağılım formu, beklenen değeri ve varyansı aşağıdaki gibi elde edilmektedir.

Binom dağılım, Bernoulli sürecinin özel bir hali olup, denemelerde birbirini engelleyen iki sonuçtan birinin gerçekleşmesiyle sonuçlanan sürecin çok kez yinelenmesiyle ilgilidir²³. Bu sürecin dağılımının önemli bir genellemesi olan rassal bir denemenin birbirinden bağımsız olarak n kez yinelenmesi sonucunda ortaya çıkan başarı sayısı ile ilgilenen Binom dağılımın olasılık kütle fonksiyonu, üstel dağılım formunda şu şekilde ifade edilmektedir²⁴;

$$\begin{aligned} f(y|n, p) &= \binom{n}{y} p^y (1-p)^{n-y} \\ &= \exp \left[\log \binom{n}{y} + y \log(p) + (n-y) \log(1-p) \right] \quad 1.10 \\ &= \exp \left[y \log \left(\frac{p}{1-p} \right) - (-n \log(1-p)) + \log \binom{n}{y} \right]. \end{aligned}$$

Binom dağılım için kanonik bağ $\theta = \log(p/(1-p))$ ile Lojit modeli oluşturmaktadır. Bu bağ fonksiyonu Lojit model başlığı altında incelenmiştir. Kanonik bağ fonksiyonunun tersi alınarak, $r(\theta)$ 'da yerine konulursa aşağıdaki eşitlik bulunmaktadır.

$$r(\theta) = [-n \log(1-p)] = n \log(1 + \exp(\theta)) \quad 1.11$$

Bilindiği gibi Binom dağılımın p ve n olmak üzere iki parametresi bulunmaktadır. Bunlardan p bilindiğinde ve üstel olasılık fonksiyonu n parametresi için geliştirilmek

²² P. McCullagh, J. A. Nelder, **a.g.e.**, p. 29.

²³ Neyran Orhunbilge, **Tanımsal İstatistik Olasılık ve Olasılık Dağılımları**, Avcıol Basım, 2000, s. 192.

²⁴ Alvin C. Rencher, G. Bruce Schaalje, **a.g.e.**, p. 514.

istendiğinde, üstel olasılık fonksiyonu olarak yazılamamaktadır²⁵.

Binom olasılık fonksiyonunun ortalaması $r(\theta) = n \log(1 + \exp(\theta))$ ve kanonik bağ fonksiyonu $\theta = \log(p/(1 - p))$ olarak kullanılırsa,²⁶

$$\begin{aligned}
 \frac{\partial}{\partial \theta} r(\theta) &= \frac{\partial}{\partial \theta} (n \log(1 + \exp(\theta))) \\
 &= n \log(1 + \exp(\theta))^{-1} \exp(\theta) \\
 &= n \left(1 + \exp \left(\log \left(\frac{p}{1-p} \right) \right) \right)^{-1} \exp \left(\log \left(\frac{p}{1-p} \right) \right) \\
 &= n(1 - p) \left(\frac{p}{1-p} \right)
 \end{aligned} \tag{1.12}$$

elde edilir. Buradan beklenen değer $E[y] = np$ olarak hesaplanmaktadır.

Binom olasılık kütle fonksiyonunun varyansı,²⁷

$$\begin{aligned}
 \text{VAR}[Y] &= \frac{\phi}{w} \frac{\partial^2}{\partial \theta^2} r(\theta) \\
 &= 1 \frac{\partial^2}{\partial \theta^2} (n \log(1 + \exp(\theta))) \\
 &= \frac{\partial}{\partial \theta} \left(n(1 + \exp(\theta))^{-1} \exp(\theta) \right) \\
 &= n \exp(\theta) \left[(1 + \exp(\theta))^{-1} - (1 + \exp(\theta))^{-2} \exp(\theta) \right] \Big|_{\theta = \log(p/(1-p))} \\
 &= n \left(\frac{p}{1-p} \right) \left[\left(1 + \frac{p}{1-p} \right)^{-1} - \left(1 + \frac{p}{1-p} \right)^{-2} \frac{p}{1-p} \right] \\
 &= np(1 - p)
 \end{aligned} \tag{1.13}$$

olarak gösterilmektedir.

Binom olasılık yoğunluk fonksiyonu için kanonik bağ iki şekilde ifade edilebilmektedir. Bu bağ fonksiyonlardan biri olan Lojit, üstel grup formunun tanımı gereği doğal olarak yer almaktadır. Diğer bağ fonksiyonu Probit ise Birikimli Standart Normal dağılıma dayanmaktadır. Genelde sosyal bilimler verisi için bu iki bağ fonksiyonu da (lojit ve probit) kullanılabilir ve bu bağ fonksiyonları benzer

²⁵ Jeff Gill, **a.g.e.**, p. 15.

²⁶ **A.e.**, p. 24.

²⁷ **A.e.**, p. 27.

şekilde sonuçlar vermektedir.²⁸ Bu modellerin sonuçlarının benzerliği ile ilgili ayrıntılı bilgi Başlık 1.1.3 içinde açıklanmaktadır.

Son olarak üstel dağılım grubunun üyesi olan dağılımlardan bazılarının normalleştirilmiş sabitleri ve varyans fonksiyonları aşağıdaki tablo ile özetlemektedir.²⁹

Tablo 1.1 Normalleştirilmiş Sabitler ve Varyans Fonksiyonları

Dağılım	$r(\theta)$	$\tau^2 = \frac{\partial^2}{\partial \theta^2} r(\theta)$
Poison	$\exp(\theta)$	$\exp(\theta)$
Binom	$n \log(1 + \exp(\theta))$	$n \exp(1 + \exp(\theta))^{-2}$
Normal	$\frac{\theta^2}{2}$	1
Gamma	$-\log(-\theta)$	$\frac{1}{\theta^2}$
Negatif Binom	$s \log(1 - \exp(\theta))$	$s \exp(\theta)(1 - \exp(\theta))^{-2}$

1.1.1 Doğrusal Olasılık Modeli

Doğrusal Olasılık Modeli bağımlı değişkenin iki kategorili olduğu durumlarda uygulanan regresyon modelidir. Yapısal modelin gösterimi aşağıdaki gibidir.

$$y_i = x_{ik}\beta_k + \varepsilon_i \quad k = 1, 2, \dots, p \quad \mathbf{1.14}$$

x_i , i . gözlemin değerlerinden oluşan vektör; β parametre vektörü ε ise hata terimidir. y bağımlı değişkeni eğer olay gerçekleşiyorsa 1, gerçekleşmiyorsa 0 değerini almaktadır. Doğrusal Olasılık Modelini anlamak için öncelikle, $E(y|x)$ 'in ne anlama geldiğini kavramak gerekmektedir. Olayın gerçekleşme olasılığı $E(y_i) = P(y_i = 1)$ 'dir. Koşullu değer olarak bahsedilen ise,

$$E(y_i|x_{ik}) = P(y_i = 1|x_{ik}) = x_{ik}\beta_k \text{ 'dir.} \quad \mathbf{1.15}$$

²⁸ Jeff Gill, **a.g.e.**, pp. 32–33.

²⁹ **A.e.**, p.29.

Buradan Doğrusal Olasılık Modelinin parametreleri şu şekilde yorumlanmaktadır: Model doğrusal olduğu sürece, x_k 'daki bir birimlik değişim, diğer açıklayıcı değişkenlerin sabit tutulduğu varsayıldığında, ilgilenilen durumun ortaya çıkma olasılığında β_k kadar beklenen değişime neden olmaktadır. Olasılıkları doğrusal olan modele *Doğrusal Olasılık Modeli* adı verilmektedir³⁰.

Doğrusal Olasılık Modelinde bağımlı değişkeninin iki kategorili olması durumunda parametre katsayılarının yorumu değişmemekle birlikte regresyon varsayımlarından bazılarını sağlaması gerekmemektedir. Bunlar; eşvaryanslılık, normallik ve fonksiyon formundaki bozukluklardır. Ayrıca Doğrusal Olasılık Modellerine alternatif olarak nitel bağımlı değişken için Lojit ve Probit modeller de vardır.

1.1.2 Lojit Model

Genel Log-Doğrusal Modelin özel bir formu olan Lojit Model, nitel bağımlı değişken analizlerini birleştiren önemli bir modeldir. “Lojit” in kelime anlamı kendi içinde saklı olup, araştırmada ilgilenilen durumun ortaya çıkma olasılığının çıkmama olasılığına oranını ifade eden odds'un doğal logaritmasının alınması anlamına gelmektedir. Literatürde “log-odds” olarak da yer almaktadır. Lojit Modelin genel gösterimi

$$\log \left[\frac{P(y = 1)}{1 - P(y = 1)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad 1.16$$

şeklinindedir. Lojit dönüşüm, bağımsız değişkenler ile orjinal olasılıklarının arasındaki doğrusal olmayan ilişkileri doğrusallaştırmaktadır. Lojit Model, hatalarının ortalaması “0” ve varyansı “ $\pi^2/3$ ” olan standart lojistik dağılıma uymaktadır³¹. Bu nedenle Lojit Modeller veriye uygulanırken bağ fonksiyonu Lojit olur, $\theta = \log(p/(1 - p))$. Başka bir ifadeyle, bağımlı değişkenin lojit fonksiyonunun açıklayıcı değişkenlerin doğrusal bir bağlantısı olduğunu söyleyen bağ fonksiyonu Lojitdir.

³⁰ J. Scott Long, **Regression Models for Categorical and Limited Dependent Variables**, USA, Sage Publication, 1997, p. 36.

³¹ A.e., p. 42.

Lojistik Regresyon Analizi için kullanılan lojit dönüşümü göreceli basitliğinin avantajı nedeniyle en yaygın kullanılan modeldir ve Lojit Model olarak adlandırılmaktadır. Lojit Modeller, İki Kategorili Lojistik Regresyon Analizi bölümü içinde daha ayrıntılı olarak ele alınacaktır.

1.1.3 Probit Model

Bağımlı değişkenin iki kategorili olduğu ve Binom dağılıma uyduğu durumlarda sık kullanılan bir diğer model Probit Modeldir. Probit Modelden elde edilen dönüşüm, Probit Analizi için istenen olayın olasılıklarını, Lojistik dağılımda Lojit yerine Birikimli Normal dağılımdaki skora dönüştürmektedir³². Probit Modellerde GLM'deki bağımlı değişkeni açıklayıcı değişkenlere doğrusal bir yapı ile bağlayan fonksiyon "Probit"tir, $\theta = \Phi^{-1}(p)$. Olasılıklar için Probit Modeller,

$$P(y = 1) = 1 - \Phi\left(-\sum_{k=1}^p \beta'_k x_k\right) = \Phi\left(\sum_{k=1}^p \beta'_k x_k\right) \quad 1.17$$

şeklinde gösterilmektedir. Lojit Model odds'ların logaritmaları alınarak yapılan dönüşüm ve Lojistik Regresyon olarak ifade edilen, olayın olasılığını modellemek üzere iki temel şekilde ifade edilebilirken; Probit Model, ilgilenilen olayın olasılıklarının standart normal dağılım tablosu kullanılarak elde edilen Z değerlerinin bir Doğrusal Regresyonu olan θ ile ifade edilebilecek şekilde göstermektedir. Z değerleri ile olasılıklar arasında doğrusal olmayan ilişki söz konusudur. Olasılık değerleri uç değerlere (0 veya 1'e) yaklaştıkça Z değerlerindeki bir birimlik değişim, fonksiyonda daha az bir değişime neden olmaktadır³³.

Lojit Analizde modelin hatasının standart sapması $\sqrt{\pi^2/3}$ 'den yaklaşık olarak "1.814" iken Probit Analizde modelin hatasının standart sapması "1" dir. İki modele ait hata varyanslarının farklı olması Probit ve Lojit modellerin katsayılarının doğrudan karşılaştırılmayacağı anlamına gelmektedir. Lojit katsayıları, Probit

³² Fred C. Pampel, **Logistic Regression A Primer**, USA, Sage Publication, 2000, pp. 54–55.

³³ Tim Futing Liao, **a.g.e.**, p.22.

katsayılarının yaklaşık olarak 1.8 katı olmaktadır.³⁴ Lojistik regresyon ve Probit katsayıları arasında fark modellerin hatalarının dağılımları olan Birikimli Lojistik ve Normal dağılımdan kaynaklanmaktadır.³⁵

1.2 İki Kategorili Lojistik Regresyon Modelleri

İki Kategorili Lojistik Regresyon Modellerinde, bağımlı değişkenler *nominal* veya *kategorik* olarak da adlandırılır ve ikili (binary) değer almaktadırlar. Bağımsız değişkenler ise nitel veya nicel değerler alabilirler. Bağımlı değişkenlerin ikili değer aldığı durumlarla sosyal bilimlerde oldukça yaygın olarak karşılaşılmaktadır.³⁶

Bağımlı değişkenin iki kategorili olduğu durumlarda $E(y|x)$ modellenmesi için birçok dağılım fonksiyonu kullanılmakla birlikte, Lojistik fonksiyon en çok tercih edilen fonksiyondur³⁷. Cox ve Snell (1989) bu fonksiyonların bazıları üzerinde durmuş ve bunların içinde Lojistik fonksiyonun kullanılma nedeni olarak öncelikli iki sebep öne sürmüşlerdir. Bunlardan birincisi matematik olarak Lojistik fonksiyonun kullanımının daha esnek ve kolay olması, ikincisi ise açıklayıcı değişkenlerin daha anlamlı bir şekilde yorumlanabilmesi olarak gösterilmiştir³⁸.

Teknik olarak, Lojit ve Lojistik terimleri dönüşümler için verilen isimlerdir. Lojit dönüşümü, $[0, 1]$ arasında değer alan p değerlerine $\log[p/(1-p)]$ dönüşümünün uygulanması ile elde edilmektedir. Lojistik dönüşüm ise gerçek seri üzerinde bir x değeri alınıp bu değere $e^x/(1+e^x)$ dönüşümü uygulanması yoluyla bulunmaktadır. Lojit dönüşümü ve Lojistik dönüşüm birbirlerinden elde edilebilirler.³⁹

x_1, x_2, \dots, x_p açıklayıcı değişkenler biliniyorken y 'nin koşullu beklenen değeri basit gösterimiyle $P(y = 1|x_1, x_2, \dots, x_p) = E(y = 1|x_1, x_2, \dots, x_p)$ olarak ifade edilirse, Lojistik Regresyon Modeli basit formda;⁴⁰

³⁴ John H. Aldrich, Forrest D. Nelson, **Linear Probability Logit and Probit Models**, USA, Sage Publication, 1984, p. 41.

³⁵ Fred C. Pampel, **a.g.e.**, p. 60.

³⁶ J. Scott Long, **a.g.e.**, p. 34.

³⁷ David W. Hosmer, Stanley Lemeshow; **a.g.e.**, p. 6.

³⁸ D. R. Cox, E. J. Snell, **Analysis of Binary Data**, Second Edition, Chapman & Hall, 1989, p. 21.

³⁹ Ronald Christensen, **a.g.e.**, pp. 116–117.

⁴⁰ Fred C. Pampel, **a.g.e.**, p. 16.

$$P(y = 1|x_1, x_2, \dots, x_p) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}} \quad 1.18$$

olarak ifade edilir.

Lojistik regresyonun temelini oluşturan $P(y = 1|x_1, x_2, \dots, x_p)$ dönüşümü *Lojit dönüşümü* olarak tanımlanmakta ve aşağıdaki şekilde ifade edilmektedir⁴¹.

$$g(x) = \log \left[\frac{P(y = 1|x_1, x_2, \dots, x_p)}{1 - P(y = 1|x_1, x_2, \dots, x_p)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

Bu dönüşümün önemi, Doğrusal Regresyon Modelinin birçok önemli özelliğine sahip olmasından kaynaklanmaktadır. Lojit, $g(x)$, parametreleri bakımından doğrusaldır. Ayrıca p değeri arttıkça *lojit(p)* değeri de artar, p değeri $[0, 1]$ arasında değer alırken *lojit(p)*, $(-\infty, +\infty)$ aralığında değer almaktadır. Son olarak da gözlemlerin sınıflara atanmasında da kullanılan önemli bir özelliği de $p < 0,5$ olduğunda *lojit(p)* < 0 ve eğer $p \geq 0,5$ ise *lojit(p)* ≥ 0 değerini alacak olmasıdır.⁴²

Lojit ile sürekli değişkenler arasındaki ilişkinin doğrusallığını tanımlamak amacıyla “tasarım değişkenleri” (design variables) ve Kesirli Polinomlar (Fractional Polynomials) olmak üzere iki yöntem önerilmektedir.⁴³

Bu yöntemlerden Royston ve Altman (1994) tarafından geliştirilen Kesirli Polinomlar bir dönüşüm yaklaşımıdır ve tanımsal istatistik bilgilerinden kartil değerleri kullanılarak çizilen grafipler üzerinden değerlendirmesi yapılan tasarım değişkenleri yöntemine göre daha analitiktir⁴⁴. Bu yaklaşımla sürekli değişkenin en iyi modelde, x 'in kaçınıcı üssü olarak yer alacağına karar verilmektedir. Kesirli Polinomlar yöntemi Çok Değişkenli Lojistik Regresyon Modellerinde de kullanılabilir ancak gösterimdeki basitliği açısından bu çalışmada, bağımsız

⁴¹ Alan Agresti, **a.g.e.**, p. 103.

⁴² Hüseyin Tatlıdil, **Uygulamalı Çok Değişkenli İstatistiksel Analiz**, Cem Ofset Ltd. Şti., Ankara, 1996, s. 293.

⁴³ David W. Hosmer, Stanley Lemeshow; **a.g.e.**, p. 99.

⁴⁴ P. Royston, D. G. Altman, “Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling”, **Applied Statistics**, Vol: 43, 1994, pp. 429–467.

değişkenin tek ve sürekli olduğu durumun modellenmesi için tanımlanmıştır. Lojit'in $g(x, \beta) = \beta_0 + x\beta_1$ şekilde olduğunda, bu fonksiyonu genellemenin bir yolu $g(x, \beta) = \beta_0 + \sum_{j=1}^J F_j(x)\beta_j$ 'dir. Burada $F_j(x)$ fonksiyonu bir çeşit güç fonksiyonu, ilk fonksiyonun değeri ise $F_1(x) = x^{p_1}$ 'dir. Teoride gücü ifade eden p_1 değeri bir reel sayıdır. Royston ve Altman (1994) bu gücü $P = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$ kümesi ile sınırlamayı önermişlerdir. $p_1 = 0$ değeri değişkenin logaritmasını ifade etmektedir⁴⁵ ve ($j = 2, 3, \dots, J$) olmak üzere fonksiyon şu şekilde gösterilir;

$$F_j(x) = \begin{cases} x^{p_j}, & p_j \neq p_{j-1} \\ F_{j-1}(x) \ln(x), & p_j = p_{j-1} \end{cases}$$

$J = 1$ olduğu durum için ($p_1 \in P$) olacak şekilde 8 farklı model yazılabilmektedir ve bu modeller içinden log-olabilirliği büyük olan seçilir. $J = 2$ için 36 farklı model oluşturulur ve yine bunlar içinden en iyi model olarak, büyük log-olabilirliğe sahip olan alınır. Daha sonra bu en iyi modelin Doğrusal Model'den anlamlı bir şekilde iyi olup olmadığının testi yapılır. $J = 1$ iken, en iyi model Doğrusal Model'in kısmi Olabilirlik Oran (Likelihood Ratio test – LR) testi yapılarak karşılaştırılır,

$$LR(1, p_1) = -2 \ln\{L(1) - L(p_1)\}.$$

Buradan $L(.)$ ile olabilirlik fonksiyonu* (likelihood function) ifade edilmektedir ve elde edilen değerler, X 'in doğrusal dağıldığı sıfır hipotezi altında 1 serbestlik dereceli Ki-kare dağılımı ile karşılaştırılmaktadır. Aynı şekilde $J = 1$ ve $J = 2$ iken elde edilen en iyi modellerin karşılaştırılmasında ise kısmi Olabilirlik Oran istatistiğinden elde edilen değer, ikinci fonksiyonun değeri sıfıra eşittir sıfır hipotezi altında 2 serbestlik dereceli Ki-kare dağılımıyla test edilmektedir. Bu şekilde değişkeni en iyi modelleyen fonksiyona karar verilir⁴⁶.

⁴⁵ W. Sauerbrei, C. Meier-Hirmer, A. Benner, P. Royston, "Multivariable Regression Model Building by Using Fractional Polynomials: Description of SAS, STATA and R Programs", **Computational Statistics & Data Analysis**, Vol: 50, 2006, p. 3466.

* Olabilirlik Fonksiyonu (Likelihood Function) ile ilgili açıklama için sayfa 22'ye, Olabilirlik Oran Testi (Likelihood Ratio Test) için sayfa 26'ya bakınız.

⁴⁶ P. Royston, D. G. Altman **a.g.e.**, pp. 429–467.

Kay ve Little (1987)⁴⁷, Copas J. B. (1983)⁴⁸ Kesirli Polinomlar yöntemini uyarlayarak, bağımlı değişken grupları içinde nicel bağımsız değişkenin dağılımına bakıp, bu değişkenlerin yaklaşık olarak nasıl modelleneceğine karar veren yeni bir yöntem önermişlerdir. Örneğin ilgili nicel değişken bağımsız değişkenlerin grupları için farklı ortalama ve farklı varyansla Normal dağılıma uyuyorsa Lojit Modelde, doğrusal terim x ve karesel terim x^2 olmalıdır. Bu da $p_1 = 1$ ve $p_2 = 2$ ile $J = 2$ olan Kesirli Polinoma uyar. Eğer x değişkeni Gamma dağılıma uyuyorsa modelde x ve $\ln(x)$ 'e ihtiyaç vardır. Bu da $p_1 = 1$ ve $p_2 = 0$ ile $J = 2$ olan Kesirli Polinoma uyar. Eğer x değişkeni bağımlı değişkenin grupları için farklı parametre değerleri ile Beta dağılımına uyuyorsa nicel değişkenin doğru bir şekilde modellenmesi için $\ln(x)$ ve $\ln(1 - x)$ 'in modelde yer alması gerekmektedir. $\ln(1 - x)$ Kesirli Polinomlar içinde gösterilememektedir. Bunun dışında bir modelde kullanılan lojistik olasılık değeri bir başka modelde nicel değişken olarak kullanılabilir. Eğer bağımsız değişken grupları yeterli örneklem büyüklüğüne sahipse, Kay ve Little'in yaklaşımını kullanarak bağımsız nicel değişkenin dağılımına karar vererek değişkeni modellemek daha uygun olabilir.

1.2.1 Lojistik Regresyon Modellerine İlişkin Varsayımlar

Lojistik Regresyon Modelinin varsayımları aşağıdaki gibidir:⁴⁹

- İlgilenilen durum için grup üyelerinin durumları bilinmelidir. Örneğin, iki durumlu değişken için "1" değerini alma olasılığı P_1 ve "0" değerini alma olasılığı $P_0 = 1 - P_1$ olarak tanımlanmalıdır.
- Bağımlı değişkenler istatistik olarak birbirleri ile ilişkisiz olmalıdır.
- Model doğru olarak tanımlanmalıdır.
- Analiz edilen kategoriler *bağımsız ve bağdaşmayan* olaylar olmalıdır. Başka bir deyişle her olay tek bir duruma ait olmalıdır.

⁴⁷ R. Kay, S. Little, "Transformation of the Explanatory Variables in the Logistic Regression Model for Binary Data", **Biometrika**, Vol: 74, 1987, pp. 495-501.

⁴⁸ J. B. Copas, "Plotting p Against", **Applied Statistics**, Vol: 32, 1983, pp. 25-31.

⁴⁹ Raymond E. Wright, **Reading and Understanding Multivariate Statistics**, American Psychological Association, Edited By Laurence G. Grimm and Paul R. Yarnold, 1994, p. 220.

- Lojistik regresyon katsayılarının hipotez testlerinin uygunluğu için büyük örneklem gerekmektedir. Bunun nedeni, En Çok Olabilirlik Tahmininin katsayıların standart hatalarını büyük örneklem (bkz. sayfa 83) için hesaplanabilmesidir.

1.2.2 Lojistik Regresyon Analizinde Model Kurma Stratejileri

Model kurarken amaç, yeterince basit ama önemli etmenleri göz ardı edecek kadar da basitleştirilmemiş bir modele ulaşmaktır. Tesadüfi değişkenlerin özelliklerine ilişkin olarak yapılan her türlü varsayım model kurmanın ilk aşamasını oluşturmaktadır. Sonraki aşama uygun tahmin yöntemleri kullanarak parametre katsayılarının tahmin edilmesidir. Üçüncü aşama ise modelin doğrulanmasıdır. Bir model ileri sürülürken, bağımlı değişkenin yapısının ve bağımsız değişkenlerin iyi belirlenmesi ile birlikte uygun fonksiyonun seçilmesi de önemlidir. Bu aşamada sorun varsa ilk adım olan model kurma aşamasına geri dönülür ve katsayılar tekrar tahmin edilerek sürece devam edilir. Son aşama olan sonuçları yorumlama ve çıkarım yapma aşamasında model kurma hataları ne kadar ciddiye, uygulanan modelden yapılan çıkarımlar da o kadar az güvenilir olacak demektir⁵⁰.

Lojistik Regresyon Modeli için değişken seçimi birkaç adımdan oluşmaktadır.⁵¹

- 1- Değişken seçimine her bir değişkenin tek başına analizi ile başlanmaktadır. Bunun için Olabilirlik Oran testi ya da onunla aynı sonucu veren Pearson Ki-kare testi kullanılabilir. Nicel değişkenler için katsayıların anlamlılığı önemlidir ve bu nedenle her bir değişkenin Wald istatistiğine bakılır. Değişkenlerin tek tek testinde bu istatistiğe denk olan t-testi de kullanılabilir. Nicel bağımsız değişkenlerin tek değişkenli olarak Lojistik regresyona uygunluğu için nicel değişkenle lojit değerleri arasında doğrusal bir ilişki olması beklenmektedir.
- 2- Değişkenlerin tek tek analizi tamamlandıktan sonra, seçilen değişkenlerin çoklu analizi yapılmaktadır. Lojistik Regresyon Analizi için genellikle α değeri 0.25 olarak alınmaktadır. Anlamlılık seviyesi için geleneksel olarak kullanılan 0.05

⁵⁰ Paul Newbold, William L. Carlson, Betty Thorne, **Statistics for Business and Economics**, Sixth edition, Prentice Hall 2007, p.539.

⁵¹ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, p. 92.

ve 0.01 değerleri alındığında model için önemli olan değişkenlerin dışarıda kalabileceğini gösteren çalışmalar bulunmaktadır⁵². Genellikle, tüm mümkün değişkenlerle modele başlama kararının uygunluğu, örnek büyüklüğüne ve aday değişkenlerin toplam sayısı ile ilişkili her bir bağımlı değişkeninin grup sayısına bağlıdır. Veri seti örnek büyüklüğü açısından uygunsuzsa analize çok değişkenli modelle başlamak uygun olabilir⁵³. Bununla birlikte, veri seti uygun değilse yani örneklem büyüklüğü yeterli değilse çok değişkenli modelle başlamak doğru değildir. Bu durumda değişken seçimi için değişkenlerin tek tek testine dayanan Wald istatistiği kullanılmamalıdır. Onun yerine, modeli oluşturacak değişkenlerin seçimi En İyi Alt Küme (Best Subsets)⁵⁴ seçimine dayandırılır.

Diğer bir değişken seçimi Adım Adım (Stepwise) yöntemidir. Uygulamada Adım Adım yöntemi, İleri Seçim (Forward Selection) ve Geriye Doğru Eleme (Backward Elimination) olmak üzere iki temel prensiple hareket etmektedir. Adım Adım, En İyi Alt Küme ve diğer mekanik seçim yöntemlerinde, modele alınan değişkenler konusunda dikkatli olunması gerekmektedir⁵⁵.

- 3- Bu adımda, çok değişkenli modelin uygunluğu, modelde bulunan her bir değişkenin önemi Wald istatistiği ile test edilerek doğrulanmalıdır. Kriteri sağlamayan değişkenler modelden çıkartılarak model yeniden oluşturulur ve yeni modelle eski modelin Olabilirlik Oran testi hesaplanarak karşılaştırılır. Ayrıca, modelde kalan değişkenlerin tahmin edilen parametreleri tam modelin parametreleri ile kıyaslanır. Bu işleme tüm önemli değişkenler modele alınana kadar devam edilir. Bu adımın sonunda ortaya çıkan modele *Başlangıç Temel Etki Modeli* (Preliminary Main Effects Model) adı verilmektedir.⁵⁶
- 4- Önemli değişkenlerin içinde olduğu model oluşturulduktan sonra, modeldeki değişkenler daha yakından incelenmelidir. Lojit ile nicel değişken arasındaki ilişki doğru olarak tanımlandıktan sonra (bu ilişki doğrusal, karesel veya

⁵² J. Mickey, S. Greenland “A Study of the Impact of Confounder-selection Criteria on Effect Estimation”, **American Journal of Epidemiology**, 129, pp. 125–137.

⁵³ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, p. 95.

⁵⁴ David W. Hosmer, B. Jovanovic, Stanley Lemeshow, “Best Subsets Logistic Regression”, **Biometrics**, Vol:45, 1989, pp. 1265–1270.

⁵⁵ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, p. 96.

⁵⁶ **A.e.**, p. 97.

doğrusal olmayan bir fonksiyon olabilir) nicel değişkenlerin lojit değerleri ile doğrusallık varsayımı kontrol edilmelidir. Doğrusallık varsayımının sağlandığı fonksiyon için parametre katsayıları elde edilmektedir. Dördüncü adım sonunda oluşturulan modele *Temel Etkili Modeller* (The Main Effects Model) adı verilmektedir.⁵⁷

- 5- Son olarak modeldeki değişkenlerin etkileşimlerinin kontrolü yapılır. Değişkenlerin etkileşimli halleri modele teker teker alınır ve belirli anlamlılık düzeyinde Olabilirlik Oran testi yapılır. Adım beşin sonunda elde edilen modele *Sonuç Model* (Final Model) denir.⁵⁸ Modeli yazmadan önce modelin uygunluk testleri yapılmalıdır (bkz. Başlık 1.2.5'e).

1.2.3 Lojistik Modellerde Parametre Tahmin Yöntemi

Parametre tahminleri açısından Doğrusal ve Lojistik Regresyon Analizi arasındaki farklardan iki tanesi çok önemlidir. $y = E(y|x) + \varepsilon$ şeklinde ifade edilen modelde ε 'la gösterilen hata teriminin, Doğrusal Regresyon Modelinde En Küçük Kareler (EKK) yöntemi ile parametre tahminlerinin geçerli olması için Normal dağılımlı ve eşit varyanslı olması ($\varepsilon \sim N(0, \sigma^2)$) gerekmektedir. Lojistik Regresyon Analizinde bağımsız değişkenin gözlenen değeri $y = P(x) + \varepsilon$ olarak gösterilebilir ve buradaki hata terimi; eğer $y = 1$ ise $P(x)$ olasılığı ile $\varepsilon = 1 - P(x)$ 'e ve eğer $y = 0$ ise $1 - P(x)$ olasılıkla $\varepsilon = -P(x)$ 'e eşit olacak şekilde iki mümkün sonuca sahiptir. Buradan da, bağımlı değişkenin koşullu dağılımının, koşullu beklenen değer $P(x)$ ile Binom dağılıma uyduğu söylenebilir.⁵⁹ Buradan da anlaşılacağı gibi Lojistik Regresyon Analizinde parametre tahminleri için, varsayımların sağlanamaması nedeniyle, EKK yönteminin kullanılması uygun değildir.

İki Kategorili Lojistik Regresyon Modellerinde parametre tahminlerinin yapılabilmesi için literatürde bazı yöntemler önerilmektedir. Bu tekniklerden uygulamada sıklıkla kullanılan En Çok Olabilirlik Tahmini'dir (Maksimum Likelihood Estimation-MLE). Bunun yanı sıra, Yeniden Ağırlıklandırılmış Tekrarlı

⁵⁷ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, p. 97.

⁵⁸ **A.e.**, p. 99.

⁵⁹ **A.e.**, p. 7.

En Küçük Kareler Tahmini (Reweighted Iterative Least Square Estimation), 1944 yılında Berkson tarafından Ağırlıklı En Küçük Kareler Tahmininden, tekrarlı veri durumunda kullanılmak üzere geliştirilmiş olan Minimum Lojit Ki-kare (Minimum Logit Chi-Square) ve Diskriminant Fonksiyonundan elde edilen Diskriminant katsayıları⁶⁰ kullanılan diğer tahmin yöntemleri arasında yer almaktadır.

Diskriminant fonksiyon tahminlerinin tutarlı olabilmesi için normallik varsayımının sağlanması gerekmektedir. Eğer modelde hem nicel hem de nitel bağımsız değişkenler birlikte bulunuyor ve nitel bağımsız değişkenlerin sayısı fazla ise, bu durum normallik varsayımının bozulmasına neden olmaktadır. Bu sebeple Diskriminant Fonksiyon Tahmin yöntemiyle elde edilen katsayı değerleri olması gerekenden daha büyük çıkacaktır⁶¹.

Bu yöntemler içinde en sık kullanılan nokta tahmin yöntemi “En Çok Olabilirlik” Tahminidir.

En Çok Olabilirlik Tahmin (Maksimum Likelihood Estimation – MLE) Yöntemi*:

Anakütleyi belirleyen olasılık yoğunluk fonksiyonu ya da olasılık fonksiyonu $f(y; \theta)$, bu anakütleden çekilmiş n gözlemlili rastgele örneklem Y_1, Y_2, \dots, Y_n ’nin gerçekleşen bir alt kümesi ise y_1, y_2, \dots, y_n olarak gösterildiğinde, En Çok Olabilirlik Tahmin yöntemi belli bir örneklem değerinin gerçekleşme olasılığını en yüksek yapan parametre değerlerini bulmaya çalışmaktadır⁶². Bir rassal örneklem söz konusu olduğunda ve bunların çekildiği anakütle dağılımının bilindiği varsayıldığında; bağımsızlık özelliğinden hareketle ortak olasılık yoğunluk fonksiyonu,

$$\begin{aligned} f(y_1, y_2, \dots, y_n; \theta) &= f(y_1; \theta) \cdot f(y_2; \theta) \cdot \dots \cdot f(y_n; \theta) \\ &= \prod_{i=1}^n f(y_i; \theta) \end{aligned}$$

⁶⁰ Hüseyin Tatlıdıl, **a.g.e.**, s. 293.

⁶¹ David W. Hosmer, Stanley Lemshow, **a.g.e.**, pp. 21–22.

* Bazı kaynaklarda En Çok Benzerlik Tahmini olarak da kullanılmaktadır. (bkz. Neyran Orhunbilge, **Çok Değişkenli İstatistik** kitabı)

⁶² Raymond E. Wright, **a.g.e.**, p. 225.

olarak yazılmaktadır⁶³.

Olabilirlik fonksiyonu, ortak olasılık fonksiyonuna verilen başka bir isimdir. Tek fark ortak olasılık fonksiyonunda θ 'nın bilindiği Y 'lerin bilinmediği, olabilirlik fonksiyonunda ise Y 'lerin bilindiği θ 'nın ise bilinmediği varsayılmaktadır⁶⁴. Olabilirlik fonksiyonu, rassal örneklemin belli bir sonucu olan y ($y = \{y_1, y_2, \dots, y_n\}$) biliniyorken θ 'yı bilinmeyen olarak ifade eden bir fonksiyondur ve aşağıdaki şekilde gösterilmektedir.

$$L(\theta|y_1, y_2, \dots, y_n) = L(\theta|y) = \prod_{i=1}^n f(y_i; \theta)$$

Bu olabilirlik fonksiyonu bağımlı değişkenin iki kategorili olduğu durum için genelleştirildiğinde; istenen durumun ortaya çıkma olasılığı $p_i = P(y_i = 1|x_i)$ ve istenmeyen durumunun ortaya çıkma olasılığı $1 - p_i = P(y_i = 0|x_i)$ olarak gösterilirse, i. gözlem için olasılık değeri,

$$P(y_i|x_i) = p_i^{y_i}(1 - p_i)^{1-y_i} \quad \mathbf{1.19}$$

olarak ifade edilmektedir. Bu olasılığın birbirinden bağımsız n gözlem için olabilirlik fonksiyonu aşağıdaki gibidir⁶⁵.

$$l(\beta|y, x_i) = \prod_{i=1}^n p_i^{y_i}(1 - p_i)^{1-y_i}$$

Bilindiği gibi En Çok Olabilirlik Tahmin yönteminin amacı açıklayıcı değişkenlerin katsayısı olan β değerlerini bağımlı değişkeni y 'nin gözlenme olabilirliğini maksimum yapacak biçimde tahmin etmektir. Kısaca, $\max_{\beta} \ln[l(\beta|y, x_i)] = \ln\left[\prod_{i=1}^n p_i^{y_i}(1 - p_i)^{1-y_i}\right]$ olarak tanımlanmaktadır.⁶⁶

⁶³ David W. Hosmer, Stanley Lemeshow; **a.g.e.**, pp. 7–9.

⁶⁴ Jeff Gill, **a.g.e.**, p. 21.

⁶⁵ Fred C. Pampel, **a.g.e.**, p. 41.

⁶⁶ Hüseyin Tatlıdil, **a.g.e.**, s. 296.

Bu maksimizasyon probleminde kolaylık sağlamak için ortak olasılık fonksiyonunun doğal logaritması kullanılarak;

$$\max \ln L(\beta|y, x_i) = \ln \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i} = \sum_{i=1}^n [y_i \ln p_i + (1 - y_i) \ln(1 - p_i)] \quad 1.20$$

fonksiyonu elde edilir.

Lojistik Regresyon Modelinden, Eşitlik 1.18’de verilen olayın gerçekleşme olasılık değeri Eşitlik 1.20’da yerine konularak fonksiyonun β parametrelerine göre birinci mertebeden türevi alınıp, sıfıra eşitlenerek elde edilen denklemlerin çözümleri $\hat{\beta}$ ’leri vermektedir. Fakat bu eşitliklerdeki β parametrelerinin olabilirlik fonksiyonu içinde doğrusal olmaması nedeniyle iteratif yolla çözüme gidilmektedir.

En Çok Olabilirlik Tahmininin, iteratif çözüm yöntemleri içinde; En Hızlı Artış (Steepest Ascent), Newton-Raphson, Derecelendirme, BHHH* ve Sayısal Türevler (Numerical Derivatives) gibi yöntemler bulunmaktadır⁶⁷. Bu yöntemlerden tercih edilen ve birçok paket programın kullandığı yöntem Newton-Raphson yöntemidir.

Tahmin edilen parametrelerin yeterlilik koşulunu sağlaması için log-olabilirlik fonksiyonunun $[\ln L(\beta|y, x_i)]$ bağımsız değişkenlerin katsayılarını ifade eden β parametrelerine göre ikinci mertebeden türevlerinin (Hessian matrisini verir) sıfırdan küçük olması gerekmektedir $\left[\frac{\partial^2}{\partial \beta^2} \ln L(\beta|y, x_i) < 0 \right]$. Örnek büyüklüğü “ ∞ ”a giderken bu koşulların gerçekleşmesi ile En Çok Olabilirlik Tahminleri tutarlı, asimptotik olarak ise normal ve etkin⁶⁸ olurlar. Bu, En Çok Olabilirlik Tahmininin küçük örneklem büyüklüğü için kötü tahmin yapacağı anlamına gelmemektedir.

En Çok Olabilirlik Tahmininde örnek büyüklüğü için literatürde farklı kriterler bulunmaktadır. Genellikle 100’den küçük örnek büyüklüğü kullanmak riskli kabul edilirken, bazı kaynaklarda büyüklüğün 500’ün üzerinde olması gerektiği değişebilsede öncelikle modelde ne kadar çok değişken varsa o kadar çok gözleme ihtiyaç

* BHHH: Berndt, E. R., Hall, B. H., Hall, R. E. ve Hausman, J. A

⁶⁷ J. Scott Long, **a.g.e.**, pp. 54–57. (daha fazla bilgi için bkz.)

⁶⁸ Damodar N. Gujarati, **Temel Ekonometri**, Literatür Yayıncılık, 1999, ss. 110–113.

olduğu söylenebilir. Bazı kaynaklarda, her bağımsız değişken için en az 50 gözlemin alınması gerektiği ifade edilmektedir⁶⁹. İkinci olarak eğer veri setindeki bağımsız değişkenler arasında yüksek korelasyon varsa ya da bağımlı değişkendeki değişiklik azsa yani neredeyse gözlemlerin çoğu gruplardan birinde yer alıyorsa örneklemin yine büyük olması beklenir. Ayrıca bağımlı değişkenin sıralı (ordinal) olduğu model analizi gibi bazı modeller daha fazla örnek büyüklüğü gerektirmektedir⁷⁰.

Lojistik regresyonda parametre tahminleri için genellikle En Çok Olabilirlik Tahmin yöntemi seçilmektedir. Lojistik regresyon katsayıları “En Çok Olabilirlik parametre tahminleri” olarak bilinmektedir. Araştırmacılar tarafından genelde En Çok Olabilirlik değeri yerine Log-olabilirlik (LL) değeri kullanılmaktadır. Log-olabilirlik değerinin “0” dan küçük değerler aldığı durumlarda 2 ile çarpılarak, büyüklüğü negatiflikten kurtarmak için de “-” ile çarpılarak⁷¹ sapma değeri (-2LL) raporlanmaktadır. Log-olabilirlik (LL) genellikle negatif, sapma değeri genellikle pozitifdir. Katsayıların olabilirliğini maksimum yapmaya çalışmak aynı zamanda LL’yi maksimum yapmak ve sapmayı minimum yapmakla eşdeğerdir⁷².

1.2.4 Parametrelerin Anlamlılık Testleri

Herhangi bir modelin katsayılarının anlamlılığını test etmede amaç “bağımsız değişken modelde olduğu zaman bağımlı değişken hakkında modelde bulunmadığı duruma göre daha mı fazla bilgi veriyor” sorusunun cevabını bulmaktır. Buradan yola çıkarak test istatistikleri, bağımsız değişkenin modele dahil edildiği ve edilmediği iki durum için modelin tahminlerinin farklarının alınmasına dayanmaktadır.

En Çok Olabilirlik Tahminleri asimptotik olarak Normal dağılmaktadırlar. Bunun anlamı örneklem büyüklüğü arttıkça, En Çok Olabilirlik Tahminleyenlerinin örneklem dağılımlarının Normal dağılıma yaklaştığıdır. Tek bir parametre için bu yaklaşım $\hat{\beta}_k \sim N(\beta_k, Var(\hat{\beta}_k))$ iken parametrelerin bir vektörü için

⁶⁹ Raymond E. Wright, **a.g.e.**, p. 221.

⁷⁰ J. Scott Long, **a.g.e.**, p. 54.

⁷¹ Neyran Orhunbilge, **Çok Değişkenli İstatistik Yöntemler**, İstanbul Üniversitesi Basım ve Yayımevi Müdürlüğü, İstanbul, 2010, s. 191.

⁷² Raymond E. Wright, **a.g.e.**, p. 230.

$\hat{\beta} \sim N(\beta, Var(\hat{\beta}))$ 'dir. Burada $Var(\hat{\beta})$, $\hat{\beta}$ 'nin kovaryans matrisidir⁷³. En Çok Olabilirlikle, tahminleri yapılan Lojistik regresyon parametrelerinin anlamlılığı Wald ve Olabilirlik Oran testleri ile test edilmektedir.

Wald Testi⁷⁴

Bu test Regresyon Analizinde regresyon katsayılarının t veya Z testlerinin benzeridir.⁷⁵ Wald testinde kısıtların doğrusal formlarını göz önüne alarak sıfır hipotezi en genel hali ile $H_0: Q\beta = r$ şeklinde gösterilmektedir. Bu eşitlikte β , testi yapılacak parametreler vektörü; Q , sabit sayı matrisi; r ise sabit sayı vektörüdür. Q ve r doğrusal kısıtlardaki değişimi ifade etmektedir.

$H_0: Q\beta = r$ hipotezi Wald istatistiği ile şu şekilde test edilebilir;

$$W = [Q\hat{\beta} - r]' [Q\widehat{Var}(\hat{\beta})Q']^{-1} [Q\hat{\beta} - r]. \quad 1.21$$

W , sabitlerin sayısı kadar yani Q 'daki satır sayısına eşit olan serbestlik derecesi ile Ki-kare dağılımına uymaktadır. Wald istatistiğinin iki bileşeni bulunmaktadır. Birincisi, formüldeki tahmin ve hipotez değeri arasındaki farkı ölçen $(Q\hat{\beta} - r)$ değeridir. Farkların karesi negatif ve pozitif farkın test istatistiği üzerinde aynı etkiye sahip olmasını sağlamaktadır. İkincisi ise tahminleyenin değişkenliğini yansıtan $[Q\widehat{Var}(\hat{\beta})Q']^{-1}$ değeridir.

İki değişkenli Lojit Model ele alınır; $\Pr(y = 1|x) = \Lambda(\beta_0 + \beta_1x_1 + \beta_2x_2)$ için $H_0: \beta_1 = \beta^*$ testi matris formunda şu şekilde ifade edilmektedir;

$$Q\hat{\beta} - r = (0 \quad 1 \quad 0) \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} - (\beta^*) = \hat{\beta}_1 - \beta^*$$

$$[Q\widehat{Var}(\hat{\beta})Q']^{-1} = \left[(0 \quad 1 \quad 0) \widehat{Var}(\hat{\beta}) \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right]^{-1} = \frac{1}{\widehat{\sigma}_{\hat{\beta}_1}^2}$$

⁷³ J. Scott Long, **a.g.e.**, p. 89.

⁷⁴ **A.e.**, pp. 89–92.

⁷⁵ Neyran Orhunbilge, **Çok Değişkenli İstatistik Yöntemler, a.g.e.**, s. 192.

Bu ifade testi yapılan parametre varyansının tersidir. Varyans büyüdükçe, hipotez ve tahmin değeri arasındaki farkı veren ağırlık azalmaktadır. Eşdeğer olarak, olabilirlik fonksiyonunun eğimi $\hat{\beta}_1$ 'nin etrafındaki bölgede değişmekte ve $(\hat{\beta}_1 - \beta^*)$ farkı daha anlamlı olmaktadır. Wald istatistiği aşağıdaki şekilde formüle edilmektedir.⁷⁶

$$W = \frac{(\hat{\beta}_1 - \beta^*)^2}{\hat{\sigma}_{\hat{\beta}_1}^2} = \left(\frac{\hat{\beta}_1 - \beta^*}{\hat{\sigma}_{\hat{\beta}_1}} \right)^2 \quad 1.22$$

Wald istatistiği, Z test istatistiğinin karesidir. “1” serbestlik dereceli Ki-kare değeri normal değer karesine eşit olmaktadır.⁷⁷

Aynı şekilde $H_0: \beta_1 = \beta_2 = 0$ hipotezi test edilmek istendiğinde Wald istatistiği aşağıdaki şekilde hesaplanmaktadır.

$$W = \sum_{k=1}^2 \frac{\hat{\beta}_k^2}{\hat{\sigma}_{\hat{\beta}_k}^2} = \sum_{k=1}^2 z_{\hat{\beta}_k}^2 \quad 1.23$$

Wald istatistiğinde Lojistik regresyon katsayıları mutlak değer olarak büyüdükçe, bu değerlerden tahmin edilen standart hatalar da çok fazla büyüyecektir.

Olabilirlik Oran Testi

Olabilirlik Oran testi (Likelihood Ratio test-LR test), Lojistik Regresyon Modeli'ndeki kısıtların* testinde kullanılmaktadır. Örnek olarak aşağıdaki Lojit Modeller ele alındığında,⁷⁸

$$M_1: \Pr(y = 1|x) = \Lambda(\beta_0 + \beta_1x_1 + \beta_2x_2)$$

$$M_2: \Pr(y = 1|x) = \Lambda(\beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3)$$

$$M_3: \Pr(y = 1|x) = \Lambda(\beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_4x_4)$$

⁷⁶ W. W. Hauck, A. Donner, “Wald’s Test as Applied to Hypotheses in Logit Analysis”, **American Statistical Association**, Vol: 72, No: 360, 1977, pp. 851–853.

⁷⁷ A. Buse, “The Likelihood Ratio, Wald and Lagrange Multiplier Tests: An Expository Note”, **The American Statistician**, Vol: 36, No: 3, 1982, pp. 153–157.

* Bu bölümde kullanılan “kısıt” kelimesi parametre değerinin “0” olması durumunu ifade etmektedir. Genellikle kaynaklarda Olabilirlik Oranı; değişkenin modelde yer almadığı olabilirlik değerinin, değişken modelde bulunduğu olabilirlik değerine oranı olarak verilmektedir.

⁷⁸ J. Scott Long, **a.g.e.**, p. 93.

$$M_4: \Pr(y = 1|x) = \Lambda(\beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4)$$

model M_1 'in, $\beta_3 = 0$ kısıtının geçerli olduğu M_2 'nin ve aynı zamanda $\beta_4 = 0$ kısıtının geçerli olduğu M_3 'ün formunda olduğu görülmektedir. Bir model kısıtlarla birlikte diğer bir modelden elde edilebiliyorsa, kısıtlanmış model kısıtlanmamış modelin içinde yer almaktadır (nested – iç içe) denmektedir. O halde M_1 , M_2 ve M_3 'ün içinde yer almaktadır. Bununla birlikte; ne M_2 'nin M_3 'ün içinde, ne de M_3 'ün, M_2 'nin içinde olduğu görülmektedir.

Olabilirlik Oran testi, β_U parametreleri ile kısıtlanmamış model M_U ve bu model içinde yer alan β_R parametreleri ile kısıtlanmış M_R modelini karşılaştırmaktadır. Sıfır hipotezi, “kısıtlanmış modeli ifade eden M_R modeli doğrudur” olarak kurulmaktadır. Kısıtlanmamış model için En Çok Olabilirlik Tahminlerinden elde edilen olabilirlik fonksiyonunun değeri $L(M_U)$ ve kısıtlanmış tahminlerin değeri $L(M_R)$ ise, Olabilirlik Oran (LR) istatistiği,

$$LR = -2 \ln[L(M_R) - L(M_U)] = 2 \ln L(M_U) - 2 \ln L(M_R)$$

olarak hesaplanmaktadır. LR , bağımsız kısıtların sayısı kadar serbestlik derecesi ile Ki-kare dağılımına uymaktadır. LR istatistiği iç içe model çiftlerinin karşılaştırılmasında kullanılmaktadır. Bu testin sonucunda hatalı karar verme riskini azaltmak için, karşılaştırılacak modellerin tamamının aynı örneklem kümesi için kullanıldığından emin olmak gerekmektedir. Herhangi bir modelin testine başlamadan önce, modelde kullanılan tüm değişkenler için gözlenmemiş değerlerin olduğu gözlem birimlerinin dışarıda bırakılarak veri kümesinin tekrar yapılandırılması önemlidir⁷⁹.

Standart istatistik paket programları genellikle LR için iki test değeri hesaplamakta ve bu hesap değerlerine ML tarafından modellenen sonuçların sunulduğu tablonun içinde yer vermektedirler. İlk test, verilen modelde tüm bağımsız değişken katsayılarının sıfıra eşit olduğunu öngören kısıtlanmış modelin karşılaştırılmasıdır. Bu testten genellikle Olabilirlik Oran Ki-kare testi olarak bahsedilmektedir. Model

⁷⁹ J. Scott Long, **a.g.e.**, p. 93.

M_β , sabit ve bağımsız değişken katsayılarının içinde bulunduğu kısıtlanmamış modeli; M_α ise tüm bağımsız değişken katsayılarının model dışı bırakıldığı ve sadece sabitin modelde olduğu kısıtlı modeli temsil etmektedir. “Bağımsız değişken katsayılarının tamamı eş zamanlı olarak sıfıra eşittir” hipotezini incelemek üzere kullanılacak test istatistiği aşağıdaki gibi yazılmaktadır.

$$LR = 2\ln L(M_\beta) - 2\ln L(M_\alpha)$$

İkinci test olan D istatistiği, 1989 yılında McCullagh ve Nelder tarafından sapma (deviance) olarak adlandırılmıştır ve bu istatistik değeri GLM’in iskeletinin oluşturulmasında geniş bir şekilde yer almıştır⁸⁰. Aynı zamanda bu değer, modellerin uyum iyiliği testleri için geliştirilen bazı yaklaşımlarda da önemli bir rol üstlenmektedir. Bununla birlikte sapma değeri, ifade olarak Doğrusal Regresyon Analizi içindeki hata kareleri toplamının benzeridir.⁸¹ Sapma, verilen modelle (M_β) tam modeli (M_F) karşılaştırmaktadır. Gözlemlenmiş veri kümesi değerleri tam olarak tahmin edildiğinde, M_F ’nin olabilirliği “1” ve log-olabilirlik “0”dır. Sapma aşağıdaki gibi yazılmaktadır.

$$\begin{aligned} D(M_\beta) &= D(M_\beta|M_F) = 2\ln L(M_F) - 2\ln L(M_\beta) \\ &= -2\ln L(M_\beta) \\ &= LR(M_\beta|M_F) \end{aligned}$$

$D(M_\beta)$ istatistiği Ki-kare dağılımına uygun olmakla birlikte, gözlemlerin sayısı ile tam modeldeki parametrelerin sayısının doğrudan artması nedeniyle $D(M_\beta)$ asimptotik olarak Normal dağılıma da uymaktadır⁸². Hosmer ve Lemeshow (2000), seyrek veriler* söz konusu olduğunda modelin uygunluk ölçüsü olarak $D(M_\beta)$ ’nin kullanılmasının uygun olmadığını vurgulamaktadır⁸³.

⁸⁰ P. McCullagh, J. A. Nelder, **a.g.e.**, pp. 33–34.

⁸¹ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, p. 13.

⁸² P. McCullagh, “The Conditional Distribution of Goodness-of-fit Statistics for Discrete Data”, **Journal of the American Statistical Association**, Vol:81, No:393, 1986, pp. 104–107.

* Seyrek veri ile bağımsız değişkenlerin değerlerinin her bir kombinasyonunun örneklemede sadece bir kere ortaya çıktığı durumdan bahsedilmektedir.

⁸³ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, pp. 143–147.

LR ve Wald testleri asimptotik olarak eşit oldukları halde, sonlu örneklem, özellikle de küçük örneklem için farklı sonuçlar vermektedirler. Genellikle hangi testin tercih edileceği çok açık değildir. Rothenberg (1984) testlerden hiçbirinin diğerine göre üstün olmadığını ifade etmiştir⁸⁴. Hauck ve Donner (1977) yaptıkları çalışmada LR ve Wald testini karşılaştırmış ve LR testinin daha güçlü olduğunu söylemişlerdir⁸⁵. Uygulamada, hangi testin kullanılacağını seçimi sıklıkla kullanılan bilgisayar programının uygunluğu tarafından belirlenmektedir. Bununla birlikte, birçok istatistikçi nispeten kolaylığı nedeniyle LR testini tercih etmektedir⁸⁶.

1.2.5 Uyum İyiliği Testleri

Uygunluk ölçüsü bir modelin yeterli olup olmadığı hususunda kabaca bir indeks (gösterge) sağlayabilir. Ancak uygunluk ölçüleri model hakkında bazı bilgileri elde etmemizi sağlasa dahi, bu değerlerin geçmiş araştırma kapsamında ve dikkate alınan modelin tahmin parametreleri içinde değerlendirilmesi gereken sadece kısmi bilgiler olduğu unutulmamalıdır.

Pearson Ki-Kare Testi

Lojistik regresyon için gözlenen ve beklenen değer arasındaki farklar yani hatalar, İki Kategorili Model yani $P_i = E(y_i|x_i) = \Pr(y_i = 1|x_i)$ için, $Var(y_i|x_i) = P_i(1 - P_i)$ ile $y_i - P_i$ sapmaları farklı varyanslıdır ve hatalar şu şekilde hesaplanmaktadır,

$$e_i = \frac{y_i - \hat{P}_i}{\sqrt{\hat{P}_i(1 - \hat{P}_i)}}.$$

e_i 'nin büyük değerleri eldeki gözlemlerin modele uygun olmadığını göstermektedir. Bu hesaplanan hatalara dayanan özet istatistik Pearson Ki-kare istatistiğidir ve formülasyonu aşağıdaki gibidir,

⁸⁴ T. J. Rothenberg, "Hypothesis Testing in Linear Models When the Error Covariance Matrix is Nonscalar", *Econometrica*, 1984, Vol: 52, pp. 827–842.

⁸⁵ W. W. Hauck, A. Donner, *a.g.e.*, pp. 851–853.

⁸⁶ J. Scott Long, Simon Cheng, *The Handbook of Data Analysis*, Edited by Hardy and Bryman pp. 261–262.

$$\chi^2 = \sum_{i=1}^N e_i^2$$

χ^2 , Ki-kare dağılımına sahiptir.

$Var(y_i - P_i) = P_i(1 - P_i)$ iken, $Var(y_i - \hat{P}_i) \neq \hat{P}_i(1 - \hat{P}_i)$ 'dir. Sonuç olarak e_i 'nin varyansı "1"e eşit değildir. Standartlaştırılmış hataların hesaplanmasında şapka (hat)⁸⁷ matris veya önkestim (predictive)⁸⁸ matrisi olarak adlandırılan H matrisi önemli bir rol oynamaktadır⁸⁹. Tahmin edilen hataların varyansları hesaplanırken, önkestim matrisi H 'nin ne olduğunun bilinmesi gerekmektedir. İkili regresyon modeli için Pregibon (1981) H matrisi aşağıdaki şekilde türetmiştir,

$$H = \hat{V}X(X'\hat{V}X)^{-1}X'\hat{V}.$$

\hat{V} , $\sqrt{\hat{P}_i(1 - \hat{P}_i)}$ 'nin köşegenleri üzerindeki değerlerden oluşan köşegen matrisi; X , örnek büyüklüğü n ve modeldeki bağımsız değişken sayısı p ile ifade edilmek üzere $n \times p$ boyutunda veri matrisidir. H 'nin sadece köşegen elemanlarına ihtiyaç duyulduğundan, hesaplama için daha basit olan

$$h_{ii} = \hat{P}_i(1 - \hat{P}_i)x_i\widehat{Var}(\hat{\beta})x_i'$$

formülü kullanılmaktadır. x_i , i . gözlem için bağımsız değişkenlerin değerlerinden oluşan satır vektörü ve $\widehat{Var}(\hat{\beta})$, En Çok Olabilirlik tahminleri $\hat{\beta}$ 'den elde edilen tahminlerin kovaryans değeridir⁹⁰. e_i 'nin varyansını tahmin etmek için $1 - h_{ii}$ kullanılır. Standartlaştırılmış Pearson hataları ise,

$$e_i^{std} = \frac{e_i}{\sqrt{1 - h_{ii}}}$$

⁸⁷ J. Scott Long, **a.g.e.**, p.99.

⁸⁸ Samprit Chatterjee, Ali. S. Hadi, Bertram Price, **Regression Analysis by Example**, Third Edition, John Wiley & Sons, 2000, p.36.

⁸⁹ Erling B. Andersen, **Introduction to the Statistical Analysis of Categorical Data**, Springer, 1997, p. 30, 177.

⁹⁰ D. Pregibon, "Logistic Regression Diagnostic", **Annals of Statistics**, Vol:9, 1981, pp. 705-724.

şeklinde ifade edilir. Uygulamada genellikle bu iki hata değeri birbirine benzemekle birlikte e^{std} , e yerine tercih edilmektedir.

Bununla birlikte McCullagh (1986), Pearson Ki-kare ve daha önce bahsedilen sapma testinin (D), modelin uygunluğunu test etmek için özellikle kapsamlı ve seyrek veri kümesi söz konusu olduğunda kullanışlı istatistikler olmadığını ve χ^2 'nin hesaplanması zor olan ortalama ve varyans ile asimptotik Normal dağılıma uyduğunu ispat etmiştir⁹¹. McCullagh'in kapsamlı ve seyrek veri kümesi ile kastettiği bağımsız gözlemlerin sayısı veya "hücreler" çok büyük olduğu halde, her bir hücrenin toplam içindeki katkısının miktarı bakımından küçük olduğu durumlardır. Ayrıca McCullagh ve Nelder (1989), χ^2 'nin kesin uyum için kullanılmaması gerektiğini ve küçük örneklem için Olabilirlik Oran testinin Pearson Ki-kare testinden daha güçlü olduğu iddiasının da tersini söylemişlerdir⁹².

Hosmer-Lemeshow Testi

Hosmer ve Lemeshow (2000), seyrek veri kümesi için alternatif test olarak grup verilerini yapılandırarak test etmeyi önermişlerdir. n birime ait tahmin edilen olasılıkları küçükten büyüğe sıraladıktan sonra, iki farklı yöntem kullanılabilir. Yöntemlerden ilki $g = 10$ 'u grup sayısı olarak kullanarak $n'_1 = n/10$ adet En Küçük Tahmin olasılıklarına sahip birimi birinci gruba yerleştirip, son grup olan $n'_{10} = n/10$ birimi de en büyük tahmin olasılık değerine sahip grup olarak yazmaktır. İkinci yöntem ise kesim noktaları $k/10$ ile $k = 1, 2, \dots, 9$ olarak tanımlanan $g = 10$ grup kullanmaktır. Her iki gruplama stratejisi için de, Hosmer-Lemeshow uyum iyiliği istatistiği, \hat{C} , gözlemlenen ve beklenen sıklıklardan oluşan $g \times 2$ tablosundan Pearson Ki-kare istatistiğinin hesaplanması ile elde edilir,

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n'_k \bar{P}_k)^2}{n'_k \bar{P}_k (1 - \bar{P}_k)}$$

⁹¹ P. McCullagh, **a.g.e.**, pp. 104–107.

⁹² P. McCullagh, J. A. Nelder, **a.g.e.**, pp. 112–122.

n'_k , k . grup içindeki birimlerin toplamı, $o_k = \sum_{j=1}^{c_k} y_j$ ile $\bar{P}_k = \sum_{j=1}^{c_k} \frac{\hat{P}_k}{n'_k} \cdot \hat{C}$, $(g - 2)$ serbestlik derecesi ile χ^2 dağılımına uymaktadır.

\hat{C} gibi özet uyum iyiliği ölçülerinin avantajı, tek bir değer ile uygunluğun kolayca yorumlanabilmesidir. Hosmer-Lemeshow testinin en büyük dezavantajı, gruplama sürecinde bireysel veri noktalarının az sayıda olması nedeniyle uyumda önemli bir sapmanın atlanmış olabileceğidir. Bu nedenle modelin uygunluğunu kabul etmeden önce, bireysel hataların analizinin ve tanımsal istatistiklerin yapılması uygun olur⁹³.

Bireysel hataların analizi ve tanımsal istatistikler incelenmeye, gözlem sayılarına karşı standartlaştırılmış artıkların *endeks grafiği* çizilerek, aykırı değerlerin araştırılmasıyla başlanır. Bu grafikten elde edilen değerler yanlış kodlanmış veya model eksik tanımlandığı için bazı değerler dışarıda kalmış olabilir. Çok büyük pozitif veya negatif hataların neden kaynaklandığı araştırılmadan, basit bir şekilde analizin dışarında bırakılmamalıdır. Ancak bu sorunun kaynağı, modelin doğru yapılandırıldığından emin olunması amacıyla ayrıca incelenmelidir. Büyük hatalar gözlemlerin pek de uygun olmadığını gösterirken, tahmin edilen β 'lar üzerinde büyük etkiye sahip olan veya olmayan gözlemleri göstermemektedir. Verilerin merkezine yakın olmak, her bir bağımsız değişken için bir gözlemin değerlerinin örneklemeindeki değişkenin beklenen değerine yakın olması demektir. Diğer taraftan, uç gözlemler çok büyük artık değerlere sahip olmasalar da tahminleri etkilemektedirler. Bu tür gözlemleri bularak, dışarıda bırakıldıkları durumda $\hat{\beta}$ 'lerde meydana gelen değişimin hesaplanması uygun bir yoldur⁹⁴. Örneklemeindeki her bir gözlemin tek tek dışarıda bırakılarak n kez modelin tahmin edilmesinin işlem yükü çok fazla olduğu için, Pregibon (1981), modeli tek seferde yaklaşık olarak tahminlemeyi amaçlamıştır. Eğer i . gözlem dışarıda bırakılırsa, $\hat{\beta}$ 'deki beklenen değişiklik ($\Delta_i \hat{\beta}$) yaklaşık olarak,

$$\Delta_i \hat{\beta} = \widehat{Var}(\hat{\beta}) x'_i \frac{y_i - \hat{P}_i}{1 - h_{ii}}$$

⁹³ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, pp. 147–156.

⁹⁴ J. Scott Long, **a.g.e.**, p.100.

ifadesine eşittir. Çıkartılan x_i 'ye uygun olan β_k 'daki standartlaştırılmış değişim, DFBETA olarak bilinmektedir,

$$DFBETA_{ik} = \frac{\Delta_i \hat{\beta}_k}{\sqrt{\widehat{Var}(\hat{\beta}_k)}}.$$

$DFBETA_{ik}$ büyük ise, i . gözlemin β_k 'nın tahmini üzerinde büyük bir etkiye sahip olduğunu göstermektedir⁹⁵.

Vektör $\hat{\beta}$ 'nın tamamı üzerinden çıkartılan i . gözlemin etkilerini özetleyen Doğrusal Regresyon Modelinde (Linear Regression Model-LRM) Cook's uzaklığına denk gelen ikinci ölçüm

$$C_i = (\Delta_i \hat{\beta})' \widehat{Var}(\hat{\beta}) (\Delta_i \hat{\beta}) = \frac{r_i^2 h_{ii}}{(1 - h_{ii})^2}$$

şeklinde ifade edilmektedir.

Tek bir gözlemin etkisinin diğer bir ölçüsü de i . gözlem çıkartıldığı zaman χ^2 'deki değişim değeri ile hesaplanmaktadır.⁹⁶

$$\Delta_i \chi^2 = \frac{r_i^2}{1 - h_{ii}}$$

Uygunluğun Sayısal Ölçümleri

Modelin uyum iyiliğinin tek bir değer ile özetlenmesi, değişkenlerin tek tek test edilmesinden daha iyi olabilmektedir. Bu tür ölçümler, alternatif modellerin karşılaştırılmasında yardımcı olmaktadır. Doğrusal Regresyon Modeli'nin (LRM) standart uygunluk ölçüsü, belirlilik katsayısı R^2 iken, kategorik sonuçların modelleri için kesin bir tercih yoktur. LRM'deki R^2 'nin benzerlerini oluşturmak için çok fazla girişim vardır. Ancak hiçbir ölçümün açık üstünlüğü veya açıklanan değişkenlik açısından net bir yorumlama avantajı bulunmamaktadır. Kategorik ve sınırlı bağımlı

⁹⁵ D. Pregibon, **a.g.e.**, pp. 705–724.

⁹⁶ J. Scott Long, **a.g.e.**, p.101.

değişken modelleri için uygunluğun birçok sayısal ölçüsü LRM içindeki belirlilik katsayısı R^2 'den yaklaşık olarak yapılandırılmıştır.

Kategorik bağımlı değişkenli modellerin uygunluk ölçümleri için birçok *Pseudo (Yaklaşık) – R^2* 'ler önerilmiştir. Aldrich ve Nelson (1984), *Pseudo – R^2* değerini tam model için hesaplanan Ki-kare istatistik değerini kullanarak,

$$R^2 = \frac{\chi^2}{\chi^2 + n}$$

şeklinde hesaplamayı önermişlerdir⁹⁷.

Efron's (1978) iki alternatifli sonuçlar için \hat{y} değeri yerine $\hat{P} = \widehat{Pr}(y|x)$ değeri kullanıldığında *Pseudo – R^2* değerini; LRM'de açıklanan değişkenliğin yüzdesi şeklinde hesaplanan R^2 değerine benzer şekilde kullanmıştır.

$$R_{Efron}^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{P}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

İki alternatifli sonuçlar için $\sum_{i=1}^N (y_i - \bar{y})^2 = (n_0 n_1)/n$ olarak tanımlanır. n_0 , örneklemdaki “0”ların sayısını, n_1 ise örneklemdaki “1”lerin sayısını ifade etmektedir⁹⁸.

McFadden (1973) ise En Çok Olabilirlik yaklaşımı ile tahmin edilen herhangi bir modelde uygulanabilecek LRM içinden, açıklanan değişim yüzdesine benzer farklı bir yol önermiştir. Bu popüler ölçüm aynı zamanda “olabilirlik oran indeksi” olarak da adlandırılmaktadır. Bu ölçüm içinde bağımsız değişkenler olmaksızın modellenen M_α 'nın log-olabilirlik değeri, toplam kareler toplamı olarak; bağımsız değişkenler ile modellenen M_β 'nin log-olabilirlik değeri ise hata kareler toplamı olarak düşünülebilir,

⁹⁷ John H. Aldrich, Forrest D. Nelson, **a.g.e.**, p. 57.

⁹⁸ Bradley Efron, “Regression and ANOVA with Zero-One Data: Measures of Residual Variation”, **Journal of the American Statistical Association**, Vol: 73, 1978, pp. 113–121.

$$R_{MCF}^2 = 1 - \frac{\ln \hat{L}(M_\beta)}{\ln \hat{L}(M_\alpha)}.$$

Eğer $M_\alpha = M_\beta$ ise R_{MCF}^2 değeri “0”a eşittir. Fakat R_{MCF}^2 değeri asla “1” olmaz. Yeni değişkenler modele ilave edildikçe R_{MCF}^2 değeri artmaktadır.

Bunu düzenlemek için Ben-Akiva ve Lerman, modeldeki parametre sayısı için düzeltilmiş R_{MCF}^2 , \bar{R}_{MCF}^2 ’yi önermiştir.

$$\bar{R}_{MCF}^2 = 1 - \frac{\ln \hat{L}(M_\beta) - K}{\ln \hat{L}(M_\alpha)}$$

Modele ilave edilen her parametre için, $\ln \hat{L}(M_\beta)$ değeri “1”den daha büyük olduğu zaman \bar{R}_{MCF}^2 değeri artmaktadır⁹⁹. Cramer (1999) da Ben-Akiva ve Lerman tarafından dikkate alınmayan dengesiz (sonuç değişkenin aldığı alternatif değerlerin sayısı bakımından) örneklemde; genellikle sıklığı az olan bağımlı değişkeninin çok kötü tahmin edildiği durumları göz önüne alarak düzeltilmiş ölçüm önermiştir¹⁰⁰.

LRM’de belirlilik katsayısını hesaplamanın başka bir yolu da parametre değerlerinin En Çok Olabilirlikle hesaplandığı model için Olabilirlik Oran dönüşümünü kullanmaktır. Buradan yola çıkarak M_α ’yı sadece sabiti olan model ve M_β ’yı da sabit ve bağımsız değişken katsayılarını içeren model olarak tanımlayıp *Pseudo – R²* şu şekilde hesaplanabilir:

$$R_{ML}^2 = 1 - \left[\frac{L(M_\alpha)}{L(M_\beta)} \right]^{2/N}$$

Maddala (1983), yukarıdaki eşitliği $G^2 = -2\ln[L(M_\alpha)/L(M_\beta)]$ değerini alacak şekilde aşağıdaki gibi ifade etmiştir,¹⁰¹

⁹⁹ J. Scott Long, **a.g.e.**, p. 104.

¹⁰⁰ J. Cramer, “Predictive Performance of the Binary Logit Model in Unbalanced Samples”, **Journal of the Royal Statistical Society**, Series D, Vol: 88, 1990, pp. 85–94.

¹⁰¹ G. S. Maddala, **Limited-Dependent and Qualitative Variables in Econometrics**, Cambridge University Press, 1983, p.39.

$$R_{ML}^2 = 1 - \exp(-G^2/N)$$

Diğer bir ifadeyle, uygunluk ölçüsü olarak gösterilen R^2 ve çeşitli *Pseudo* – R^2 'ler genellikle hipotez testleri ile yakından ilişkilidir. Magee (1990), Wald ve skor testlerine dayanan uygunluk ölçülerini,¹⁰²

$$R^2 = \frac{W}{n + W} \equiv R_W^2, \quad R^2 = 1 - \exp(-LR/n) \equiv R_{LR}^2$$

olarak bulmuştur. McFadden'ın haricinde, hesaplanan *Pseudo* – R^2 'yi uygulamalarında kullanan bulunmamaktadır. Bunun nedeni, *Pseudo* – R^2 'nin “0” ve “1” dışındaki değerleri için açık bir yorumlanma imkanı olmamasıdır. Bu değer sadece belirlilik katsayısı büyüdükçe modelin uygunluğunun gelişmekte olduğunu göstermektedir.

Sınıflama Tabloları

Kategorik sonuçların gözlemlenen ve beklenen değerlerini karşılaştırmak, modellerin uyum iyiliğini değerlendirmede diğer bir yaklaşımdır. Gözlemlenen ve beklenen değerlere dayanan sınıflandırma tablolarında doğru tahmin oranları “*Count (Sayılan) – R²*” ile ifade edilmektedir.

$$R_{Count}^2 = \frac{1}{N} \sum_j n_{jj}$$

En büyük satır marjına göre R_{Count}^2 değerinin

$$R_{AdjCount}^2 = \frac{\sum_j n_{jj} - \max_r(n_{r+})}{N - \max_r(n_{r+})}$$

şekilde düzeltilmesi gerekmektedir. n_{r+} , r . satırın marjinal değerini; $\max_r(n_{r+})$ ise satır marjı en büyük olanı belirtmektedir¹⁰³.

¹⁰² Lonnie Magee, “ R^2 Measures Based on Wald and Likelihood Ratio Joint Significance Tests”, **The American Statistician**, Vol: 44, 1990, pp. 250–253.

¹⁰³ J. Scott Long, **a.g.e.**, p. 106–108.

Bilgi Ölçüleri

Bilgi ölçüleri, İç İçe Modellerin karşılaştırılmasında kullanışlı olan ve en çok bilinen Akaike Bilgi Kriteri (Akaike Information Criteria-AIC) ile popüleritesi artmakta olan Bayes Bilgi Kriteridir (Bayesian Information Criteria-BIC).

Akaike 1973 yılında bilgi kriterini şu şekilde tanımlamıştır:¹⁰⁴

$$AIC = \frac{-2\ln\hat{L}(M_\beta) + 2P}{n}$$

P ; modeldeki parametrelerin sayısı; $\hat{L}(M_\beta)$, verilerin modeldeki olabilirliklerini gösterir ki, büyük değer daha iyi uyum anlamına gelmektedir. $-2\ln\hat{L}(M_\beta)$ $[0, +\infty)$ arasında değer almakta ve küçük değerler daha iyi uyum olduğunu göstermektedir. Modeldeki parametre sayıları arttığında, $-2\ln\hat{L}(M_\beta)$ küçülür. AIC'ye $2P$ 'nin ilave edilmesi parametre sayısındaki artış için bir ceza değeridir. Gözlemlerin sayısı $-2\ln\hat{L}(M_\beta)$ değerini etkilediği için düzeltilmiş $-2\ln\hat{L}(M_\beta)$, her bir gözlemin katkısını elde etmek amacıyla örnek büyüklüğüne bölünmektedir.

AIC genellikle, farklı örnek büyüklüklerini veya Olabilirlik Oran (LR) testleri ile karşılaştırılmayacak, İç İçe Modelleri karşılaştırmada kullanılmaktadır. Diğer her şey aynı iken, küçük AIC değerli model en uygun model olarak kabul edilmektedir¹⁰⁵.

Bayes Bilgi Kriteri (BIC) 1995 yılında Raftery tarafından bir modelin uygunluğunu kapsamlı bir şekilde değerlendirmek ve hem İç İçe (Nested) hem de İç İçe Olmayan (Nonnested) Modellerin karşılaştırılmasına olanak vermek amacıyla önerilmiştir.

BIC istatistiği, Bayes faktör tahmini kullanılarak hesaplanması uygun bir yaklaşımdır. BIC ölçüsü, model M_1 model M_2 içinde yer alıyorsa yani İç İçe Modeller ise şu şekilde hesaplanmaktadır,

¹⁰⁴ H. Akaike, "Information Theory and an Extension of the Maximum Likelihood Principle", **Perspective in Statistics Selected Paper of Hirotugu Akaike**, Springer, 1998, p. 209.

¹⁰⁵ J. Scott Long, **a.g.e.**, p. 109.

$$BIC_{12} = \chi_{12}^2 - df_{12} \log n$$

Model M_1 'e göre model M_2 karşılaştırılırken hesaplanan standart Olabilirlik Oran test istatistik değeri χ_{12}^2 , $df_{12} = df_2 - df_1$ ile testin serbestlik derecesine eşittir.

Genellikle modeller karşılaştırılırken, temel model olarak M_D ile gösterilen Doymuş (Saturated) Model ya da sadece sabit terimin yer aldığı M_α modeli kullanılmaktadır.

Temel model olarak Doymuş Model (M_D) alınır, model M_k için hesaplanan BIC değeri BIC_k ile gösterilmektedir ve

$$BIC_k = D_k - df_k \log n$$

ifadesine eşittir. Burada D_k , model M_k için sapma değeri; df_k ise sapma istatistiğine ait serbestlik derecesidir. Doymuş Model'e ait BIC değeri "0" a eşit olduğu için, $BIC_k > 0$ olduğunda Doymuş Model kullanılmaktadır. $BIC_k < 0$ olduğunda ise, veri seti için daha uygun olan, model M_k tercih edilmektedir.

Temel model olarak sadece sabitin yer aldığı M_α modeli alındığında ise, BIC yerine BIC'_k kullanılır. LR(M_k)'nin diğer hesaplama yolunda LR'deki df'_k , modeldeki bağımsız değişken sayısı ile Ki-kare dağılımına uymaktadır.¹⁰⁶

$$BIC'_k = -LR(M_k) - df'_k \log n$$

LR(M_k), sadece sabit değer α 'nın olduğu model M_α 'ya karşı model M_k 'nin Olabilirlik Oran test istatistik değeridir. df'_k , model M_k 'daki bağımsız değişken sayısıdır. $BIC'_\alpha = 0$ yokluk hipotezi altında, $BIC'_k > 0$ olduğu zaman M_α modeli, $BIC'_k < 0$ olduğu zaman ise M_k modeli geçerli olmaktadır.

Bayes Bilgi Kriteri olarak hesaplanan BIC veya BIC' değerlerinden hangisinin kullanılacağına, model M_α 'ya karşı diğer model karşılaştırılırken, "sapma istatistik değeri"nin mi yoksa "olabilirlik oran test istatistik değeri"nin mi hesaplandığına göre karar verilmektedir. Eğer sapma değeri kullanılıyorsa BIC, olabilirlik oran değeri kullanılıyorsa BIC' değeri kullanılmalıdır.

¹⁰⁶ J. Scott Long, a.g.e., p. 111.

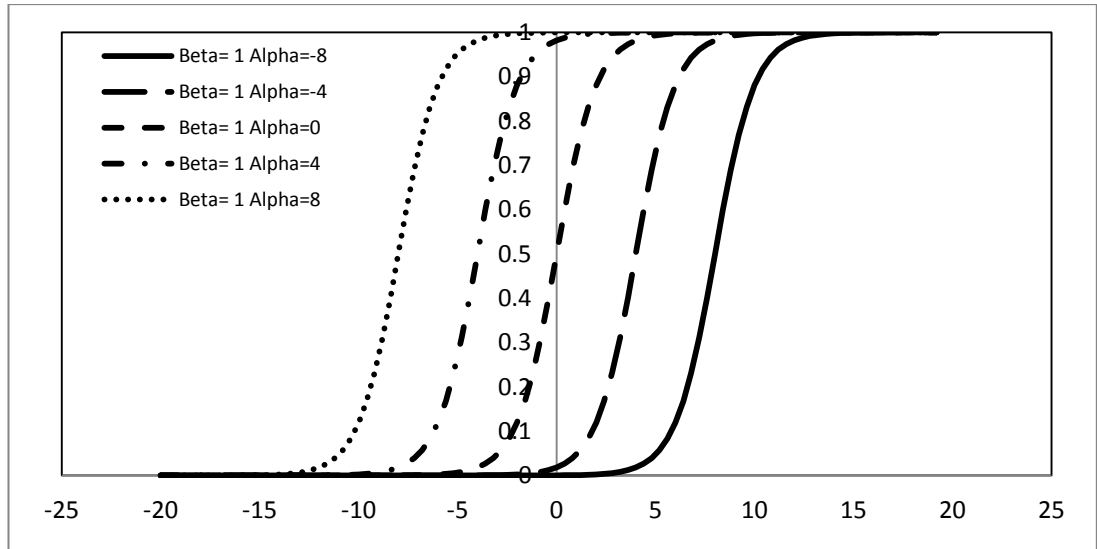
M_j ve M_k gibi herhangi iki model karşılaştırılırken, BIC_{jk} değeri iki modelin BIC değerlerinin farkı alınarak hesaplanmaktadır.

$$BIC_{jk} = BIC_k - BIC_j$$

Model seçiminde BIC değeri küçük olan model; yani $BIC_k - BIC_j < 0$ ise model M_k , $BIC_k - BIC_j > 0$ ise model M_j tercih edilmektedir.¹⁰⁷

1.2.6 Model Parametrelerinin Yorumu

Lojistik regresyon eğrisi S şeklindedir. Ancak “sabit” ve “eğim (bağımsız değişken)” katsayılarının farklı değer çiftleri için farklı lojistik regresyon eğrileri oluşmaktadır. Grafik olarak modeldeki sabit ve bağımsız değişkenlere ait parametre katsayılarının, olasılık değeri üzerindeki etkisini görmek için Lojit Modeli tek bağımsız değişkenli $P(y = 1|x) = F(\alpha + \beta x)$ olacak şekilde ifade ettiğinde, olasılık eğrisi üzerindeki “sabit”in etkisi Şekil 1.1’de görülmektedir. Burada eğim katsayısı β ’ya sabit bir değer verildiğinde, α değeri büyüdükçe lojistik regresyon eğrisinin sola, küçüldükçe de sağa doğru yer değiştirdiği görülmektedir.

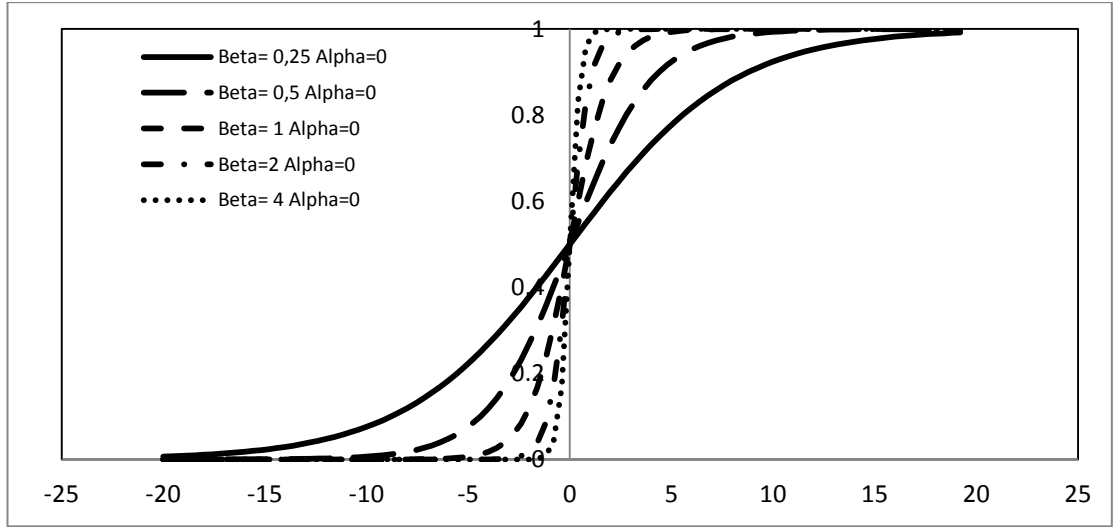


Şekil 1.1 Modeldeki “sabit” in olasılık üzerindeki etkisi.¹⁰⁸

¹⁰⁷ Adrian E. Raftery, “Bayesian Model Selection Social Research”, *Sociological Methodoloji*, Vol: 25, 1995, pp. 111–165.

¹⁰⁸ J. Scott Long, a.g.e., pp. 62–63.

Şekil 1.2’de ise modeldeki eğim katsayısındaki değişimin etkisi görülmektedir. Modelde α ’nın olmadığı durumlar için, β ’nin aldığı değer küçüldükçe lojistik regresyon eğrisi esnemektedir. Örneğin, $\beta = 0.25$ olduğunda, x ’in -20 ile +20 arasında muntazam bir şekilde arttığı görülürken, $\beta = 0.5$ için eğrideki değişimin hızı başlarda daha yavaşken x değeri “0”a yaklaştıkça eğimdeki değişimin hızının arttığı gözlemlenmektedir.¹⁰⁹



Şekil 1.2 Modeldeki “eğim” katsayısının olasılık eğrisi üzerindeki etkisi.¹¹⁰

İki Kategorili Lojistik Regresyon Modelinde birden fazla bağımsız değişkenin olması bu değişkenlere ait parametrelerin ilgilenilen olayın olasılığı üzerindeki etkilerini grafikte görmeyi güçleştirecektir.

Doğrusal Olasılık Modelinde, “diğer bağımsız değişkenlerin değeri sabit¹¹¹ olduğu varsayımı altında”* x_k ’daki bir birimlik değişimin $P(y = 1)$ olasılığı üzerindeki etkisi $\hat{\beta}_k$ kadar olacaktır ve x_k ’nın tüm değerleri için bu etki sabittir. Lojit Model’de $P(y = 1)$ ile her bir x_k arasındaki doğrusal olmayan ilişki nedeniyle, $P(y = 1)$

¹⁰⁹ J. Scott Long, **a.g.e.**, p. 63.

¹¹⁰ **A.e.**, p. 63.

¹¹¹ J. Scott Long, Simon Cheng, **a.g.e.**, p. 260.

* Modeldeki parametre katsayılarının yorumunda, herhangi bir bağımsız değişkenindeki değişimin olasılık üzerindeki etkisinden bahsedildiğinde, her zaman diğer değişkenlerin modele sabit bir değer olarak alındığı kabul edilmektedir. Sabit değerleri elde etmenin iki yolu vardır; ya ilgilenilen değişkenler dışındaki bağımsız değişkenlerin değerlerinin kendi ortalamaları modele sabit değer olarak alınmakta veya doğrudan her bir gözlemin ilgilenilen değişken dışındaki diğer değişkenlerin değerleri tek tek modelde yerine konularak marjinal değişimler hesaplanıp sonra tüm gözlemler için ortalaması hesaplanmaktadır.

üzerindeki x_k 'daki değişimin etkisini yorumlamak çok kolay olmamaktadır. Bu etkiyi yorumlamanın bir yolu, eğer ilgilenilen x_k bağımsız değişkeni nicel ise bu değişkendeki değişimin, $P(y = 1)$ olasılığı üzerindeki değişim oranını ölçmek için kısmi türevini almaktır.

$$\begin{aligned}\frac{\partial P(y = 1|x)}{\partial x_k} &= \frac{\exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)}{1 + \exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)} \frac{1}{1 + \exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)} \hat{\beta}_k \\ &= P(y = 1|x)[1 - P(y = 1|x)]\hat{\beta}_k\end{aligned}$$

Yukarıdaki formülde $\hat{\beta}_k$ değeri çarpımsal faktör olarak eşitlikte yer almaktadır ve $\hat{\beta}_k$ 'nin işareti x_k 'daki değişimin $P(y = 1)$ üzerindeki etkisinin yönünü tayin etmektedir. Etkinin büyüklüğü ise x_k 'ların tümünü içine alan $(\beta_0 + \sum_{k=1}^p \beta_k x_k)$ değerinin büyüklüğüne bağlı olmaktadır. Doğrusal olmayan $(\beta_0 + \sum_{k=1}^p \beta_k x_k)$ değeri x_k 'daki değişimin $P(y = 1)$ üzerindeki etkisinin değerini düşürmektedir. $(\beta_0 + \sum_{k=1}^p \beta_k x_k) = 0$ iken etkinin değerindeki düşüş minimumdur. Bu değer “0”dan (-) veya (+) tarafa uzaklaştıkça etkinin değerindeki düşüş de artmaktadır.¹¹²

x_k bağımsız değişkeni nitel ve “0” ile “1” olmak üzere iki değer alan bir kukla değişken ise; x_k 'daki değişimin $P(y = 1|x)$ üzerindeki etkisi şu şekilde hesaplanmaktadır.

$$\frac{\Delta P(y = 1|x)}{\Delta x_k} = P(y = 1|x, x_k = 1) - P(y = 1|x, x_k = 0)$$

Parametre katsayılarının yorumlanmasında bir başka yol da, önce bağımlı ve bağımsız değişkenler arasındaki fonksiyonel ilişkinin tanımlanarak bağ fonksiyonunun oluşturulması ve sonra da her bir bağımsız değişkendeki bir birimlik değişikliğin bu tanımlanan bağ fonksiyonu üzerindeki etkisinin uygun bir şekilde ifade edilmesidir¹¹³. Lojistik Regresyon Modeli şu şekilde ifade edilmekteydi,

¹¹² John H. Aldrich, Forrest D. Nelson, **a.g.e.**, pp.41–44.

¹¹³ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, pp. 47–48.

$$P(y = 1|x) = \frac{\exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)}{1 + \exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)}$$

Bu ifade ile Doğrusal Olmayan Olasılık Modeli'nde, olasılık değerleri odds şekline dönüştürülebilmektedir. Lojistik regresyon katsayılarının düzgün yorumlanabilmesi için “odds” kavramının iyi anlaşılması gerekmektedir. İki Kategorili Lojistik Regresyon Modeli için odds değerini $\Omega(x)$ ile gösterildiğinde,

$$\Omega(x) = \frac{P(y = 1|x)}{1 - P(y = 1|x)} = e^{\beta_0 + \sum_{k=1}^p \beta_k x_k} = e^{\beta_0} \prod_{k=1}^p e^{\beta_k x_k}$$

eşitliğin sol tarafı odds, sağ tarafı da $\exp(\beta_k)$ üsteli ile odds üzerindeki x_k 'nin marjinal etkisini vermektedir. Odds değeri olasılık değerinden farklıdır. Olasılık [0, 1] aralığında değerler alırken, odds değeri $[0, +\infty)$ aralığında değerler almaktadır.¹¹⁴

Lojistik Regresyon Modeli'nde katsayıların yorumlanmasında diğer önemli bir kavram da Odds Oranı (Odds Ratio-OR)'dir. OR, yorumlamadaki kolaylığı nedeniyle Lojistik regresyonda genellikle ilgilenilen bir parametredir. Odds değeri $\Omega(x, x_k) = \exp(x_k \beta_k) = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \dots + \beta_p x_p}$ olarak tanımlandığında OR değeri, örneklemdaki herhangi bir gözlemin, modeldeki diğer değişkenlerin sabit olduğu varsayımı altında, bağımsız değişken x_k 'daki δ birimlik artışın ilgilenilen olayın ortaya çıkmasını $\exp(\beta_k \delta)$ kadar daha mümkün kıldığını ifade etmektedir.¹¹⁵

$$OR = \frac{\Omega(x, x_k + \delta)}{\Omega(x, x_k)} = \exp(\beta_k \delta)$$

Buradaki sorun, OR tahmininin (\widehat{OR}) çarpık bir dağılıma sahip olmasıdır. Bu sorunun kaynağı, OR parametre değerinin “1”e eşitliği test edilirken, odds değerlerinin $[0, +\infty)$ arasında değerler alıyor olmasıdır. Teoride yeterince büyük örneklem için \widehat{OR} 'nin dağılımı normaldir, ancak birçok çalışma, \widehat{O} değerlerinin Normal dağılıma uygun olmasını sağlayacak örnek büyüklüğüne sahip değildir. Bu nedenle çıkarımlar, genellikle çok daha küçük olan örneklem büyüklüğü için Normal

¹¹⁴ Raymond E. Wright, **a.g.e.**, pp. 222–224.

¹¹⁵ J. Scott Long, **a.g.e.**, p.80.

dağılıma sahip olma eğiliminde olan $\ln(\widehat{OR})$ 'nin örneklem dağılımına dayanmaktadır.¹¹⁶ Buradan OR parametre değeri için nokta tahmini yapılabildiği gibi $\ln(\widehat{OR}) = \hat{\beta}_k$ için $\%(1 - \alpha)$ düzeyinde uç değer tahmini de yapılabilmektedir.

$$\exp[\hat{\beta}_k \pm z_{1-\alpha/2} \widehat{SE}(\hat{\beta}_k)]$$

Bu aralık “1” değerini içermiyorsa aralık değerlerine bakılıp örnekleme ait OR 'nin sağa ya da sola çarpık olduğuna karar verilir.

1.3 İki Kategorili Probit Analizi

Probit Analizi olayın olasılıklarını Lojistik dağılımdaki Lojit yerine, Birikimli Normal dağılımda skorlara dönüştürmektir. Literatürdeki birçok çalışma Lojit ve Probit analizinin benzerliklerini vurgulamaktadır. Probit dönüşümü “0” ve “1” uç değerlerinden oluşan olasılıkları sınırlar olmaksızın skorlara dönüştürmek için Lojistik dağılım yerine Birikimli Standart Normal dağılımın değerlerini kullanmaktadır. Birikimli Standart Normal dağılım eğrisi Lojistik dağılım eğrisinden farklı olarak “0” ve “1” değerlerini alması¹¹⁷ dışında benzer dağılımlardır. Başka bir ifadeyle, Lojistik Birikimli dağılım fonksiyonu uç bölgelerde Probit Model'e göre daha geniştir.

1.3.1 Probit Modelinin Kurulması

Lojistik regresyonda olduğu gibi, Probit Analiz de, iki kategorili bağımlı değişken üzerindeki ilişkiyi sürekli değişken üzerindeki ilişkiye benzer yapmak için, dönüşüme dayanmaktadır. Tahmin edilen Probit, bir veya daha fazla bağımsız değişkenle tanımlanan doğrusal eşitlikteki bağımlı değişkendir. Probit Model aşağıdaki gibi ifade edilmektedir.

$$P(y = 1|x) = 1 - \Phi\left(-\sum_{k=1}^p \beta_k x_k\right) = \Phi\left(\sum_{k=1}^p \beta_k x_k\right)$$

¹¹⁶ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, p. 52.

¹¹⁷ John E. Freund, **a.g.e.**, p. 54.

$$= \int_{-\infty}^{\sum \beta_k x_k} \frac{1}{\sqrt{2\pi}} \exp - \left(\frac{u^2}{2} \right) du$$

Burada $\sum_{k=1}^p \beta_k x_k \equiv Z$, Birikimli Standart Normal dağılım kullanılarak elde edilen Z değerlerinin doğrusal olmayan olasılıklarıdır¹¹⁸. Probit Model; parametreleri bakımından doğrusal, olasılık değerleri bakımından doğrusal olmayan bir istatistik modeldir. Probit Analiz, Standart Normal dağılım eğrisinin formülünün karmaşıklığı nedeni ile daha zordur. Probit Analiz için doğrusal olmayan eşitlikte $P(y = 1|x)$ olasılık değeri, Birikimli Standart Normal dağılımın formülündeki Z fonksiyonu gibidir.

Birikimli Standart Normal dağılıma dayanan, herhangi Z değeri ile ilişkili birikimli olasılıklar;

$$P(y = 1|x) = \int_{-\infty}^Z \frac{1}{\sqrt{2\pi}} \exp - \left(\frac{u^2}{2} \right) du$$

$u \sim N(0,1)$ ile standartlaştırılmış normal değişkendir. Formülün amacının, Probit eşitliğindeki doğrusal olarak tanımlanan Z 'yi doğrusal olmayan şekilde tanımlanan olasılık değerlerine geri dönüştürmek olduğu unutulmamalıdır.¹¹⁹

Lojistik regresyonda, $lojit = \log\left(\frac{P(y=1|x)}{1-P(y=1|x)}\right)$ dönüşümü ile bağımsız değişkenlerle olasılıklar arasındaki ilişkiyi doğrusal formda gösterilebildiği ilgili bölümde belirtilmiştir. Probit Analizi'nde de Birikimli Standart Normal dağılımının tersi alınarak fonksiyonun doğrusal şekli elde edilmektedir. Birikimli Standart Normal dağılımı Φ ile gösterildiğinde $P(y = 1|x) = \Phi(Z)$,

$$\sum_{k=1}^p \beta_k x_k \equiv Z = \Phi^{-1}(P)$$

¹¹⁸Tim Futing Liao, **a.g.e.**, p. 21.

¹¹⁹Fred C. Pampel, **a.g.e.**, p. 59.

eşitliği elde edilmekte ve bu Probit Regresyon Modeli olarak adlandırılmaktadır. Basit bir formül ile temsil edilememesine rağmen Birikimli Standart Normal dağılımın olasılık değerlerinin tersi, Probit Analizi'nde bağımlı değişkeni gösteren doğrusal Z değerlerini vermektedir. Bağımlı değişkene probit dönüşümü uygulanarak elde edilen katsayıların tahmin değerleri, olasılıklardaki değişimden ziyade Birikimli Standart Normal dağılımın tersinden elde edilen Z değerlerindeki değişimi göstermektedirler.

İki Kategorili Probit Modeller için varsayımlar şunlardır:¹²⁰

- y_i bağımlı değişkeni örneklemdaki tüm gözlemler için ($i = 1, 2, \dots, n$) "0" veya "1" değerini almalıdırlar.
- Örneklemdaki birimler Birikimli Standart Normal dağılım fonksiyonu şeklinde ($P(y_i = 1|x_i) = \Phi(\sum_{k=1}^p \beta_k x_{ik})$) ifade edilmelidirler.
- $y_1, y_2 \dots, y_n$ örneklemdaki her bir birime ait bağımlı değişkenin değeri istatistik olarak bağımsız olmalıdır.
- Tüm x_{ik} 'ler arasında tam veya yaklaşık doğrusal bağımlılık olmamalıdır.

En Çok Olabilirlik Tahmin Yöntemi

Parametre değerlerinin tahmininde Probit Analizi de, Lojistik Regresyon Analizi'nde olduğu gibi En Çok Olabilirlik Tahmin tekniğini kullanmaktadır. Daha önce tanımlandığı gibi buradaki amaç probit olabilirlik değerini maksimum yapacak parametre değerlerini tahmin etmektir. Lojistik regresyondan farkı ise, bağımsız değişkenler ve parametre tahminleri için olasılık değerlerini Lojistik dağılım yerine Birikimli Standart Normal dağılımı kullanarak elde etmesidir.¹²¹ Probit olabilirlik fonksiyonu aşağıdaki gibi ifade edilmektedir.

$$L(y|x, \beta) = \prod_{i=1}^n [\Phi(\beta_k x_{ik})]^{y_i} [1 - \Phi(\beta_k x_{ik})]^{1-y_i}$$

¹²⁰ John H. Aldrich, Forrest D. Nelson, **a.g.e.**, p. 49.

¹²¹ Alan Agresti, **a.g.e.**, pp. 78–79.

Çarpımlar yerine toplamlarla çalışmak daha kolay olduğu için yukarıdaki ifadenin doğal logaritmasını alarak işlemlere devam etmek daha uygundur.

$$\sum_{i=1}^n \frac{[y_i - \Phi(\beta_k x_{ik})]\phi(\beta_k x_{ik})}{\Phi(\beta_k x_{ik})[1 - \Phi(\beta_k x_{ik})]} x_{ij} = 0 \quad j = 1, 2, \dots, p$$

$\phi(\beta_k x_{ik})$, Standart Normal dağılım için olasılık yoğunluk fonksiyonudur.¹²² “ p ” (parametre sayısı) denklemin iterasyonla çözümüyle model parametre tahminleri elde edilmektedir.

1.3.2 Tahmin Parametrelerinin ve Modelin Uyum İyiliği Testleri

Katsayıların tek tek testinde $\beta_k = 0$ sıfır hipotezi altında t-test istatistiği kullanılmaktadır.

$$t = \frac{\hat{\beta}_k - \beta^*}{\hat{\sigma}_{\hat{\beta}_k}}$$

Bu test istatistiği $(n - p - 1)$ serbestlik derecesi ile Student’s t dağılımına uymaktadır. Modelin uygunluk testi olarak Olabilirlik Oran testi kullanılmaktadır.

$$LR = -2\ln(M_\alpha/M_\beta) = 2\ln L(M_\beta) - 2\ln L(M_\alpha)$$

$L(M_\beta)$, tam model veya uygunluğu test edilen modelin olabilirlik fonksiyon değeri ve $L(M_\alpha)$, sabit hariç tüm parametre katsayılarının sıfır olduğu temel olabilirlik fonksiyonunun maksimum değeridir. χ^2 dağılımına uyan bu test istatistiği tüm bağımsız değişkenlerin katsayıları sıfıra eşittir sıfır hipotezini $p - 1$ (test edilen katsayı sayısı) serbestlik dereceli χ^2 değeri ile test etmektedir. Model uygunluğunu test etmek için ayrıca log olabilirlik değerinin kullanıldığı birkaç farklı *Pseudo - R²* katsayısı hesaplanmaktadır¹²³. Bu katsayılar ve bunların model üzerindeki etkilerinin testleri Başlık 1.2.5 altında incelenen “Uygunluğun Sayısal Ölçümleri” ile aynıdır.

¹²² John H. Aldrich, Forrest D. Nelson, **a.g.e.**, pp. 51–52.

¹²³ Fred C. Pampel, **a.g.e.**, p. 65.

Bunlara ek olarak 1984 yılında Aldrich ve Nelson, işlem kolaylığı ve $[0, 1]$ arasında değer alması nedeni ile aşağıdaki *Pseudo R²*'yi önermişlerdir.¹²⁴

$$Pseudo - R^2 = \frac{LR}{LR + n}$$

1.3.3 Model Parametrelerinin Yorumu

Probit Model içinde x_k 'daki değişimin $P(y = 1|x)$ üzerindeki etkisini yorumlamak Lojistik Regresyon Modeli'nde olduğu gibi çok kolay değildir.

Probit Model'de x_k bağımsız değişkenindeki değişimin $P(y = 1|x)$ olasılığı üzerindeki etkisini veren $\hat{\beta}_k$ değeri aynı zamanda Z değeri üzerindeki etkisini de ifade etmektedir. Yani x_k bağımsız değişkenindeki bir birimlik değişimin etkisi beklenen Z değerlerinin farklarından elde edilebilmektedir.

Probit eşitliğindeki bağımsız değişkenlere ait görece büyüklüğün etkisini göstermek için basit hesaplamalar yapılabilir. İlgilenilen bağımsız değişkenlerin her birinin standart sapmaları probit katsayıları ile çarpılır ve böylece; bağımsız değişkendeki bir standart sapmalık artış, bağımlı değişkenin Birikimli Standart Normal dağılım olasılık değerinin tersi kadar değişimi göstermektedir. Bunun dışında, x 'lerin standart sapmalarını y 'nin standart sapmasına oranlayıp tahmin edilen parametre katsayıları ile çarparak probit katsayılarının standartlaştırılmış değerleri elde edilebilmektedir.¹²⁵

$$\beta_{yx}^* = \hat{\beta}_{yx}(s_x/s_y)$$

Probit katsayılarının standartlaştırılmaları için kullanılan diğer yöntemler Lojistik regresyonda kullanılan yöntemlere benzerlik göstermektedir. Probit Model de aynı Lojistik Regresyon Modeli'nde olduğu gibi parametre değerlerinin yorumlanmasında nicel bağımsız değişkenler için kısmi türevleri, nitel bağımsız değişkenler için kesikli

¹²⁴ John H. Aldrich, Forrest D. Nelson, **a.g.e.**, pp. 54–59.

¹²⁵ Fred C. Pampel, **a.g.e.**, pp. 60–61.

değişimi kullanmaktadır. Buradaki fark $P(y = 1|x)$ 'nin x_k değerine göre kısmi türevinin ifade edilme şeklindedir¹²⁶.

$$\frac{\partial P(y = 1|x)}{\partial x_k} = \frac{1}{\sqrt{2\pi}} \exp \left[- \left(\sum_{k=1}^p \beta_k x_k \right)^2 / 2 \right] \hat{\beta}_k \equiv \phi \left(\sum_{k=1}^p \beta_k x_k \right) \hat{\beta}_k$$

Bu eşitlikte $\hat{\beta}_k$, etkinin yönünü vermekte ve etkinin büyüklüğü $\phi(\cdot)$ değerine bağlı olarak değişmektedir. $(\sum_{k=1}^p \beta_k x_k)$ değerinin “0”a yakın olduğu yerde, olasılıklar üzerindeki anlık etki değerindeki düşüş minimum iken, “0”dan uzaklaştıkça Normal dağılımdaki yoğunluğu azalacaktır ve bu da $\hat{\beta}_k$ 'nin, olasılık üzerindeki anlık etkisini azaltacaktır.¹²⁷

Olasılık değerinin x_k değişkenine göre kısmi türevinin alınması nitel bağımsız değişkenler için bir anlam ifade etmemektedir. Bir nitel değişken için olasılıklardaki değişimi hesaplamak amacıyla bağımlı değişkenin ortalamasını (beklenen değerini), dışarıda bırakılan grubun tahmin olasılığı gibi ele almak gerekmektedir. Bu ortalama değer, Birikimli Standart Normal dağılım tablosu kullanılarak Z değerine dönüştürülür. Daha sonra nitel değişkenin probit katsayısına bu Z değeri eklenerek elde edilen yeni Z değerinin Standart Normal dağılım tablosundan yeni olasılık değerleri bulunarak, bu değer ortalamadan çıkartılıp iki grubun tahmin olasılıkları arasındaki fark hesaplanabilmektedir.¹²⁸ İki değer alan bir nitel değişken için x_k 'daki değişimin $P(y = 1|x)$ üzerindeki etkisi şu şekilde gösterilmektedir.

$$\frac{\Delta P(y = 1|x)}{\Delta x_k} = P(y = 1|x, x_k = 1) - P(y = 1|x, x_k = 0)$$

Sonuç olarak Probit Analizi, bağımlı değişkenin kategorik olduğu durumlarda dağılım problemini, Birikimli Standart Normal dağılıma dayanan dönüşümle çözmektedir. Probit Analizi'nde OR'nin hesaplanması mümkün değildir ve olasılıklardaki değişimin hesaplanması Lojistik regresyondan daha zordur. Birçok durum için, Lojistik Regresyon Analizi tercih edilmektedir. Ancak alternatif olarak

¹²⁶ J. Scott Long, **a.g.e.**, p. 72.

¹²⁷ John H. Aldrich, Forrest D. Nelson, **a.g.e.**, pp. 41–44.

¹²⁸ J. Scott Long, **a.g.e.**, p. 78.

Probit Analizi'nin gereklerini tartışmak, bağımlı değişkenlerin kategorik olduğu durumların analiz stratejilerini anlamada daha faydalı olmaktadır.¹²⁹

1.4 Çok Kategorili Lojistik Regresyon Modelleri

Sosyal bilimlerde birçok model, Çoklu Regresyon veya İkili Lojit ve Probit modeller kullanılarak tahminlenebilmekle beraber, bağımlı değişkenin doğası gereği modellerin birçoğunu tahminleyebilmek için bu teknikler yeterli olmamaktadır. Bağımlı değişken ikiden fazla sıklı nitel olduğu durumlar için model tahminlerinde, İki Kategorili Lojistik Regresyon Model'in uzantısı olan Çok Kategorili Lojistik Regresyon model teknikleri kullanılarak yapılması gerekmektedir. Çok Kategorili Lojistik Regresyon Modeller kendi içinde, bağımlı değişkenin çok terimli (multinomial) veya sıralı (ordinal) olmasına göre farklı şekilde tahminlenmekte ve analizleri yapılmaktadır.¹³⁰

1.4.1 Çok Terimli Lojistik Regresyon Modelleri

Bağımlı değişkenin ikiden fazla kategorisi olduğu durumlar için 1974'de McFadden Lojistik Regresyon Modelini uyarlayarak “Kesikli Tercih Modelini (Discrete Choice Model)” önermiştir. Ekonometri literatüründe bu isimle adlandırılırken, sosyal ve sağlık bilimlerinde *Çok Terimli Lojit (Multinomial Logit - MNL)* veya *Çoklu Grup¹³¹ Lojistik (Polytomous Logistic)* model olarak adlandırılmaktadır.¹³²

Çok Terimli Lojistik Modeller, bağımlı değişken kategorilerinin tüm ikili lojitlerinin eş zamanlı tahminleri olarak adlandırılmaktadırlar. Modelin “yeterli (efficient)” tahmini tüm mümkün çiftlerin aynı anda tahmin edilmesini gerektirmektedir.¹³³

Bağımlı değişkenin, P_1, P_2, \dots, P_m olasılıkları ile m tane kategorik alternatiften oluştuğu varsayıldığında (burada alternatiflere verilen sayıların sıra değerleri yoktur),

¹²⁹ Fred C. Pampel, **a.g.e.**, p. 68.

¹³⁰ Scott Menard, **Applied Logistic Regression Analysis**, 2nd Edition, Sage Publications, 2002, p. 91.

¹³¹ Hüseyin Tatlıdil, **a.g.e.**, s. 304.

¹³² David W. Hosmer, Stanley Lemeshow, **a.g.e.**, p. 260.

¹³³ Simon Cheng, J. Scott Long, “Testing for IIA in the Multinomial Logit Model”, **Sociological Methods & Research**, Vol:35, Number 4, May 2007, pp. 583–600.

ikiden fazla alternatifi olan bağımlı değişkenler için $m - 1$ tane doğrusal ilişkiyi gösteren Lojistik Model şu şekilde ifade edilmektedir

$$\log \left[\frac{P(y = j)}{P(y = m)} \right] = \beta_{0j} + \beta_{1j}x_1 + \beta_{2j}x_2 + \cdots + \beta_{pj}x_p \quad j = 1, 2, \dots, m - 1 \quad 1.24$$

Yukarıdaki fonksiyon kümesinden de açıkça görülebileceği gibi Çok Terimli Lojistik Modeller'de m alternatif için her bir lojite ait olacak şekilde, $m - 1$ tane regresyon katsayısı kümesi bulunmaktadır.

Buradan, ikiden fazla şıklı kategorik değişkenlere ait lojistik fonksiyonlar en genel hali ile¹³⁴

$$P_j = \frac{e^{\beta_{0j} + \sum_{k=1}^p \beta_{kj}x_k}}{1 + \sum_{j=1}^{m-1} e^{\beta_{0j} + \sum_{k=1}^p \beta_{kj}x_k}} \quad j = 1, 2, \dots, m - 1$$

$$P_m = \frac{1}{1 + \sum_{j=1}^{m-1} e^{\beta_{0j} + \sum_{k=1}^p \beta_{kj}x_k}} \quad 1.25$$

şeklinde ifade edilebilir. Buradan elde edilen her lojit için sabit dahil tüm parametre katsayıları birbirinden farklıdır.

1.4.1.1 Model Parametrelerinin Tahmin Yöntemleri

Çok Terimli Lojistik Regresyon Modelleri'nin parametre tahminleri ve anlamlılık testleri İki Kategorili Lojistik Regresyon Modelleri'yle oldukça benzer olmakla birlikte parametre tahminleri için bu modellerde de bazı hesaplama farklılıkları ile En Çok Olabilirlik yöntemi kullanılmaktadır.

Eşitlik 1.25'de verilen lojistik fonksiyonda, i . birimine ait x 'lerin gözlemlenen değerlerinin vektörü x_i ile gösterildiğinde ve eşitlikte x 'ler yerine x_i 'ler kullanıldığında i . birime ait olasılıklar da P_{ij} ($j = 1, 2, \dots, m - 1$) ve P_{im} ile ifade edilmektedir. Ayrıca bağımlı değişken kümesinde i . birim j . kategoriye düşüyorsa

¹³⁴ Tim Futing Liao, a.g.e., pp. 48–49.

$y_{ij} = 1$, diğer durumlar içinse $y_{ij} = 0$ olacak şekilde tanımlandığında birbirinden bağımsız n gözlem için Çok Terimli Lojit Modelin olabilirlik fonksiyonu,¹³⁵

$$L = \prod_{i=1}^n P_{i1}^{y_{i1}} P_{i2}^{y_{i2}} \dots P_{im}^{y_{im}}$$

şeklinde, log-olabilirlik fonksiyonu da

$$\log L = \sum_{i=1}^n \sum_{j=1}^m y_{ij} \log P_{ij} \quad 1.26$$

şeklinde gösterilmektedir.

i . birime ait alternatiflerin olasılık değerleri;

$$\left[P_{ij} = \frac{\exp(x_i' \beta_j)}{1 + \sum_{j=1}^{m-1} \exp(x_i' \beta_j)} \quad j = 1, 2, \dots, m-1 \right] \quad \text{ve} \quad \left[P_{im} = \frac{1}{1 + \sum_{j=1}^{m-1} \exp(x_i' \beta_j)} \right]$$

iken, kısmi türevlerinin basitleştirilmiş şekli aşağıdaki gibi gösterilmektedir;

$$\frac{\partial P_{ij}}{\partial \beta_j} = P_{ij}(1 - P_{ij})x_i \quad j = 1, 2, \dots, m-1$$

$$\frac{\partial P_{ij}}{\partial \beta_k} = -P_{ij}P_{ik}x_i \quad j \neq k = 1, 2, \dots, m-1$$

$$\frac{\partial P_{im}}{\partial \beta_j} = -P_{ij}P_{im}x_i$$

Buradan da, log-olabilirliğin kısmi türevleri;¹³⁶

$$\begin{aligned} \frac{\partial \log L}{\partial \beta_k} &= \sum_{i=1}^n \left[\frac{y_{ik}}{P_{ik}} P_{ik}(1 - P_{ik}) + \sum_{\substack{j=1 \\ j \neq k}}^m \frac{y_{ij}}{P_{ij}} (-P_{ij}P_{ik}) \right] x_i \\ &= \sum_{i=1}^n \left[y_{ik} - P_{ik} + \sum_{j=1}^m y_{ij} \right] x_i \end{aligned}$$

¹³⁵ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, pp. 262–263.

¹³⁶ G. S. Maddala, **a.g.e.**, p. 36.

$$= \sum_{i=1}^n (y_{ik} - P_{ik})x_i$$

olarak bulunmaktadır. Elde edilen kısmi türev denklemleri, En Çok Olabilirlik tahminlerini elde etmek için sıfıra eşitlenerek çözülür.

$$\left[\sum_{i=1}^n (y_{ij} - P_{ij})x_i = 0 \quad j = 1, 2, \dots, m - 1 \right]$$

Bu denklemler basit lojistik modele benzer olup P_{ij} , β 'ların doğrusal olmayan bir fonksiyonu olduğu için ancak Newton-Raphson metodu gibi iteratif (tekrarlı) bir yöntemle çözülebilmektedir. Başlangıç değeri olarak bilgisayar programlarının çoğu tarafından diskriminant fonksiyonunun katsayıları kullanılmaktadır.¹³⁷

1.4.1.2 Çok Terimli Lojistik Regresyon Modellerine İlişkin Varsayım

Çok Terimli Lojistik Regresyon Modelleri, İki Kategorili Lojistik Regresyon Modelleri'nin parametre tahminlerine ve yorumlanmasına benzerliği nedeniyle en çok kullanılan regresyon modelleri olmasına rağmen bağımlı değişkenin kategori sayısı ve parametre sayısı çok olduğunda karmaşıklaşmaktadır. Ayrıca Çok Terimli Lojistik Regresyon Modelleri, İki Kategorili Lojistik Regresyon Modellerinin sağlaması gereken varsayımlara ilave olarak İlgisiz Alternatiflerin Bağımsızlığı (Independence of Irrelevant Alternatives – IIA) varsayımını da sağlaması gerekmektedir.

Bu varsayım altında; x biliniyorken sonuç j 'nin gözlenme olasılığı,

$$P(y = j|x) = \frac{\exp(x\beta_j)}{\sum_{j=1}^m \exp(x\beta_j)} \quad j = 1, 2, \dots, m$$

şeklinde hesaplanmaktadır. Burada $\beta_j = (\beta_{0j} \dots \beta_{kj} \dots \beta_{pj})'$ vektörüdür ve Çok Terimli Lojit Modeller'in her bir m ve n gibi alternatif çiftlerinin odds değeri,

¹³⁷ G. S. Maddala, a.g.e., p. 37.

$$\Omega_{m|n} = \exp[x(\beta_m - \beta_n)]$$

olarak yazılabilir. Yukarıdaki eşitlikde olduğu gibi n 'ye karşı m 'nin seçilmesi olasılığı diğer sonuçlara bağlı olmaksızın yalnızca m ve n 'nin katsayı vektörleri (β_m ve β_n) ile belirleniyorsa, buna “İlgisiz Alternatiflerin Bağımsızlığı” (IIA) özelliği adı verilmektedir.¹³⁸

Kısaca; İlgisiz Alternatiflerin Bağımsızlığı (IIA), her şey aynı iken, iki alternatifli sonuçlar arasındaki kişinin tercihinin mevcut diğer alternatiflerden etkilenmemesi anlamına gelmektedir. Bu da kalan alternatifleri dikkate almadan herhangi iki alternatifin göreceli odds değeri üzerinde, değişkenlerin etkilerinin karşılaştırılabilir olmasını sağlamaktadır.

Sezgisel olarak, IIA özelliği MNL (Çok Terimli Lojit) Model içinde, EKK regresyonu için hata terimlerinin bağımsızlığı varsayımına benzer bir rol oynamaktadır ve IIA özelliğinin ihlali tahmin edilen seçenek olasılıklarının sistematik hatalarına neden olmaktadır. IIA özelliği belirli bir seçim durumu için değil belirli bir model için geçerli ya da geçersizdir.¹³⁹

IIA Varsayımının Testi

IIA varsayımının geçerliliği için önerilen testlerin temel özelliği, bağımsız değişkenlerin alabileceği tüm değerlerin içinde olduğu Çok Terimli Lojistik Regresyon Modellerin tahminlerinin sonuçları ile sadece bazı sonuç değerlerini içeren sınırlı tahmin değerlerinin karşılaştırılması üstüne kurulmuştur. Burada, sınırlı modelin parametre tahminleri $\hat{\beta}^r = (\hat{\beta}_2^r, \dots, \hat{\beta}_{m-1}^r)$, tam modelin parametre tahminleri ise $\hat{\beta}^f = (\hat{\beta}_2^f, \dots, \hat{\beta}_{m-1}^f)$ vektörlerinden oluşmaktadır. $\hat{\beta}^f$ içinde, sınırlı modelin parametre tahminler setinde yer almayan m kategorisine ait parametre tahminleri bulunmamaktadır. Bu parametre tahmin setleri;

¹³⁸ Simon Cheng, J. Scott Long, **a.g.e.**, pp. 583–600

¹³⁹ Daniel McFadden, Kenneth Train and William B. Tye, “An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model”, **Transportation Research Board Record**, 637, 1981, pp. 39–46.

$$P(y = j|x) = \frac{\exp(x\beta_j)}{\sum_{k=1}^m \exp(x\beta_k)} \quad j = 1, 2, \dots, m$$

ile tüm değerlerin eşzamanlı karşılaştırılması yoluyla elde edilmiştir ve $\hat{\beta}_j^f$ ile gösterilmektedir. Üst indis f , parametrelerin tüm bağımlı değişken kategorilerini içeren tam modelden tahmin edildiğini ifade etmektedir. Sınırlanmış tahmin değerleri tam modelle aynı şekilde, sadece bağımlı değişkenin m . kategori dışarıda bırakılarak, şu şekilde yazılabilir,

$$P(y = j|x) = \frac{\exp(x\beta_j)}{\sum_{k=1}^{m-1} \exp(x\beta_k)} \quad j = 1, 2, \dots, m - 1$$

Sınırlanmış tahmin değerleri $\beta_1 = 0$ olduğu varsayımı altında m kategorisi dışarıda bırakılarak yukarıdaki ifadeye belirtilen haliyle gösterilebildiği gibi, herhangi bir kategoride dışarıda bırakılarak gösterilebilmektedir.

IIA varsayımının özünde ifade ettiği, sınırlanmış tercih kümesinden elde edilen $\hat{\beta}_j^r$ ile gösterilen tahminlerin tutarlı fakat yetersiz (etkin olmayan) tahminler olduğudur. Tam modelden tahmin edilen katsayılar ile sınırlanmış modelden tahmin edilen katsayılar benzer ise IIA varsayımı sağlanmış olur.¹⁴⁰ Başka bir ifade ile eğer test istatistiği anlamlı ise, IIA varsayımı reddedilir ve bu da Çok Terimli Lojit Modelin sonuçlarının uygun olmadığı anlamına gelmektedir.

IIA'nın ilk testi McFadden, Train ve Tye (1981)¹⁴¹ tarafından sunulmuş MTT testidir. Bu test, sınırlı modelin log-olabilirliği içine tam model tahminlerini koyarak elde edilen değerleri sınırlı tahminden gelen log-olabilirlik değeriyle karşılaştırılmaktadır.

$$MTT = -2[L_r(\hat{\beta}^f) - L_r(\hat{\beta}^r)]$$

¹⁴⁰ Simon Cheng, J. Scott Long, **a.g.e.**, pp. 583–600.

¹⁴¹ Daniel McFadden, Kenneth Train and William B. Tye, **a.g.e.**, pp. 39–46.

Burada L_r , sınırlanmış tahminlerin log-olabilirlik fonksiyonudur. MTT, $\hat{\beta}^r$ 'deki satır sayısına eşit serbestlik derecesi ile χ^2 dağılımına uyar.¹⁴²

Bununla birlikte 1985'de Small ve Hsiao, MTT testinin yanlı sonuçlar verdiğini gösterip, Olabilirlik Oran testinin alternatifi olan ve yanlılığı ortadan kaldıran Small ve Hsiao (SH) testini kullanılmasını önermişlerdir.

$$SH = -2[L_r(\hat{\beta}_{AB}^f) - L_r(\hat{\beta}_B^r)]$$

SH, sınırlanmış seçim setindeki parametre sayısı kadar serbestlik derecesi ile χ^2 dağılımına uymaktadır.¹⁴³

Small ve Hsiao, bu yanlılığı engellemek için MTT'yi uyarlayarak sunmuşlardır. İlk olarak, örneklem büyüklüğü eşit olacak şekilde A ve B gibi iki alt örneklem için tam model eşitliğinden parametreleri tahmin etmişlerdir. Bu tahmin değerlerini $\hat{\beta}_A^f$ ve $\hat{\beta}_B^f$ olarak ifade ederek iki örneklemin katsayılarının ağırlıklandırılmış ortalamalarını şu şekilde hesaplatmışlardır;

$$\hat{\beta}_{AB}^f = \frac{1}{\sqrt{2}}\hat{\beta}_A^f + \left[1 - \frac{1}{\sqrt{2}}\right]\hat{\beta}_B^f$$

IIA varsayımının tespiti için Small ve Hsiao (SH) testi ile birlikte önerilen bir test de, 1984'de Hausman ve McFadden tarafından önerilen HM¹⁴⁴ testidir. Bu test, tüm alternatiflerin modelde olduğu eşitlikteki parametre seti $\hat{\beta}^f$ ile alternatiflerin alt setlerinden oluşan sınırlanmış modelden elde edilen parametre seti $\hat{\beta}^r$ ile elde edilen katsayıların karşılaştırılmasıdır.

HM istatistiği şu şekilde hesaplanmaktadır,¹⁴⁵

$$HM = (\hat{\beta}^r - \hat{\beta}^f)'[\widehat{Var}(\hat{\beta}^r) - \widehat{Var}(\hat{\beta}^f)]^{-1}(\hat{\beta}^r - \hat{\beta}^f)$$

¹⁴² Simon Cheng, J. Scott Long, **a.g.e.**, pp. 583–600.

¹⁴³ Kenneth A. Small, C. Hsiao, "Multinomial Logit Specification Tests", **International Economic Review**, Vol: 26, No: 3, 1985, pp. 619–627.

¹⁴⁴ J. Hausman, D. McFadden, "Specification Tests for the Multinomial Logit Model", **Econometrica**, Vol: 52, No: 5, 1984, pp. 1219–1240.

¹⁴⁵ Simon Cheng, J. Scott Long, **a.g.e.**, p.589.

Burada $\widehat{Var}(\hat{\beta}^r)$ ve $\widehat{Var}(\hat{\beta}^f)$ tahmin edilen kovaryans matrisleridir. HM, sınırlı modeldeki parametre sayısı kadar serbestlik derecesi ile χ^2 dağılımına uymaktadır. Eğer hesaplanan HM değeri χ^2 tablo değerinden büyükse, iki eşitliğin katsayılarının eşit olduğunu ileri süren sıfır hipotezi ($H_0: \hat{\beta}^f = \hat{\beta}^r$) reddedilir. Yani IIA varsayımı sağlanamamıştır.¹⁴⁶ Bu durumda Çok Terimli Lojit Modelin kullanılmaması gerekmektedir. Onun yerine modelin İç içe Lojit veya Çok Terimli Probit Model kullanılarak tekrar tahmin edilmesi gerekmektedir.

Yukarıdaki her testin çeşitli versiyonları sınırlı tercih setinin oluşturulmasında farklı alternatifler elenerek oluşturulabilir. Örneğin, eğer tek kategori çıkararak sınırlanmış modelden tahminler elde ediliyorsa; 1. versiyonda, ilk kategori dışarıda bırakarak sınırlandırılmış tahminler oluşturulabilir ve 2. versiyonda, ikinci kategori dışarıda bırakılarak bu şekilde devam edilebilir.

Fry ve Harris (1996)'de yaptığı ampirik çalışmada, Monte Carlo Simulasyonu kullanarak, IIA varsayımının testi için yukarıda verilen MTT, SH ve HM testlerini altı farklı sınırlanmış alternatif set sonuçları için karşılaştırmıştır. SH testinin olması gerekenden büyük çıktığını ve HM testinin kabul edilebilir büyüklükte olduğunu bulmuşlardır. Ayrıca, MTT testi değerleri olması gerekenin altında bulunmuştur.¹⁴⁷

Bu testler IIA varsayımı geçerli ise, karmaşık olan İç içe Lojit Model veya Çok Terimli Probit Model yerine, Çok Terimli Lojit Model'i tahmin etmeyi mümkün kıldığı için kullanışlıdır.¹⁴⁸ İç içe Lojit Modeller, karşılaştırılmasının zor olması ve uygulama yapan araştırmacılar tarafından pek bilinmemesi nedeni ile uygulamalarda nadiren kullanılmaktadır. Çok Terimli Probit için de durum, uygulamadaki zorlukları bakımından benzerdir.

¹⁴⁶ J. Hausman, D. McFadden, **a.g.e.**, pp. 1219–1240.

¹⁴⁷ Tim R. Fry, Mark N. Harris, "A Monte Carlo Study of Tests for the Independence of Irrelevant Alternatives Property", **Transportation Research Part B: Methodological**, Vol: 30, 1996, pp. 19–30.

¹⁴⁸ William H. Crown, **Statistical Models for the Social and Behavioral Sciences**, Praeger, p. 135.

1.4.1.3 Uyum İyiliği Testleri¹⁴⁹

Uyum iyiliği ölçüleri; gözlemlenen verilerin, bir modelle yaklaşık olarak doğruluğunu belirleyen özet istatistiklerdir. Bağımlı değişkenlerin nitel olduğu durumlar için uyum iyiliği testlerinin doğruluğuna hesaplanan olasılıklar ve gözlenen tepki frekansları arasındaki uyum açısından veya gözlenen tepkileri tahmin eden model açısından karar verilebilmektedir.¹⁵⁰

R^2 ve *Pseudo* – R^2 değerleri LR (Olabilirlik Oran) testlerinden türetilebilir. R^2 sadece bağımlı değişkenlerin iki değer aldığı durumlar için kullanılmalıdır. Çok Terimli Lojit Modeller için R^2 hesaplanamamaktadır. Bunun yerine LR test istatistiğine dayalı bir uyum iyiliği ölçüsü tanımlanmalıdır. R^2 'nin Doğrusal Regresyon Modeli için LR test istatistiği ile ilişkili olarak tanımlandığı varsayıldığında; L_Ω , $\alpha, \beta_1, \beta_2, \dots, \beta_p$ gibi tüm parametreler açısından maksimize edilmiş olabilirlik fonksiyonunun maksimumu ve L_w , sadece α 'in maksimize edilmiş maksimum değeri ise uyum iyiliği ölçüsü,¹⁵¹

$$R^2 = 1 - \left(\frac{L_w}{L_\Omega} \right)^{2/n}$$

olarak formüle edilmektedir.

j . kategoride gözlemlenen birim n_j ile gösterilirse $j = 1, 2, \dots, m - 1$ için

$\exp(\hat{\alpha}_j) = \frac{n_j}{n_m}$ olacağı açıkça görülmektedir.

Buradan

$$L_w = \prod_{j=1}^m \left(\frac{n_j}{n} \right)^{n_j} \quad 1.27$$

elde edilir. Yukarıdaki R^2 değerinin Lojit Modelin uygunluğunu ölçmek için iyi bir ölçüt olduğu düşünülse de aslında $R^2 \rightarrow 0$ değeri ancak $L_\Omega \rightarrow L_w$ 'ya yaklaşırsa

¹⁴⁹ G. S. Maddala, **a.g.e.**, pp. 34–41.

¹⁵⁰ Paul Newbold, William L. Carlson, Betty Thorne, **a.g.e.**, p. 613.

¹⁵¹ G. S. Maddala, **a.g.e.**, p. 39.

sağlanacaktır ve model doğrusal olmadığı sürece R^2 'nin üst sınırınının 1'den oldukça küçük olması gibi bir sorun bulunmaktadır. Ayrıca olabilirlik fonksiyonu ancak $L = \prod_{i=1}^n P_{i1}^{y_{i1}} P_{i2}^{y_{i2}} \dots P_{im}^{y_{im}}$ iken mutlak maksimum 1'i ($L_{\max} = 1$) sağlayacaktır. Buradan aşağıdaki eşitsizlikler elde edilmektedir;

$$\prod_{j=1}^m \left(\frac{n_j}{n}\right)^{n_j} = L_w \leq L_\Omega \leq L_{\max} = 1$$

$$\frac{L_w}{L_{\max}} \leq \frac{L_w}{L_\Omega} \leq 1$$

Daha önce verilen $R^2 = 1 - \left(\frac{L_w}{L_\Omega}\right)^{2/n}$ eşitliğinin yardımı ile $0 \leq R^2 \leq 1 - (L_w)^{2/n}$ olarak bulunur. Bu da gösteriyor ki, model veri setine mükemmel şekilde uysa da R^2 değeri her zaman 1'den küçük olacaktır.¹⁵²

Daha iyi bir ölçüm olan *Pseudo* – R^2 şu şekilde tanımlanır,

$$Pseudo - R^2 = \frac{1 - (L_w/L_\Omega)^{2/n}}{1 - (L_w/L_{msx})^{2/n}} = \frac{L_\Omega^{2/n} - L_w^{2/n}}{1 - L_w^{2/n}}$$

L_w değeri Eşitlik 1.27'den hesaplanmatadır, $L_{\max} = 1$ ve L_Ω de istatistik paket programı tarafından hesaplanan maksimum olabilirlik fonksiyonu değeridir.

1.4.1.4 Model Parametrelerinin Yorumu

Çok Terimli Lojistik Regresyon Model, parametrelerin yorumu bakımından İki Kategorili Lojistik Regresyon Modele oldukça benzemekle birlikte bazı farklılıkları da içermektedir.

Çok Terimli Lojistik Model'de her bir birim için tüm alternatiflerin tahmin olasılıklarının tamamının toplamı "1" ($\sum_{j=1}^m \hat{p}_{ij} = 1$) olmalıdır¹⁵³. Sonuç olarak, Çok Terimli Lojistik Regresyon Modeller bağımlı değişkenin "m" sayıdaki kategorisi için

¹⁵² G. S. Maddala, **a.g.e.**, p. 40.

¹⁵³ J. Scott Long, **a.g.e.**, p. 153.

“m-1” katsayı setini içermektedir. Her bir eşitlik $\log(P_m/P_0)$, açıklayıcı değişkenle ilişkilidir.

İki Kategorili Lojit Modeller için, belirli bir değişkenin katsayısının işareti bağımlı değişkenin odds değerlerinin logaritmalarının artıyor veya azalıyor olduğunu göstermektedir. Odds değerlerinin logaritmaları ile bağımlı değişken kategorilerinin olasılıkları aynı yönde hareket etmektedir. Fakat bunun Çok Terimli Lojistik Regresyon Modeller için doğru olması beklenmemektedir. Parametre tahminlerinin işaretlerinin istatistik anlamlılığı yalnızca referans grupla ilişkili etkilerin yönü için güvenilir göstergeler olarak kullanılmaktadır.¹⁵⁴ Örneğin; açıklayıcı değişkenin pozitif bir parametre tahminine sahip olması, değişken ile referans grubuna göre ilgilenilen sonuç kategorisinin lojit (log-odds) değeri arasındaki ilişkinin pozitif olduğunu göstermekle birlikte, aynı açıklayıcı değişkenin, sonucun kendi olasılığı ile ilişkisinin negatif olması da mümkündür. Bu, Çok Terimli Lojit Modeller’in açıklayıcı değişkenlerinin kısmi türevlerinin test edilmesiyle anlaşılmaktadır.¹⁵⁵

Kısmi türevler, belirli bir değişkenin üzerindeki olay olasılığının, odds değerlerinin logaritma formülasyonu ile karışıklığa neden olmaksızın incelenmesini sağlamaktadır. Çok Terimli Lojit Model için kısmi türevler bağımlı değişkenin üç kategorili olduğu varsayıldığında, iki tahmin eşitliği ile şu şekilde yazılmaktadır;¹⁵⁶

$$\log\left(\frac{P_{1i}}{P_{0i}}\right) = \hat{\beta}_{10} + \sum_{k=1}^p \hat{\beta}_{1k} x_{ki}$$

$$\log\left(\frac{P_{2i}}{P_{0i}}\right) = \hat{\beta}_{20} + \sum_{k=1}^p \hat{\beta}_{2k} x_{ki}$$

Hesaplamalardan sonra,

$$P_{0i} = \frac{1}{1 + e^{\hat{\beta}_{10} + \sum_{k=1}^p \hat{\beta}_{1k} x_{ki}} + e^{\hat{\beta}_{20} + \sum_{k=1}^p \hat{\beta}_{2k} x_{ki}}}$$

¹⁵⁴ Vani K. Borooah, **Logit and Probit Ordered and Multinomial Models**, Sage Publications, 2002, p. 51.

¹⁵⁵ William H. Crown, **a.g.e.**, p. 129.

¹⁵⁶ **A.e.**, pp. 129–130.

$$P_{1i} = \frac{e^{\hat{\beta}_{10} + \sum_{k=1}^p \hat{\beta}_{1k} x_{ki}}}{1 + e^{\hat{\beta}_{10} + \sum_{k=1}^p \hat{\beta}_{1k} x_{ki}} + e^{\hat{\beta}_{20} + \sum_{k=1}^p \hat{\beta}_{2k} x_{ki}}}$$

$$P_{2i} = \frac{e^{\hat{\beta}_{20} + \sum_{k=1}^p \hat{\beta}_{2k} x_{ki}}}{1 + e^{\hat{\beta}_{10} + \sum_{k=1}^p \hat{\beta}_{1k} x_{ki}} + e^{\hat{\beta}_{20} + \sum_{k=1}^p \hat{\beta}_{2k} x_{ki}}}$$

örneklemedeki n birim için her bir kategoride olma olasılıkları elde edilmektedir. Belirli bir eşitlikteki x_{jk} bağımsız değişkeninin bu olasılıklara bağlı olarak kısmi türevinin alınması ile aşağıdaki eşitlik elde edilir,

$$\frac{\partial P_j}{\partial x_{jk}} = P_j(1 - P_j)\hat{\beta}_{jk}$$

Diğer eşitlikteki x_{lk} bağımsız değişkeninin belirli bir sonucunun olasılığına göre kısmi türevi de

$$\frac{\partial P_j}{\partial x_{lk}} = -P_j P_l \hat{\beta}_{lk}$$

olarak gösterilmektedir.

Burada yine, türev alma işlemi bir değişkendeki çok ufak bir değişimin diğer değişken üzerindeki etkisini vermekte olduğu için açıklayıcı değişken nitel değişken olduğu zaman kısmi türevlerin hesaplanması ile ilgili problemler vardır. Çok Terimli Lojit Modeller’de de bu problemi gidermek için temel yaklaşım, İki Kategorili Lojit Model’de olduğu gibi OR değerlerinin hesaplanmasıdır. Çok Terimli Lojit Modeller için OR, İki Kategorili Lojit Modeller’deki ile aynı şekilde tanımlanır ve yorumlanırlar.¹⁵⁷

¹⁵⁷ Tim Futing Liao, **a.g.e.**, p.51.

1.4.2 Çok Terimli Probit Modeller

Çok Terimli Probit Model Aitchison ve Bennett (1970)¹⁵⁸ tarafından geliştirilmiş ve Hausman ve Wise (1978)¹⁵⁹ tarafından bağımlı değişkenin üç kategorili durumu için uygulaması yapılmıştır.

İki Kategorili Lojit ve Probit modellerinde olduğu gibi Çok Terimli Lojit ve Probit modeller için de olasılıkların dağılımlarının farklı olduğu kabul edilmektedir. Çok Terimli Lojit Modellerin, Çok Terimli Birikimli Lojit dağılıma uyduğu varsayılırken Çok Terimli Probit Modellerin, Çok Terimli Birikimli Normal dağılıma sahip olduğu varsayılmaktadır. Çok Terimli Probit Modeller, teorik hesaplamadaki zorlukları nedeniyle, özellikle bağımlı değişken üç kategoriden fazla olduğu durumlarda çok nadiren model tahminlemek için kullanılmaktadır. Genellikle IIA varsayımı geçerli olduğu sürece Çok Terimli Lojit Modellerin kullanılması tercih edilmekle birlikte varsayımın geçerli olmadığı durumlarda Çok Terimli Lojit Modeller yerine İç içe Lojit Modeller* kullanılmalıdır.

Bağımlı değişken iki kategoriden fazla değer aldığı zaman, Çok Terimli Lojit Modeller ve İç içe Lojit Modeller arasında seçim IIA varsayımının istatistik olarak anlamlı olmasına göre şekillenmektedir. IIA varsayımının reddedilmesi için istatistik olarak kanıt yok ise, Çok Terimli Lojit Model kullanılmaktadır. Eğer IIA varsayımı reddediliyorsa, Çok Terimli Probit, İç içe Lojit, Sıralı Lojit veya Sıralı Probit modeller tahmin edilmelidir.¹⁶⁰

1.4.3 Sıralı Lojistik Regresyon Modelleri

Sosyal bilimlerde yapılan araştırmalarda bağımlı değişkenin aldığı değerler çoklu kategoride olabileceği gibi, bazı durumlarda kategorik (nominal) değil sıralı (ordinal)

¹⁵⁸ J. Aitchison, J. A. Bennett, "Polychotomous Quantal Response by Maximum Indicant", **Biometrika**, Vol: 57, No: 2, 1970, pp. 253–262.

¹⁵⁹ Jerry A. Hausman, David A. Wise, "A Conditional Probit Model for Qualitative Choice: Discrete Decision recognizing Interdependence and Heterogeneous Preferences", **Econometrica**, Vol: 46, No: 2, 1978, pp. 403–426.

* İç içe Lojit Model'e, "1.4.5. Çok Kategorili Bağımlı Değişken için Diğer Modeller" başlığı altında yer verilmiştir.

¹⁶⁰ William H. Crown, **a.g.e.**, pp. 125–158.

şekilde bulunmaktadır. Sıralı bağımlı deęişken için modeller ilk olarak Aitchison ve Silvey (1957) tarafından ele alınmıştır. Sıralı bağımlı deęişkenler için en yaygın örnekler; sosyal ve kamu fikri ile ilgili sorulara verilen cevaplarda genellikle “kesinlikle katılmıyorum’dan kesinlikle katılıyorum’a” veya politikada adayların görüşleri ile ilgili ne düşündükleri sorulduğunda “güçlü olarak desteklemiyorum’dan güçlü olarak destekliyorum” aralığında likert tipi ölçekler kullanıldığı durumlar olarak verilebilir. Bunun dışında kişilerin işlerindeki başarı düzeyi, eğitim durumları, istihdam durumu (işsiz, yarı zamanlı ve tam zamanlı) gibi veriler veya iş performansı için (yetersiz, yeterli, üstün) yapılan deęerlendirmeler sıralı bağımlı deęişkene verilebilecek dięer örneklerdir.

Bağımlı deęişken deęerleri sıralı kategoride olan modeller, klasik regresyonun varsayımlarında meydana gelecek sapmalar nedeniyle klasik regresyonla modellenemediđi gibi Çok Terimli Lojit Modeller de bağımlı deęişkenin sıralanış doęasını açıklamada başarısız olacaklarından dolayı Sıralı Lojit Modellerinin kullanılması daha uygun olacaktır.¹⁶¹

Sıralı Lojit Modelleri’nde kategoriler arasında çok açık sıralama olmakla birlikte bitişik kategoriler arasındaki mesafe aynı veya deęişmez olarak kabul edilmemektedir. Başka bir deyişle, sonuçların niteliđi nedeniyle cevapların güçlerinin farkları ile ilgili herhangi saklı bir anlam yoktur, yani bağımlı deęişkenin $y_i = 2$ olan deęeri, $y_i = 1$ olan deęerinden iki kat daha güçlü anlamına gelmemektedir.¹⁶²

Sıralı Lojit Modeller, bağımlı deęişkenin iki kategorili olduđu modellerin Çok Terimli Lojit Modeller gibi doęal bir uzantısı olup, İki Kategorili Lojit Modellerin’dekiyle aynı çerçevede oluşturulur:

$$y^* = \sum_{k=1}^p \beta_k x_k + \varepsilon$$

¹⁶¹ Tim Futing Liao, **a.g.e.**, p. 37.

¹⁶² Vani K. Borooah, **a.g.e.**, p.5.

Bağımlı değişkenin iki kategorili olduğu modeldeki gibi y^* ; gözlenemeyen ama olayın altında yatan eğimi veren değerdir ve ε hata teriminin de ortalaması sıfır olan Lojistik dağılıma uyduğu varsayılır.

y bağımlı değişkeninin gözlenen m sayıdaki sıralı kategorisi aşağıdaki gibi tanımlanmaktadır.¹⁶³

$$\begin{aligned} y = 1 & \quad \text{eğer } y^* \leq \alpha_1 (= 0) \\ y = 2 & \quad \text{eğer } \alpha_1 < y^* \leq \alpha_2 \\ & \quad \vdots \\ y = m & \quad \text{eğer } \alpha_{m-1} < y^* \end{aligned}$$

α ve β 'lar, tahmin edilen bitişik kategorileri ayıran bilinmeyen parametrelerdir. p tane bağımsız değişkenden oluşan x vektörü biliniyorken, bağımlı değişkenin j . kategoriye düşme olasılığı için genel gösterim şekli şöyledir,¹⁶⁴

$$P(y = j) = F\left(\alpha_j - \sum_{k=1}^p \beta_k x_k\right) - F\left(\alpha_{j-1} - \sum_{k=1}^p \beta_k x_k\right) \quad 1.28$$

Eşitlik 1.28'de α 'lar ve β 'lar Sıralı Lojit ve Sıralı Probit model ile tahmin edilebilir. Lojit ve Probit model için bu denklemdaki tek fark, Lojit Model'de hataların dağılımı Lojistik dağılıma uyduğu için $F(\cdot)$ yerine $\Lambda(\cdot)$, Probit Model'de hataların dağılımı Standart Normal dağılıma uyduğu için $\Phi(\cdot)$ ile birikimli dağılım değerlerinin hesaplanıyor olmasıdır. Tüm olasılıklar sırası ile pozitif olmalıdır ($0 < \alpha_2 < \alpha_3 < \dots < \alpha_{m-1}$). İlk eşik parametresi olan α_1 , ölçek keyfi olarak belirlendiği ve herhangi bir değerle başlayabileceği için genellikle sıfır olarak alınır ve böylece bir parametre daha az tahmin edilir. Normalde eşik sayısı kategori sayısından bir az ($m - 1$) olacak şekilde eşik değerinin (α) tahmin edilmesi gerekirken $\alpha_1 = 0$ olarak kabul edildiğinde, tahmin edilecek α sayısı ($m - 2$) tanedir. Ayrıca, bazı bilgisayar programları yukarıdaki gibi hem ilk eşik değerinin (α_1) hem de sabit terimin (β_0) birlikte sıfır olduğunu varsayarken, bazı bilgisayar programları ilk eşik değerini sıfır

¹⁶³ J. Scott Long, **a.g.e.**, p. 116.

¹⁶⁴ Tim Futing Liao, **a.g.e.**, p.38.

kabul edip modeldeki sabit terimi tahmin eder. Burada parametrenin seçimi eğim katsayılarının (β_i) tahminini etkilemez, β_0 ve α 'ların tahminini etkiler¹⁶⁵.

Sıralı Lojit Modeller ile genellikle $y \leq j$ kategorilerinin olasılıkları ile $y > j$ kategorilerinin olasılıklarını karşılaştırmak üzere doğrusal ilişkiyi gösteren Sıralı Lojit Modeller şu şekildedir;¹⁶⁶

$$\log \left[\frac{P(y \leq j|x)}{1 - P(y \leq j|x)} \right] = \alpha_j - \sum_{k=1}^p \beta_k x_k, \quad j = 1, 2, \dots, m - 1$$

Sıralı ve İki Kategorili Lojit modeller arasındaki tek fark; bağımlı değişkenin sıralı olduğu modellerde, bağımsız değişkenlere ait katsayılar (β_k 'lar) aynı iken sabit terimleri ifade eden α 'lar ile farklı tanımlanan lojitlerin (log-odds) ardışık olmasına izin veriyor olmasıdır. Eşitliğin sol tarafı literatürde *Birikimli Lojit* (Cumulative Logit), *Devam Oranları* (Continuation Ratios) veya *Oransal Odds* (Proportional Odds) olarak adlandırılmaktadır. Yukarıdaki eşitlikten Lojistik Regresyon Modelleri (veya Lojistik dağılım için birikimli dağılım fonksiyonunu) şu şekilde ifade edilmektedir;¹⁶⁷

$$P(y \leq j) = P(y^* \leq \alpha_j) = \frac{e^{\alpha_j - \sum_{k=1}^p \beta_k x_k}}{1 + e^{\alpha_j - \sum_{k=1}^p \beta_k x_k}}$$

Buradan Sıralı Lojit Modeller'in olasılık terimleri ise aşağıdaki gibi ifade edilmektedir;

$$\begin{aligned} P(y = 1) &= \Lambda \left(- \sum_{k=1}^p \beta_k x_k \right) \\ P(y = 2) &= \Lambda \left(\alpha_2 - \sum_{k=1}^p \beta_k x_k \right) - \Lambda \left(- \sum_{k=1}^p \beta_k x_k \right) \\ &\vdots \end{aligned}$$

¹⁶⁵ J. Scott Long, Simon Cheng, **a.g.e.**, p. 269.

¹⁶⁶ Tim Futing Liao, **a.g.e.**, p. 40.

¹⁶⁷ J. Scott Long, **a.g.e.**, p.119.

$$P(y = m) = 1 - \Lambda \left(\alpha_{m-1} - \sum_{k=1}^p \beta_k x_k \right)$$

Sıralı bağımlı deęişken modelinde, bağımlı deęişken kategorisi j 'nin deęerine bakılmaksızın x 'in etkisini sabit kabul etmektedir. Buna, "paralel doęrular" (parallel lines) varsayımı adı verilmektedir. Bazı kaynaklarda "paralel eęim" (parallel slope)¹⁶⁸ olarak da geçmektedir. Bu, lojit durum için genellikle "orantılı oranlar" (proportional odds) varsayımı ve probit durum için de "eşit eęimler" (equal slopes) varsayımı adını almaktadır.¹⁶⁹

Sıralı Lojistik Regresyon Modelleri, bağımlı deęişkenlerin sıralı olduęu durumlar için en sık kullanılan modeldir. Bu modelin sonuçlarının doęru bir şekilde yorumlanabilmesi için paralel eęriler veya oransal odds olarak adlandırılan varsayımın saęlanması gerekmektedir.¹⁷⁰

1.4.3.1 Sıralı Lojistik Regresyon Analizinde Model Kurma Stratejileri

Sıralı Lojistik Regresyon Modelleri oluşturulurken izlenecek yol İki Kategorili Lojistik Regresyon Modellerin'deki ile oldukça benzerdir. İlk olarak modele alınan bağımsız deęişkenlerin anlamlılık testleri yapılır ve çoklu doęrusal bağlantı sorunu olup olmadıęı kontrol edilerek başlanır veya modele girecek bağımsız deęişkenlerin seçimi Adım Adım (stepwise) yöntemiyle yapılır. Daha sonra, Sıralı Lojistik Regresyon Modellerinde yer alan sürekli bağımsız deęişkenlerin ölçęi kontrol edilir ve İkili Kategorili Lojistik Regresyonda olduęu gibi $y = k$ 'ya karşı $y = "0"$ ın p tane ayrı kesirli polinomial analizi uygulanarak doęrusal olup olmadıęı kontrol edilip modele alınacak deęişkenlere karar verilir. Son olarak; seçilmiş modelde etkileşimli deęişkenlerin yer alıp almayacağına, deęişkenlerin etkileşimleri tek tek modele sokulup tahmin edilen parametrelerin testleri tekrar yapılarak karar verilir. Uygun model seçimi için, temel modelle karşılaştırma yapmaya olanak veren olabilirlik oranı testi kullanılmaktadır.¹⁷¹ Ayrıca paralel eęriler varsayımı da kontrol edilmeli ve

¹⁶⁸ Vani K. Borooah, **a.g.e.**, p. 14.

¹⁶⁹ Tim Futing Liao, **a.g.e.**, p. 41.

¹⁷⁰ Simon Cheng, J. Scott Long, **a.g.e.**, p. 270.

¹⁷¹ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, pp. 305–308.

bu aşamanın sonrasında kullanılacak modele karar verilmelidir. Sıralı Lojistik Regresyon Modelleri tanımsal istatistik ve uyum iyiliği testleri kullanmak için çok uygun olmadığı için modele karar verirken ayrı ayrı İkili Lojistik Regresyon Modeller yaklaşımı kullanılabilir. Bu yaklaşımın en büyük dezavantajı gerçek uygun modeli kontrol etmeyip sadece yaklaşık olanını kontrol ediyor olmasıdır. Bununla birlikte bu yöntem, örneklemdaki birimlerin en az bağımsız değişkenle açıklanabilecek şekilde modellenmesi konusunda etkili olacaktır. Sonuç olarak OR tahminine dayanarak elde edilen sonuçlar Sıralı Lojistik Regresyon Modellerini yorumlamakta kullanılabilir.

Model kurma sürecine ilave olarak eğer bağımlı değişkenin sıralı olup olmadığından emin olunamıyorsa; duyarlı olan yol, sıralı değilmiş gibi kabul edilmesi ve bağımlı değişken için Çok Terimli Lojit Model'in kullanılarak parametre tahminlerinin yapılmasıdır.¹⁷²

Sıralı olmayan bağımlı değişkenine sıralı gibi davranılması, yanıtlar üzerine bir sıralama ve sahip olmadıkları paralel eğriler varsayımının sağlanmasını gerektireceği fakat büyük bir olasılıkla bu kuralın sağlanamayacak olması nedeniyle, parametre tahminlerinin yanlış olmasına neden olacaktır. Diğer taraftan gerçekte sıralı olan bağımlı değişkene sırasızmış gibi davranılması, cevaplar üzerinde meşru bir sıralamayı yapmakta başarısız olunması sonucunu doğuracaktır. Bu odds değerinde etki kaybına yol açması mümkün olmakla birlikte, parametre tahminlerinin yanlışlığına neden olması pek mümkün değildir. Bu iki olası hata karşısında etki kaybı hatası, tahminlerin yanlışlığı hatası yanında daha az ciddi bir hatadır.¹⁷³ Kısacası, bağımlı değişkeninin sıralanması için iyi bir neden sahip olunmadığı sürece cevaplara sıralanmamış olarak davranılması ve çok terimli bağımlı değişken olarak modellenmesi daha uygun olacaktır.

¹⁷² Vani K. Borooah, **a.g.e.**, p. 7.

¹⁷³ **A.e.**, p. 6.

1.4.3.2 Model Parametrelerinin Tahmin Yöntemleri

Sıralı değişkenler seti; $i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$ iken, Y_i j 'inci kategori içinde ise $Z_{ij} = 1$ ve diğer durumlar için $Z_{ij} = 0$ değerini alacak şekilde tanımlandığında $Z_{ij} = 1$ olma olasılığı şu şekilde hesaplanabilmektedir,¹⁷⁴

$$Pr(Z_{ij} = 1) = \Lambda(\alpha_j - \beta'x_i) - \Lambda(\alpha_{j-1} - \beta'x_i)$$

Gözlemlenen örneklemin olabilirlik fonksiyonu gözlemlerin bağımsız olduğu varsayımı altında olasılıklarının çarpımıdır,

$$L = \prod_{i=1}^n \prod_{j=1}^m [\Lambda(\alpha_j - \beta'x_i) - \Lambda(\alpha_{j-1} - \beta'x_i)]^{z_{ij}}$$

ve log olabilirlik fonksiyonu,

$$L^* = \log L = \sum_{i=1}^n \sum_{j=1}^m z_{ij} \log [\Lambda(\alpha_j - \beta'x_i) - \Lambda(\alpha_{j-1} - \beta'x_i)]$$

şeklinde ifade edilmektedir. Hata teriminin birikimli olasılık yoğunluk dağılımı $F(x) = Pr(\varepsilon_i < x)$ 'dir. Sıralı Lojit Modeller'de hata terimlerinin Lojistik dağılıma uyduğu varsayımı altında, yukarıdaki denklemin sırasıyla bilinmeyen tüm parametreleri için türevleri alınıp, bulunan eşitliklerin tamamı sıfıra eşitlenerek çözüldüğünde, örneklem birimlerinin gözlemlenen En Çok Olabilirlik değerini verecek " $\hat{\beta}$ " parametre tahminleri elde edilir.

L^* 'in ikinci mertebeden kısmi türevlerinden elde edilecek matrisin işaretlerinin tersi alındığında bilgi matrisi elde edilir. Bilgi matrisinin tersi alınarak, tahmin edilen parametrelerin ($\hat{\beta}$) kovaryans tahmin matrisi elde edilerek, istenilen her anlamlılık düzeyi için parametre ve güven aralığı testleri yapılabilmektedir.

¹⁷⁴ G.S. Maddala a.g.e., p. 48.

Pratt (1981)¹⁷⁵, İki Kategorili Probit Model için L*'ın ikinci mertebeden kısmi türevlerinin negatif tanımlı olduğunu göstermiştir. Parametre tahminleri, Newton-Raphson tekrarlı yöntemi ile, olabilirlik fonksiyonunun genellikle 4 ile 10 tekrar arasında global maksimuma yaklaşıldığını söylemişlerdir.

β_k katsayılarının tahmini için $\widehat{\beta}_k$ değeri kullanılmaktadır ve buradan örneklemden her bir birey için $\widehat{Z}_i = \sum_{k=1}^p \widehat{\beta}_k X_{ik}$ tahmini hesaplanır. Tahmin edilen olasılık değerleri şu şekilde yazılmaktadır;¹⁷⁶

$$\hat{p}_{i1} = P(\varepsilon_i \leq \hat{\alpha}_1 - \widehat{Z}_i) = \Lambda(\hat{\alpha}_1 - \widehat{Z}_i)$$

$$\hat{p}_{i2} = P(\hat{\alpha}_1 - \widehat{Z}_i < \varepsilon_i \leq \hat{\alpha}_2 - \widehat{Z}_i) = \Lambda(\hat{\alpha}_2 - \widehat{Z}_i) - \Lambda(\hat{\alpha}_1 - \widehat{Z}_i)$$

⋮

$$\hat{p}_{im} = P(\varepsilon_i \geq \hat{\alpha}_{m-1} - \widehat{Z}_i) = 1 - \Lambda(\hat{\alpha}_{m-1} - \widehat{Z}_i)$$

$\sum_{j=1}^m \hat{p}_{ij} = 1$, tüm $i = 1, 2, \dots, n$ için sağlanmalıdır. Yukarıda tanımlanan model aynı zamanda *Orantılı-Odds* (Proportional Odds) modeli olarak adlandırılmaktadır.

1.4.3.3 Sıralı Lojistik Regresyon Modellerine İlişkin Varsayım

Sıralı Lojistik Regresyon Modelleri'nin, İki Kategorili Lojistik Regresyon Modelleri varsayımlarına ilave olarak sağlaması gereken önemli bir varsayımı da paralel eğriler veya Sıralı Lojit için oransal odds olarak bilinen varsayımdır.

Paralel Eğriler (Orantılı Odds) Varsayımı

Bu varsayım, bağımlı değişkeni etkileyen bağımsız değişkenlerin katsayılarının farklı kategoriler için karşılaştırmalarının tamamında aynı olmasının gerektiğini ifade etmektedir. Eğer varsayım geçersiz ise; belli bir değişkenle ilişkili parametre katsayısı, kategorilerin tamamı için farklı olacaktır. Bu nedenle modelleme için Sıralı Lojit ve Sıralı Probit metotların kullanılması uygun olmayacaktır. Sıralı Lojit

¹⁷⁵ J. W. Pratt, "Concavity of the Log-Likelihood", **Journal of the American Statistical Association**, Vol: 76, 1981, pp. 103–106.

¹⁷⁶ Vani K. Borooah, **a.g.e.**, p. 10.

Modeller içinde paralel eğriler varsayımının uygunluğu Birikimli (Cumulative) Lojit Modeller için aşağıdaki gibi ifade edilmektedir.¹⁷⁷

$$\log \frac{P_j}{1 - P_j} = \alpha_j + \sum_{k=1}^p \beta_k x_{ik}$$

$$P_1 + P_2 + \dots + P_m = 1$$

($j = 1, 2, \dots, m - 1$) olmak üzere $P_j = \Pr(Y_i \leq j)$ 'dir. Bu, bağımlı değişkenin gözlenen değeri Y_i 'nin j . kategori ve bu kategoriden daha düşük kategoriler içinde olma olasılığını ifade etmektedir.

Paralel Eğriler Varsayımının Test Edilmesi

Birikimli lojitler için β regresyon katsayılarının tahminleri yapıp eşit olup olmadıklarına bakılabildiği gibi, paralel ilişki varsayımı için biçimsel testler de uygulanmaktadır. Bunlar; Skor testi, yaklaşık Olabilirlik Oran (LR) testi ki bunlar çok amaçlı testlerdir, eğer sadece bazı bağımsız değişkenler için problem varsa bunların hangileri olduğu saptanamaz. Wald testi ise, β tahminlerin hem tümünün birden hem de ayrı ayrı test edilmesine izin vermektedir.¹⁷⁸

Olabilirlik Oran testi, paralel eğri varsayımının geçerliliğini veriler üzerinden Çok Terimli Lojit Modellerin tahmini ile test etmektedir. Çok Terimli Lojit Modeller $j = 1, 2, \dots, m$ kategoriler arasında tahmin edilen eğim katsayıları β_k 'nin farklı olmasına izin vermektedir. Sıralı Lojit Modeller p tane parametreyi tahmin ederken, Çok Terimli Lojit Modeller $p(m - 1)$ tane parametre değerini tahmin eder. Eğer L_1 , Sıralı Lojit Modeller için log-olabilirlik değeri ve L_2 , Çok Terimli Lojit Modeller için log-olabilirlik değeri olarak kabul edilirse; buradan $2(L_2 - L_1)$ değeri hesaplanıp $p(m - 2)$ serbestlik dereceli χ^2 tablo değeri ile karşılaştırılarak yorumlanabilir. Ancak burada şunun farkında olunmalıdır ki bu, doğrudan Olabilirlik Oran testi değildir. Çünkü Sıralı Lojit Modeller, Çok Terimli Lojit Modeller içinde

¹⁷⁷ Ann A. O'Connell, **Logistic Regression Models for Ordinal Response Variables**, Sage Publication, 2006, pp. 30–31.

¹⁷⁸ Simon Cheng, J. Scott Long, **a.g.e.**, p. 270.

yer almamaktadır. Bu test sadece fikir vermek içindir: “çok büyük” bir χ^2 değeri paralel eğim varsayımının sağlanmadığı yönünde endişe uyandırırken, “kabul edilebilir büyüklükteki bir değer endişe için bir temel sağlamamaktadır.”¹⁷⁹

1.4.3.4 Uyum İyiliği Testleri

Genelde hipotezde, modeldeki bağımsız değişkenlerin katsayılarının sıfır olduğu durum ile ilgilenilmekte ve “tam” model ile “sadece sabit” değeri içeren modelin karşılaştırması olan $2(L_1 - L_0)$ değeri raporlanmaktadır. Burada L_0 , açıklayıcı değişken olarak sadece sabit terim modelde olduğu zaman hesaplanan log-olabilirlik fonksiyonunun değeri ve L_1 , açıklayıcı değişkenlerin tümü modelde olduğu zaman hesaplanan log-olabilirlik fonksiyonunun değeridir. Buradan hesaplanan serbestlik derecesi değeri ile tahmin edilen eğim katsayısını veren χ^2 değeri karşılaştırılarak modelin uyum iyiliği testleri yapılabildiği gibi, bazı “*Pseudo - R²*” değerlerinin sonuçları da modelin uygunluğu için kullanılabilir.

Bu “*Pseudo - R²*” değerlerinden birkaçı şunlardır:

- 1973’de McFadden¹⁸⁰ tarafından önerilen “*Pseudo - R²*” değeridir bu değer hesaplanan log-olabilirlik fonksiyonları değerlerinin oranlarına dayanmaktadır ve *Pseudo - R²* değeri şu şekilde hesaplanmaktadır,

$$R_{MCF}^2 = 1 - \frac{L_1}{L_0}$$

- 1975’de McKelvey ve Zavoina¹⁸¹ modelin uyum iyiliği ölçüsü olarak en uygununun hesaplanma şeklinin, bağımlı değişkenin sıralı alternatiflerden oluştuğu durum için, regresyon modelinde kullanılan R^2 değerinin aşağıdaki gibi uyarlanarak hesaplandığı *Pseudo - R²* değeri olduğunu söylemişlerdir,

¹⁷⁹ Vani K. Borooah, **a.g.e.**, p. 15.

¹⁸⁰ D. McFadden, “Conditional Logit Analysis of Quantitative Choice Behavior”, Edited by Zarembka, P., **Frontiers in Econometrics**, New York Academic Press, 1973, p. 121.

¹⁸¹ R.D. McKelvey, W. Zavoina, “A Statistical Model for the Analysis of Ordinal Level Dependent Variable”, **Journal of Mathematical Sociology**, Vol: 4, 1975, pp. 103–120.

$$R_{M\&Z}^2 = \frac{\widehat{Var}(\hat{y}^*)}{\widehat{Var}(y^*)} = \frac{\widehat{Var}(\hat{y}^*)}{\widehat{Var}(\hat{y}^*) + \widehat{Var}(\hat{\varepsilon})}$$

Bu formül iki nedenle Doğrusal Regresyon Modelinden farklıdır. İlk olarak, gözlemlenen y yerine gizli (latent) değişken y^* 'dan tahmin edilen varyanslar kullanılmaktadır. \hat{y}^* 'ın varyansı En Çok Olabilirlik yönteminden elde edilen β tahmin değerleri kullanılarak şu şekilde hesaplanmaktadır,¹⁸²

$$\widehat{Var}(\hat{y}^*) = \hat{\beta}'\widehat{Var}(x)\hat{\beta} + Var(\varepsilon)$$

$\widehat{Var}(x)$, x 'lerden oluşan tahmini kovaryans matrisidir. İkinci olarak da ε 'un varyansı varsayım gereği tahmin edilebilir olmak yerine, sabittir. Lojit model için, $Var(\varepsilon) = \pi^2/3$ ve Probit model için $Var(\varepsilon) = 1$.

1992'de yaptıkları simülasyon çalışmasında Hagle ve Mitchell,¹⁸³ McKelvey ve Zavoina tarafından önerilen $R_{M\&Z}^2$ değerinin modelin uyum iyiliği kararında kullanılmasını önermişlerdir.

Yukarıda verilen *Pseudo* – R^2 değerleri [0, 1] aralığında tanımlı olmakla birlikte, daha önce de vurgulandığı gibi “0” ile “1” dışındaki değerlerin doğal bir yorumu yoktur. Bu değer, *Pseudo* – R^2 değeri yükseldikçe modelin uygunluğunun artmakta olduğunu ifade etmektedir. Uyum iyiliği ölçüsü olarak R^2 değeri dışında doğru tahmin olasılıklarının ortalama değerlerinin kullanılması da mümkündür.

1.4.3.5 Model Parametrelerinin Yorumu

Sıralı Lojit Modellerin sonuçlarının yorumlanmasında marjinal etkiler kullanılabilirliği gibi yorumlanmasının daha anlamlı olacağı düşünülen OR'de kullanılabilirlikte.

Marjinal değişim, bağımsız değişkenin olasılık eğrisinin yaklaşık doğrusal olduğu bir bölgesi üzerinde değişmediği sürece, herhangi bir bağımsız değişkendeki (x_k) bir

¹⁸² J. Scott Long, Simon Cheng, **a.g.e.**, p. 271.

¹⁸³ T.M. Hagle, G.E. Mitchell II, “Goodness-of-fit Measures for Probit and Logit”, **American Journal of Political Science**, Vol:36, No:3, 1992, pp. 762–784.

birimlik deęişim için gözlenen olasılıklardaki deęişimi göstermez. Eđim yaklaşık olarak doğrusal ise, marjinal etki bağımsız deęişkendeki bir birimlik deęişikliđin bağımlı deęişkenin olasılığı üzerindeki etkisini özetlemekte kullanılabilir. ¹⁸⁴ Sıralı Lojit Modellerin olasılıklarının marjinal etkisi x_k 'ya göre kısmi türevleri alınarak gösterilmektedir. Genel olarak herhangi bir deęişkene ait kısmi türev şu şekilde ifade edilebilir, ¹⁸⁵

$$\frac{\partial P(y = j)}{\partial x_k} = \left[f \left(\alpha_{j-1} - \sum_{k=1}^p \beta_k x_k \right) - f \left(\alpha_j - \sum_{k=1}^p \beta_k x_k \right) \right] \beta_k$$

Burada $f(\cdot)$, olasılık yoğunluk fonksiyonunu ifade eder. $\Lambda'(x) = d\Lambda(x)/dx$, Lojistik dağılımın olasılık yoğunluk fonksiyonunu ifade edecek şekilde Lojistik dağılımın yoğunluk fonksiyonu şu şekildedir, ¹⁸⁶

$$P(X \leq x) = \Lambda(x) = \frac{\exp(x)}{[1 + \exp(x)]} = \frac{1}{[1 + \exp(-x)]}$$

$$\Lambda'(x) = \frac{d}{dx} \left[\frac{\exp(x)}{[1 + \exp(x)]} \right] = \frac{[1 + \exp(x)] \exp(x) - [\exp(x)]^2}{[1 + \exp(x)]^2} = \Lambda(x)[1 - \Lambda(x)]$$

Sürekli bağımsız deęişken x_{ik} deęerindeki küçük bir deęişikliđin i . birey için y 'nin alabileceđi kategorik deęerlerin olasılık üzerindeki marjinal etkisi Lojistik dağılım altında şu şekilde hesaplanmaktadır, ¹⁸⁷

$$\frac{\partial P(Y_i = 1)}{\partial x_{ik}} = -\Lambda' \left(\alpha_j - \sum_{k=1}^p \beta_k x_k \right) \beta_k$$

$$\frac{\partial P(Y_i = j)}{\partial x_{ik}} = \left[\Lambda' \left(\alpha_{j-1} - \sum_{k=1}^p \beta_k x_k \right) - \Lambda' \left(\alpha_j - \sum_{k=1}^p \beta_k x_k \right) \right] \beta_k$$

$$\frac{\partial P(Y_i = m)}{\partial x_{ik}} = \Lambda' \left(\alpha_j - \sum_{k=1}^p \beta_k x_k \right) \beta_k$$

¹⁸⁴ Simon Cheng, J. Scott Long, **a.g.e.**, p. 272.

¹⁸⁵ Tim Futing Liao, **a.g.e.**, p. 45.

¹⁸⁶ Vani K. Borooah, **a.g.e.**, p. 13.

¹⁸⁷ **A.e.**, pp. 12–13.

Tanımlanan k . değişkenin değerindeki küçük bir miktar artış, $\beta_k > 0$ olması durumunda Lojit Modelin olasılık değerlerini düşürmeyecek demektir. $P(Y_i = 1)$ 'in türevi β_k ile ters işaretlidir. $P(Y_i = m)$ 'nin türevi β_k ile aynı işaretli olduğu için değeri artmaktadır. Bununla beraber, birinci ve sonuncu kategoride olma olasılığı dışındaki olasılık değerlerinin ne olduğu açık değildir. Bu değerler artabilir, düşebilir veya değişmeyebilirler. Sadece iki uç değer olasılıklarının yönündeki değişiklik açıkça belirlenebilmektedir.¹⁸⁸

Bunun yanı sıra marjinal etkilerin yorumlamada kullanılması, tanımlanan değişkenler sürekli olduğu zaman uygun olsa da, bağımsız değişkenin kategorik tanımlanmış olduğu zaman yanıltıcı olabilir. Kategorik bağımsız değişkenin marjinal etkisini bulabilmek için değişkenin aldığı herhangi iki değer arasındaki olasılık farklarının hesaplanarak değerlendirilmesi gerekmektedir. Sürekli değişkenlerin ölçümlerinde ister kısmi türevler ister tahmin edilen olasılıkların farkları alınsın iki durumda da sonuçlar yaklaşık olarak aynıdır.

Kategorik bağımsız değişkendeki değişim için başlangıç değeri x_S ve bitiş değeri x_E olacak şekilde x_k değişkeni için değişimin tahmin edilen olasılığı şu şekildedir,¹⁸⁹

$$\frac{\Delta \Pr(y = j|x)}{\Delta x_k} = \Pr(y = j|x, x_k - x_E) - \Pr(y = j|x, x_k - x_S)$$

Bu eşitlik diğer x 'ler sabitken, x_k , x_S 'den x_E 'ye değiştiği zaman bağımlı değişkenin beklenen olasılığı kadar değişir. Kategorik bağımsız değişkenin değeri x_k 'nin başlangıç değerine x_k 'daki değişim miktarına ve diğer değişkenlerin değerine bağlıdır.

Bağımsız değişken için m tane kategorik değişim katsayısı, tüm bağımlı değişken kategorilerindeki değişim değerlerinin mutlak ortalamaları tarafından hesaplanarak özetlenebilmektedir;¹⁹⁰

¹⁸⁸ Vani K. Borooah, **a.g.e.**, pp. 13–14.

¹⁸⁹ J. Scott Long, Simon Cheng, **a.g.e.**, p. 272.

¹⁹⁰ Simon Cheng, J. Scott Long, **a.g.e.**, p. 273.

$$\bar{\Delta} = \frac{1}{m} \sum_{j=1}^m \left| \frac{\Delta \Pr(y = j|x)}{\Delta x_k} \right|$$

$\bar{\Delta}$, ortalama mutlak kesikli deęişim (the average absolute discrete change) deęeri olarak adlandırılmaktadır.

Ayrıca hesaplanan sonuçların doğruluęunu kontrol etmek için olasılık toplamlarının “1” olması gerektięi gibi, marjinal etkilerin toplamlarının da “0” olması koşulunun kontrol edilmesi oldukça faydalıdır.

Sıralı Lojit Modeller, anlamlı katkıda bulunacaęı düşünülerek, lojit tahminlerin yorumlanmasında kullanılmaktadır. Paralel eğriler varsayımı geçerli olduęunda, x 'in üzerindeki kısmi etkisi baęımlı deęişkeninin kategorisi ne olursa olsun deęişmeyecek demektir.

$$\log \left[\frac{P(y \leq j|x)}{1 - P(y \leq j|x)} \right] = \alpha_j - \sum_{k=1}^p \beta_k x_k$$

Paralel eğriler varsayımı sağlanıyorsa, yukarıdaki eşitlikten tahmin edilen deęişkenlere ait parametre katsayıları β 'lar baęımlı deęişkenin her kategorisi için aynı olacaktır. Baęımlı deęişkeninin üç kategorili olduęu bir durum ele alındıęında, yukarıdaki eşitlikteki x_k deęişkeninin etkisi dięer herşey aynıyken kategori 1'in kategori 2 ve 3'e veya kategori 1 ve 2'nin kategori 3'e göre ortaya çıkma odds deęeri x_k 'daki bir birimlik artış için $\exp(\beta_k)$ katı kadar artacak demektir.¹⁹¹

Modelin parametre tahminlerinin yorumlanmasında dięer bir yol da, x_k 'daki tahmin edilen β_k parametresindeki bir standart sapmalık artışın dięer deęişkenler sabitken odds deęerinde ortaya çıkartacaęı artış miktarının yüzdesel olarak ifade edilebilmesidir. Bu deęer şu şekilde hesaplanmaktadır,¹⁹²

$$100[\exp(-s_k \times \beta_k) - 1] \quad \mathbf{1.29}$$

¹⁹¹ Tim Futing Liao, **a.g.e.**, p. 43.

¹⁹² Simon Cheng, J. Scott Long, **a.g.e.**, p. 273.

Bu deęerin yorumu baęımlı deęiřkeninin üç kategoriden oluřtuęu bir durum ele alındıęı dūřınılduęunda, x_k 'daki bir standart sapmalık artıř, kategori 2 ve 3'e karřı kategori 1'in odds deęerini dięer deęiřkenler sabit tutulurken, Eřitlik 1.29'dan "hesaplanan deęer" kadar arttırmaktadır. Aynı řekilde, kategori 3'e karřı kategori 1 ve 2'nin odds deęeri x_k 'daki her bir standart sapmalık artıř için "hesaplanan deęer" kadar arttıracak demektir.

1.4.4 Sıralı Probit Modeller

Sıralı Probit Modellerin Sıralı Lojit Modellerden tek farkı hata terimi ε_i 'lerin varsayılan daęılımlarından kaynaklanmaktadır. Sıralı Lojit Modeller ε_i 'nin sıfır ortalama ve $\pi^2/3$ varyansı ile Lojistik daęılıma uyduęunu varsayarken, Sıralı Probit Modeller ise ε_i 'nin sıfır ortalama ve 1 varyansla Normal daęıldığını varsayar. Lojistik daęılım kuyruklar dıřında normal daęılıma benzer bir daęılım olup, tek farkı Normal daęılıma gōre kuyruklarda daha yoęun olmasındır. Bunun dıřında tahmin ve parametre yorumları Sıralı Lojistik Regresyon Modelleri için yapılan yorumlarla benzerlik gōstermektedir.

Sıralı deęiřkenler setin de $i = 1, 2, \dots, n; j = 1, 2, \dots, m$ olacak řekilde deęerler alıyorken $y_i, j.$ kategori iinde ise $z_{ij} = 1$ ve dięer durumlar iin $z_{ij} = 0$ deęerini alacak řekilde tanımlandığında $z_{ij} = 1$ olma olasılıęı Sıralı Probit Modelleri iin ařaęıdaki gibi hesaplanabilmektedir,¹⁹³

$$Pr(z_{ij} = 1) = \Phi(\alpha_j - \beta'x_i) - \Phi(\alpha_{j-1} - \beta'x_i)$$

Φ , Birikimli Standart Normal daęılımı gōstermektedir. Genel olarak $P(y = j)$, Sıralı Probit Modeller kullanılarak birbiri ardına gelen iki birikimli olasılıęın farkı alınarak elde edilmektedir. Őzel durumlar olan ilk ve son kategori iinse $P(y \leq 1) = P(y = 1)$ ve $P(y \leq m) = 1$ 'e eřittir. Buradan modelin olabilirlik fonksiyonu,¹⁹⁴

$$L = \prod_{i=1}^n \prod_{j=1}^m [\Phi(\alpha_j - \beta'x_i) - \Phi(\alpha_{j-1} - \beta'x_i)]^{z_{ij}}$$

¹⁹³ J. Scott Long, **a.g.e.**, p. 121.

¹⁹⁴ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, p. 291.

ve log-olabilirlik fonksiyonu

$$L^* = \log L = \sum_{i=1}^n \sum_{j=1}^m z_{ij} \log[\Phi(\alpha_j - \beta' x_i) - \Phi(\alpha_{j-1} - \beta' x_i)]$$

şeklinde ifade edilmektedir.

$\frac{\partial \Phi(x)}{\partial x} = \phi(x)$ ve $\frac{\partial \phi(x)}{\partial x} = -x\phi(x)$ olduğu biliyorken, eğer $j = m$ ise $\delta_{jm} = 1$ ve diğer durumlarda $\delta_{jm} = 0$ iken, log-olabilirlik fonksiyonunun kısmi türevleri alındığında ve sıfıra eşitlendiğinde,¹⁹⁵

$$\frac{\partial L^*}{\partial \beta} = \sum_{i=1}^n \sum_{j=1}^m z_{ij} \frac{\phi_{i,j-1} - \phi_{ij}}{\Phi_{ij} - \Phi_{i,j-1}} x_i = 0$$

$$\frac{\partial L^*}{\partial \alpha_k} = \sum_{i=1}^n \sum_{j=1}^m z_{ij} \frac{\delta_{j,m} \phi_{i,j-1} - \delta_{j-1,m} \phi_{ij}}{\Phi_{ij} - \Phi_{i,j-1}} = 0$$

denklemleri elde edilmektedir. Bu denklemler Newton-Raphson yöntemi kullanılarak parametre tahminleri için tekrarlı yöntemle çözümler. L^* 'ın ikinci mertebeden kısmi türevleri aşağıdaki denklemler ile ifade edilmektedir.¹⁹⁶

$$\frac{\partial^2 L^*}{\partial \beta \partial \beta'} = \sum_{i=1}^n \sum_{j=1}^m \frac{z_{ij}}{(\Phi_{ij} - \Phi_{i,j-1})^2} [(\Phi_{ij} - \Phi_{i,j-1})(Y_{i,j-1} \phi_{i,j-1} - Y_{i,j} \phi_{ij}) - (\phi_{i,j-1} - \phi_{ij})^2] x_i x_i'$$

$$\frac{\partial^2 L^*}{\partial \beta \partial \alpha_k} = \sum_{i=1}^n \sum_{j=1}^m \frac{z_{ij}}{(\Phi_{ij} - \Phi_{i,j-1})^2} [(\Phi_{ij} - \Phi_{i,j-1})(Y_{i,j} \phi_{ij} \delta_{j,m} - Y_{i,j-1} \phi_{i,j-1} \delta_{j-1,m}) - (\phi_{i,j-1} - \phi_{ij})(\phi_{i,j} \delta_{j,m} - \phi_{i,j-1} \delta_{j-1,m})] x_i$$

¹⁹⁵ G. S. Maddala, **a.g.e.**, p. 48.

¹⁹⁶ **A.e.**, pp. 48-49.

$$\frac{\partial^2 L^*}{\partial \alpha_k \partial \alpha_l} = \sum_{i=1}^n \sum_{j=1}^m \frac{z_{ij}}{(\Phi_{i,j} - \Phi_{i,j-1})^2} [(\Phi_{i,j} - \Phi_{i,j-1})(Y_{i,j-1} \phi_{i,j-1} \delta_{j-1,m} \delta_{j-1,l} - Y_{i,j} \phi_{i,j} \delta_{j,m} \delta_{j,l}) - (\phi_{i,j} \delta_{j,m} - \phi_{i,j-1} \delta_{j-1,m})(\phi_{i,j} \delta_{j,l} - \phi_{i,j-1} \delta_{j-1,l})]$$

İki Kategorili Probit Modeller için L^* 'ın ikinci mertebeden kısmi türevlerinin negatif tanımlı olduğu 1981 yılında Pratt¹⁹⁷ tarafından gösterilmiştir. Newton-Raphson tekrarlı yöntemi ile olabirlik fonksiyonunun 4 ile 10 tekrar arasında global maksimumuna yaklaştığı görülmektedir.¹⁹⁸

Parametre tahminleri yorumlanırken; nicel bağımsız değişken x_{ik} 'nin değerindeki küçük bir değişikliğin i . kişi için y 'nin alabileceği kategorik değerlerin olasılık üzerindeki marjinal etkisini ölçmek için, Sıralı Lojit Modellerde olduğu gibi, Normal dağılımdan elde edilen olasılık değerlerinin farklarından hesaplanan sonuçlar kullanılmaktadır.¹⁹⁹

$$\frac{\partial P(Y_i = 1)}{\partial x_{ik}} = -\Phi' \left(\alpha_1 - \sum_{k=1}^p \beta_k x_k \right) \beta_k$$

$$\frac{\partial P(Y_i = j)}{\partial x_{ik}} = \left[\Phi' \left(\alpha_{j-1} - \sum_{k=1}^p \beta_k x_k \right) - \Phi' \left(\alpha_j - \sum_{k=1}^p \beta_k x_k \right) \right] \beta_k$$

$$\frac{\partial P(Y_i = m)}{\partial x_{ik}} = 1 - \Phi' \left(\alpha_{m-1} - \sum_{k=1}^p \beta_k x_k \right) \beta_k$$

Burada $\Phi'(x) = d\Phi(x)/dx$ 'i ifade etmektedir.

Sıralı Probit Modellerde de, Sıralı Lojit Modellerde olduğu gibi, bağımsız değişken nicel değer almadığında marjinal etkilerin ölçülmesi için kısmi türevinin alınması uygun olmamaktadır. Onun yerine Lojit Modellerde olduğu gibi kukla değişkenin

¹⁹⁷ J. W. Pratt, **a.g.e.**, pp. 103–106.

¹⁹⁸ G.S. Maddala **a.g.e.**, p. 49.

¹⁹⁹ Vani K. Borooah, **a.g.e.**, p. 12.

etkisi yorumlanırken, değişkenin alacağı herhangi iki değer arasındaki farkın karşılaştırılması ile elde edilen olasılık sonucunun kullanılması daha uygundur.

Ayrıca parametre tahmin değerlerinin yorumlanmasında OR, Sıralı Lojit Modellerden tahmin edilen parametre değerlerinin Sıralı Probit Modellerden elde edilen tahmin değerlerin yaklaşık olarak 1.6 ile 1.8 katı kadar olması dışında benzerdir²⁰⁰.

Hem Sıralı Lojit hem de Sıralı Probit modeller SAS ve STATA bilgisayar programları ile tahmin edilebilmektedirler. SAS ve STATA uygulamasında sıralı modellerin tahmininde olasılıklar hesaplanırken kesim değerleri farklı hesaplanmaktadır. SAS²⁰¹ kesim değerlerini hesaplarken modelin sabit değerini ilave ederek modeli vermekte, STATA²⁰² ise kesim değeri ve sabit terimi ayrı ayrı vermektedir. Kesim değerlerinin farklı hesaplanması En Çok Olabilirlik Tahminlerini (Maksimum Likelihood Estimation) etkilememektedir. Ayrıca SAS uygulamalarında PROC LOGISTIC ile paralel eğriler varsayımını test eden Score test ve Çok Terimli Lojit ve Sıralı Lojit için olabilirlik değerlerinin karşılaştırmasını yapan χ^2 değeri de verilmektedir²⁰³. SAS ve STATA dışında Sıralı Probit Modeller için LIMDEP de kullanılabilir.

1.4.5 Çok Kategorili Bağımlı Değişken için Diğer Modeller

Sosyal bilimlerde çok kategorili bağımlı değişkenler için en sık kullanılan yöntemlerden yukarıda bahsedilmiştir. Ancak bu yöntemlerin, modellerin sağlaması gereken varsayımları sağlamaması durumunda kullanılması çok da uygun olmamaktadır.

Çok Terimli Lojit Modeller için IIA varsayımının sağlanmaması durumunda Çok Terimli Probit veya İç içe Lojit modeller kullanılabilir. Bu modele bazı kaynaklarda Bitişik Kategorili (Adjacent Category) Modeller adı verilmektedir. Paralel eğriler

²⁰⁰ J. Scott Long, **a.g.e.**, p.129.

²⁰¹ Tim Futing Liao, **a.g.e.**, p. 41.

²⁰² Vani K. Borooah, **a.g.e.**, p. 11.

²⁰³ Alfred Demaris, **Logit Modeling**, Sage University Paper Series on Quantitative Applications in the Social Sciences, 1992, p. 76.

varsayımı sağlanmadığında ise, literatürde genellikle Birikimli (Cumulative) Sıralı Lojit ve Birikimli Sıralı Probit modeller olarak adlandırılan Sıralı Lojit ve Probit modeller yerine bağımlı değişkenin sıralı düzende olduğu durumlar için farklı modeller önerilmektedir. Yapılan çalışmalarda en çok kullanılan modeller, Sürekli Oran (Continuation Ratio) ve Oransal Odds (Proportional Odds) modelleridir. J. Scott Long ve Simon Cheng²⁰⁴, Sürekli Oran ve Oransal Odds modelleri ile Stereo tip Lojistik modellere yer verirken, paralel eğriler varsayımının sağlanmaması durum için alternatif olarak, sıralı bağımlı değişkeninin çok terimli bağımlı değişkenin modellenmesine benzer şekilde ele alınmasını önermişlerdir²⁰⁵.

Bitişik Kategoriler Modeli (Adjacent Categories Model)

Çok Terimli Lojit Modelin özel bir durumudur. Bu modellerin taşıması gereken en önemli varsayım, bağımlı değişkenin kavramsal olarak farklı ve istatistik olarak birbirinden bağımsız olmasıdır. Bitişik Kategoriler Modeli, bağımlı değişkenin seviyelerinin her birinin kendisinden daha büyük olanla karşılaştırmasına olanak vermektedir. Buradaki varsayım log-odds cevaplayıcıya bağlı olmadığı ve Bitişik Kategorili Lojitlerin (log-odds) parametreleri bakımından doğrusal olduğudur.²⁰⁶

$$\Omega_{j+1|j}(x) = \frac{\Pr(y = j + 1|x)}{\Pr(y = j|x)} = \exp(\alpha_{j+1} + x\beta) \quad \mathbf{1.30}$$

Burada, diğer değişkenler sabit tutulduğunda, x 'deki bir birimlik artışın odds değerinde ($\Omega_{j+1|j}$), $\exp(\beta)$ faktörlük değişime neden olacağı gösterilmektedir.

Eşitlik 1.30'dan tahmini olasılıklar, aşağıdaki eşitliklerin çözülmesi ile elde edilmektedir;

$$\Pr(y = j + 1|x) = \frac{\exp(\alpha_{j+1} + x\beta)}{1 + [\exp(\alpha_{j+1} + x\beta)]} \quad j = 1, 2, \dots, m - 1$$

²⁰⁴ J. Scott Long, Simon Cheng, **a.g.e.**, pp. 274–276.

²⁰⁵ **A.e.**, p. 274.

²⁰⁶ Ann A. O'Connell, **a.g.e.**, p. 76.

Bitişik Kategorili Lojit Modeller İki Kategorili Lojit Modellerle şu şekilde ifade edilmektedir.

$$\begin{aligned} \ln \left[\frac{P(y = m|x)}{P(y = 0|x)} \right] &= \ln \left[\frac{P(y = 1|x)}{P(y = 0|x)} \right] + \ln \left[\frac{P(y = 2|x)}{P(y = 1|x)} \right] + \dots + \ln \left[\frac{P(y = m|x)}{P(y = m-1|x)} \right] \\ &= (\alpha_1 + x\beta) + (\alpha_2 + x\beta) + \dots + (\alpha_m + x\beta) \\ &= (\alpha_1 + \alpha_2 + \dots + \alpha_m) + (mx\beta) \end{aligned}$$

Yukarıdaki eşitlikte Bitişik Kategorili Modellerin sabit $\beta_{m0} = (\alpha_1 + \alpha_2 + \dots + \alpha_m)$ ve parametre katsayılarını veren $\beta_m = m\beta$ ile İki Kategorili Lojistik Regresyon Modellerinin bir versiyonu olduğu gösterilebilmektedir.²⁰⁷

Devam Eden Oranlı Model (The Continuation Ratio Model)²⁰⁸

Devam Eden Oranlı Model 1980 yılında Fienberg tarafından önerilmiştir ve sıralı bağımlı değişkenlerin tasarımında kategoriler, süreçteki her bir ilerlemenin aşamalarını betimlemektedir. Örneğin; bağımlı değişken öğretim üyeliğinin sırası, aşamalar da asistanlık, doçentlik ve profesörlük ise, sürecin önemli özelliği her bir bireyin her aşamayı teker teker geçmesinin gerekmesidir.

Eğer $P(y = j|x)$, x biliniyorken aşama j 'de olma olasılığı ise, aşama j veya sonrasında olma olasılığı $\Pr(y \geq j|x) = \sum_{j=1}^m \Pr(y = j|x)$ şeklindedir.

Aşama j veya daha sonrasında olduğunun bilinmesi koşulu altında, aşama j 'nin içinde olma olasılığı aşağıdaki gibi ifade edilebilmektedir,*

$$\Pr(y = j | y \geq j, x) = \frac{\Pr(y = j|x)}{\Pr(y \geq j|x)} \quad j = 1, 2, \dots, m - 1$$

Buradan aşama j 'nin sonrasında olma olasılığında,²⁰⁹

²⁰⁷ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, p. 289.

²⁰⁸ J. Scott Long, Simon Cheng, **a.g.e.**, p. 276.

* Koşullu Olasılık değeri $P(A|B) = P(A \cap B)/P(B)$ gibi ifade edilmektedir.

²⁰⁹ Ann A. O'Connell, **a.g.e.**, p. 30.

$$\Pr(y > j | y \geq j, x) = \frac{\Pr(y > j | x)}{\Pr(y \geq j | x)} = 1 - \Pr(y = j | y \geq j, x)$$

Bu olasılıkları kullanarak, cevaplayıcının aşama j ya da sonrasında olduğu biliniyorken aşama j 'de olmasına karşın aşama j 'yi geçmiş olma odds değeri aşağıdaki eşitlikten hesaplanabilmektedir,²¹⁰

$$\frac{\Pr(y = j | y \geq j, x)}{\Pr(y > j | y \geq j, x)} = \frac{\Pr(y = j | x)}{\Pr(y > j | x)}$$

Odds değerinin logaritması alınarak, model şu şekilde dönüştürülür,

$$\log \left[\frac{\Pr(y = j | x)}{\Pr(y > j | x)} \right] = \alpha_j + x\beta \quad j = 1, 2, \dots, m - 1$$

Burada eşitliğin iki tarafının antilogaritması alınırsa sağ tarafta bulunan $\exp(\beta_k)$ değeri, bireyin aşama j veya daha sonrasında olduğu bilindiğinde, x_k 'daki bir birimlik artışın, j . aşamada olmaya karşın daha yüksek kategoride olmasının odds değerleri diğer değişkenler sabit tutulduğundaki etkisi olarak yorumlanabilecektir.

$y = 1$ olma olasılığı İki Kategorili Lojit Modellerde olduğu gibi, $\frac{\Pr(y=1|x)}{\Pr(y>1|x)} = \exp(\alpha_1 + x\beta)$ eşitliğinden hesaplanabilmektedir;²¹¹

$$\Pr(y = 1 | x) = \frac{\exp(\alpha_1 + x\beta)}{1 + \exp(\alpha_1 + x\beta)}$$

Buradan $y = 2$ olma olasılığı,

$$\Pr(y = 2 | x) = \frac{\exp(\alpha_2 + x\beta)}{[1 + \exp(\alpha_2 + x\beta)][1 + \exp(\alpha_1 + x\beta)]}$$

eşitliğinden hesaplanabilmektedir. Genel olarak $y = j$ olma olasılığı,

$$\Pr(y = j | x) = \frac{\exp(\alpha_j + x\beta)}{\prod_{j=1}^{m-1} [1 + \exp(\alpha_j + x\beta)]} \quad j = 1, 2, \dots, m - 1$$

²¹⁰ J. Scott Long, **a.g.e.**, p. 146.

²¹¹ J. Scott Long, Simon Cheng, **a.g.e.**, p. 276.

$$\Pr(y = m|x) = 1 - \sum_{j=1}^{m-1} \Pr(y = j|x)$$

olarak gösterilmektedir. Tahmin edilen bu olasılıklar, modelin yorumlanması için kullanılabilirler.

Stereotype Sıralı Regresyon Modeli (The Stereotype Ordered Regression Model)

Sıralı Lojistik Regresyon Modellerinin paralel eğriler varsayımına karşılık, 1984 yılında Anderson²¹² tarafından Stereotype Sıralı Regresyon Modeli önerilmiştir. Bu model şu şekilde formüle edilmektedir,

$$\log \left[\frac{\Pr(y = j)}{\Pr(y = m)} \right] = (\alpha_j - \alpha_m)\beta_0 + (\phi_j - \phi_m)x\beta \quad j = 1, 2, \dots, m$$

α 'lar ve ϕ 'ler bağımlı değişken kategorileri ile ilişkili ölçek faktörleridir. β_0 sabit terim ve β bağımsız değişkenlerle ilişkili katsayı vektörüdür.

Bu model, eşitliğin sol tarafındaki bağımlı değişken değerinin kullanılan ikisine bağlı olarak her bir bağımsız değişken ile ilişkili katsayıların farklı olmasına izin vermektedir. Aynı şekilde α 'lar da her bir bağımlı değişken çifti için sabit değerlerin farklı olmasına izin vermektedir. Model kullanıldığında, çok fazla sınırlanmamış α ve ϕ değerleri bulunmasından dolayı değerler olması gerekenden çok farklı çıkabilir. Modelin tanımlanması için çok farklı yollar vardır. Örneğin; $\phi_1 = 1$, $\phi_m = 0$, $\alpha_1 = 1$, $\alpha_m = 0$ olduğu varsayılabilir veya sıralı bağımlı değişken değerleri için log-doğrusal modellerin yaklaşımları kullanılır. $\sum_{j=1}^m \phi_j = 0$ ve $\sum_{j=1}^m \phi_j^2 = 1$ sınırlaması ile model tanımlanabilir.²¹³

Buraya kadar bahsedilen modellerde herhangi bir sıra sınırlaması istenmemekle birlikte Stereotype Modellerde genellikle tercih edilen budur. Stereotype Modellerde, bağımlı değişkenin sırasını sağlayacak şekilde, ilave ($\phi_1 = 1 > \phi_2 > \dots > \phi_{m-1} > \phi_m = 0$) sınırlaması da sağlanmalıdır.

²¹² J. A. Anderson, "Regression and Ordered Categorical Variables", **Journal of the Royal Statistical Society Series B**, Vol: 46, No: 1, pp. 1–30.

²¹³ J. Scott Long, Simon Cheng, **a.g.e.**, pp. 274–275.

Eşitlik $\log \left[\frac{\Pr(y=j)}{\Pr(y=m)} \right] = (\alpha_j - \alpha_m)\beta_0 + (\phi_j - \phi_m)x\beta$ 'da aşağıdaki gibi hesaplanan tahmin olasılıkları kullanabilmektedir,²¹⁴

$$\Pr(y = j|x) = \frac{\exp(\alpha_j\beta_0 + \phi_jx\beta)}{\sum_{j=1}^m \exp(\alpha_j\beta_0 + \phi_jx\beta)} \quad j = 1, 2, \dots, m$$

Bu model aynı zamanda m 'ye karşı j bağımlı değişken değerinin olasılığının x_k 'daki değişim etkisi açısından da yorumlanmasına izin verir.

Odds değerleri açısından $\log \left[\frac{\Pr(y=j)}{\Pr(y=m)} \right]$ eşitliği tekrar yazıldığında,

$$\Omega_{j|m}(x, x_k) = \frac{\Pr(y=j)}{\Pr(y=m)} = \exp[(\alpha_j - \alpha_m)\beta_0 + (\phi_j - \phi_m)x\beta]$$

eşitliği elde edilmektedir.

Buradan, $\frac{\Omega_{j|m}(x, x_{k+1})}{\Omega_{j|m}(x, x_k)} = e^{(\phi_j - \phi_m)\beta_k} = \left(\frac{e^{\phi_j}}{e^{\phi_m}} \right)^{\beta_k}$ değeri hesaplanmaktadır.

Yukarıdaki eşitlikten anlaşılacağı üzere, m 'ye karşı j 'nin odds değeri üzerinde x_k 'nın etkisi, ölçekleme katsayısı ϕ 'ye göre tüm bağımlı değişken değerlerinin karşılaştırılmasında farklıdır.²¹⁵

1.5 Lojistik Regresyon Modellerinin Örnek Büyüklüğü

Lojistik Regresyon Modelleri için örnek büyüklüğü belirlemekle ilgili literatürde çok az çalışma bulunmaktadır. Lojit Modellerin değerlendirilmesinde temel unsurlardan biri, her bir bağımsız değişken için yeterli sayıda olayın gerçekleşmesinin beklenmesidir. Örnek büyüklüğünü belirlemek için yöntemlerden bahsederken, aynı zamanda her bir bağımsız değişkendeki olayların yeterli sayıda olmasının öneminin de farkında olmak oldukça önemlidir.

Örnek büyüklüğü belirlenirken cevaplanması gereken “Herhangi bir parametre katsayısının sıfıra eşit olduğunu ileri süren sıfır hipotezine karşı bu katsayının belli

²¹⁴ J. Scott Long, Simon Cheng, **a.g.e.**, p. 275.

²¹⁵ **A.e.**, p. 275.

bir değere eşit olduğunu ifade eden karşı ön savı test etmek için ne kadar örnek büyüklüğüne gereksinim olduğu?" sorusudur. İki kategorili tek bağımsız değişkenli Lojistik Regresyon Modellerinde örnek büyüklüğünün hesaplanması için 1998'de Hsieh ve arkadaşları tarafından önerilen formülasyon şu şekildedir,²¹⁶

$$n = \frac{(Z_{1-\alpha/2} + Z_{1-\theta})^2}{P_0(1 - P_0)\beta^{*2}}$$

Burada n , gereken örnek büyüklüğü; $\beta^* = \Delta/\sigma$ ifadesindeki Δ , hesaplanan iki grup ortalamasının farkıdır ve P_0 , standartlaştırılmış X 'in veya $X = 0$ iken olasılık değeri ve $(1 - \theta)$, testin gücü iken; $Z_{1-\theta}$ ve $Z_{1-\alpha/2}$ sırası ile θ ve $\alpha/2$ 'in Standart Normal dağılım tablosundaki kuyruk olasılık değerleridir.

Bunun dışında, bağımlı değişkenin iki kategorisine ait oran eşitliğinin test edilmesine dayanan geleneksel örnek büyüklüğü yöntemi kullanılabileceği gibi alternatif bir yol da 1981 yılında Whitemore tarafından önerilmiştir.²¹⁷

İki kategorili tek bağımsız değişkenli Lojistik Regresyon Modeli için oran eşitliğinin test edilmesine dayanan örnek büyüklüğü, sıfır hipotezi $H_0: P_0 = P_1$ 'in α anlamlılık düzeyindeki alternatif hipotez $H_a: P_0 \neq P_1$ olacak şekilde, testin gücü $1 - \theta$ için aşağıdaki eşitlik tarafından hesaplanabilmektedir,²¹⁸

$$n = \frac{\left(z_{1-\alpha} \sqrt{2\bar{P}(1 - \bar{P})} + z_{1-\theta} \sqrt{P_0(1 - P_0) + P_1(1 - P_1)} \right)^2}{(P_1 - P_0)^2}$$

Burada $\bar{P} = (P_0 + P_1)/2$ ve $z_{1-\alpha}$ ve $z_{1-\theta}$ sırası ile Standart Normal dağılımda α ve θ yüzde değerinin üst kısmını göstermektedir yani Standart Normal dağılım tablosundaki kuyruk olasılık değerleridir. Yukarıdaki örnek büyüklüğü bağımlı değişkenin her bir grubu için olması gereken örnek büyüklüğü değeridir.

²¹⁶ F. Y. Hsieh, Daniel A. Bloch, Michael D. Larsen, "A Simple Method of Sample Size Calculation for Linear and Logistic Regression", **Statistics in Medicine**, Vol:17, 1998, pp. 1623-1634.

²¹⁷ Alice S. Whitemore, "Sample Size for Logisitic Regression with Small Response Probability", **Journal of the American Statistical Association**, Vol: 76, 1981, pp. 27-32.

²¹⁸ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, p. 339.

Whitemore çalışmasında örnek büyüklüğünü bulurken, farklı anlamlılık düzeyleri ve testin güç düzeyleri için bağımsız değişkenin Normal, Üstel, Poisson ve Bernoulli dağılıma uyduğu durumlar için ayrı ayrı örnek büyüklüğü değerleri hesaplamıştır²¹⁹. İki kategorili tek bağımsız değişken içeren Lojistik Regresyon Modelinde bağımsız değişkenin Bernoulli dağılımına uyduğu durumlar için (x , “0” ve “1” olarak kodlanmaktadır) hipotez testleri $H_0: \beta_1 = 0$ 'a karşı $H_a: \beta_1 = \beta_1^*$ için gereken toplam örnek büyüklüğü şu şekilde hesaplanmaktadır,²²⁰

$$n = (1 + 2P_0) * \frac{\left(z_{1-\alpha} \sqrt{\frac{1}{1-\pi} + \frac{1}{\pi}} + z_{1-\theta} \sqrt{\frac{1}{1-\pi} + \frac{1}{\pi e^{\beta_1^*}}} \right)^2}{P_0 \beta_1^{*2}}$$

$\pi = P(X = 0)$. Buradan hesaplanan örnek büyüklüğü bağımlı değişkenin her grubu için eşit olacak şekilde yuvarlanmaktadır.

Whitemore ayrıca, bağımsız değişkenin nicel olduğu durumlar için örnek büyüklüğünü hesaplarken, bağımsız değişkenin ortalaması “0” ve varyansı “1” olacak şekilde standartlaştırılmış değerlerin kullanılmasını önermiştir. α anlamlılık düzeyi ve $1 - \theta$ testin gücü ile tek taraflı test değerinin kullanılarak $H_0: \beta_0 = 0$ karşı $H_a: \beta_1 = \beta_1^*$ hipotezinin testini yapabilmek için örnek büyüklüğü şu şekilde hesaplanmaktadır,

$$n = (1 + 2P_0 \delta) * \frac{\left(z_{1-\alpha} + z_{1-\theta} e^{-0.25 \beta_1^{*2}} \right)^2}{P_0 \beta_1^{*2}}$$

Bu eşitlikte $\delta = \frac{1+(1+\beta_1^{*2})e^{1.25\beta_1^{*2}}}{1+e^{-0.25\beta_1^{*2}}}$ ve $P_0 = \frac{e^{\beta_0}}{1+e^{\beta_0}}$ yani P_0 standartlaştırılmış bağımsız değişkenin beklenen değerinin lojistik olasılığı olarak hesaplanan değeridir.

Hsieh ve arkadaşlarının 1998 yılında yaptıkları çalışmadaki Çok Değişkenli Lojistik Regresyon için önerdikleri örnek büyüklüğü yukarıdaki denkleme uyarlandığında aşağıdaki eşitlik elde edilmektedir,

²¹⁹ Alice S. Whitemore, **a.g.e.**, pp. 27–32.

²²⁰ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, p. 341.

$$n = \frac{(1 + 2P_0\delta)}{1 - \rho^2} * \frac{(z_{1-\alpha} + z_{1-\theta}e^{-0.25\beta_1^2})^2}{P_0\beta_1^2}$$

ρ^2 , ilgilenilen deęişkenin (örneğin, x_k), modelde yer alan dięer $p - 1$ deęişkene göre hesaplanan kısmi korelasyon katsayısının karesidir.²²¹

2010’da Alam v.d.²²² tarafından yapılan çalışmada, Whittemore (1981) ile Hsieh v.d. (1998) tarafından önerilen örnek büyüklüğü hesaplama yöntemlerinin birleşiminden elde edilen yeni bir yöntem önerilmiştir. Bu simülasyon çalışmasında, bağımsız deęişken Bernoulli dağılımına uyduęu durumlarda, Hsieh ve arkadaşlarının uyarladığı yöntemle yeni önerilen yöntemin eşit düzeyde iyi sonuç verdięi elde edilmekle birlikte güç bakımından yeni yöntemin daha iyi sonuçlar verdięi gözlemlenmiştir.

Lojistik Regresyon Modellerinde regresyon katsayılarının tahminlerinin güvenilirliğini sağlamak için her bir bağımsız deęişken için kaç tane birime gerek olduęu konusu 1996’da Peduzzi v.d. tarafından Monte Carlo çalışması olarak incelenmiştir²²³. Söz konusu çalışmada, her bir parametre için birim sayısı 2, 5, 10, 15, 20 ve 25 olacak şekilde simülasyon çalışması yapılmıştır. Yapılan çalışma sonucunda, her bir parametre için birim sayısının en az 10 olmasının, varyansın yüksek ya da düşük tahmin edilmesinden dolayı, Wald tarafından önerilen katsayıların testlerinde meydana gelebilecek problemleri ortadan kaldırmak için yeterli olduęu gösterilmiştir. Veri setinin yeterli olması için modeldeki parametre sayısı $p + 1 \leq \min(n_1, n_0) / 10$ ’dan daha çok olmamalıdır. Bununla beraber, model temelli tahminlerin performansı genellikle deęişkenlerin birim sayısı yerine toplam örnek büyüklüğü tarafından belirlenmektedir.²²⁴

²²¹ F. Y. Hsieh, Daniel A. Bloch, Michael D.Larsen, **a.g.e.**, pp. 1623–1634.

²²² M. Khorshed Alam, M. Bhaskara Rao, Fu-Chih Cheng, “Sample Size Determination in Logistic Regression”, **The Indian Journal of Statistics**, Vol: 72-B, 2010, pp. 58-75.

²²³ P. N. Peduzzi, J. Concato, E. Kemper, T. R. Holford, A. Feinstein, “A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis”, **Journal of Clinical Epidemiology**, Vol: 49, 1996, pp. 1373–1379.

²²⁴ David W. Hosmer, Stanley Lemeshow, **a.g.e.**, p. 343.

İKİNCİ BÖLÜM: DİSKRİMİNANT ANALİZİ

Bağımlı değişkenin nitel, bağımsız değişkenlerin çoğunun nicel olduğu durumlar için kullanılması uygun olan çok değişkenli istatistik yöntemlerden biri de Diskriminant Analizi'dir.

İlk olarak 1936'da Fisher²²⁵ tarafından iki grubu sınıflamak için geliştirilmiş olan Diskriminant Analizi, 1948 yılında Rao²²⁶ tarafından ikiden fazla grup için geliştirilmiştir. 1949'da Hoel ve Peterson²²⁷ Diskriminant Analizi'nin, grupların farklarının kaynağını saptamak ve birimlerin hangi gruba ait olduğuna karar vermek gibi iki temel amaç için kullanıldığına dikkat çekmişlerdir. Sonrasında yapılan bazı çalışmalarda Diskriminant Analizi, uygulama amacına göre farklı isim almaktadır. Eğer çalışmada birbirinden farklı iki veya daha fazla grup üyesini en iyi ayırt etme yolu ile değişkenlerin tanımlanması ve fonksiyonun elde edilmesi amaçlanıyorsa *Tanımlayıcı Diskriminant Analizi* (Descriptive Discriminant Analysis), amaç hangi gruptan geldiği bilinmeyen birimlerin ait olduğu grubun saptanması ise *Tahmin Edici Diskriminant Analizi* (Predictive Discriminant Analysis) adı verilmektedir.²²⁸ Bununla birlikte birçok çalışmada bu iki amaç iç içe girmiş durumda olup birlikte ele alınmaktadır. Bu çalışmada Diskriminant Analizi denildiğinde bu iki durumun da birlikte ele alındığı analizden bahsedilmektedir.

Kısaca Diskriminant Analizi, bağımsız değişkenler açısından birbirleri içine girmiş ortak özelliklere sahip grupları birbirinden ayırmak için grup ortalama vektörlerini

²²⁵ R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems", **Annals of Eugenics**, 7(2), 1936, pp. 466–475.

²²⁶ C. R. Rao, "The Utilization of Multiple Measurements in Problem of Biological Classification", **Journal of the Royal Statistical Society**, Series B (Methodological), 1948, Vol. 10, No. 2, pp. 159–203.

²²⁷ P. G. Hoel, R. P. Peterson, "A Solution to the Problem of Optimum Classification", **The Annals of Mathematical Statistics**, 1949, Vol. 20, No. 3 (Sep.), p. 433.

²²⁸ A. Pedro Durate Silva, Antonie Stam, **Reading and Understanding Multivariate Statistics**, Edited by Laurence G. Grimm, Paul R. Yarnold, American Psychological Association Washington, DC., 1994, p. 277.

kullanarak fonksiyonlar geliřtiren ve geliřtirdiđi fonksiyonların sınıflandırmadaki gücünü test eden bir yöntemdir²²⁹.

2.1 İki Kategorili Diskriminant Analizi

İki Kategorili Diskriminant Analizi'nde, iki grup arasındaki ayrımı en iyi yapacak faktörlerin doğrusal bileřimini belirlemek amacıyla diskriminant fonksiyonunun yazılması amaçlanmaktadır. Diskriminant fonksiyonunu yazarken, gruplararası deđişimin gruplarıiçi deđişime oranından elde edilen özdeđer veya karakteristik kök (λ) deđerini maksimum yapacak fonksiyonun yazılmasına çalıřılmaktadır.²³⁰

Tüm Diskriminant Analizi problemleri ařađıdaki dört ortak belirleyici özellik ile tanımlanmaktadır:²³¹

- Diskriminant Analizi'nde incelenen birimler iki gruptan sadece birine ait olması ve grupların da bađdařmaz ve bađımsız olması gerekmektedir.
- Her bir Diskriminant Analiz grubu iyi tanımlı olmalı ve gruplar birimler arasındaki gerçek farkları yansıtmalıdır. Ayrıca gruplar ikiden fazla ve sıralı ise Diskriminant Analizi yerine bu durum için geliřtirilen Sıralı Lojit Analiz tekniđinin kullanılması tercih edilmelidir.
- Diskriminant Analizi'nde, gruplar veri seti toplanmadan önce tanımlanmış olmalıdır. Bunun zıttı durumlar için, veri setinin belirleyici özelliklerine dayanarak grupları belirlemede ise Kümeleme Analizi kullanılmaktadır.
- Son olarak da, bađımsız deđiřkenler seti mümkün olduđunca tam ve birimleri tanımlamada dođru ayırım yapacak şekilde oluřturulmalıdır. Ancak bazı belirleyici özellikler bađımlı deđiřkeni açıklamada anlamlı iken, grupları ayırma veya sınıflandırmada etkili olmayabilir. Model yorumlanırken bu durum göz önünde bulundurulmalıdır.

²²⁹ Kazım Özdamar, **Paket Programlar ile İstatistiksel Veri Analizi**, Cilt 2, Yenilenmiş 5. Baskı, 2004, s. 355.

²³⁰ Subhash Sharma, **Applied Multivariate Techniques**, John Wiley & Sons, Inc., 1996, p. 241.

²³¹ A. Pedro Durate Silva, Antonie Stam, Edited by Laurence G. Grimm, Paul R. Yarnold, **a.g.e.**, p. 277.

Uygulamada sıklıkla yapılan, bağımlı değişken gruplarının ayırımını sağlamada önemli olabileceği düşünülen tüm değişkenlerden oluşan büyük bir set oluşturmak ve analiz sırasında bu set içinde bağımlı değişkeni açıklamada anlamlı olanları saptamaktır. Diskriminant Analizi'nde önemli ve zor olan bir süreçtir. Bunun için, araştırmacının kullanacağı istatistik teknikle birlikte konu hakkındaki bilginin birleştirmesinin en etkili yöntem olacağı düşünülmektedir.

Diskriminant Analizi'ne ilk olarak modele girecek ayırıcı değişkenlere karar verilerek başlanmalıdır. Daha sonra varsayımların kontrolü ile devam edilmektedir. Ayırıcı değişkenlerin seçiminden sonra diskriminant fonksiyonunun elde edilerek testleri yapılır ve yorumlanır. Son olarak da Diskriminant Analizinin başarısı değerlendirilir.

2.1.1 Ayırıcı Değişkenlerin Seçimi

Diskriminant fonksiyonunun oluşturulmasına geçmeden önce modele girecek uygun bağımsız değişkenlerin seçilmesi gerekmektedir. Bu amaçla ilk önce, birimlerin gruplandırılmasında anlamlı olabileceği düşünülen bağımsız değişkenlerin grup farklarının anlamlı olup olmadığının saptanmasına olanak veren “Grup Ortalamalarının Eşitlik Testi” yapılarak grupları ayırmada etkin olan değişkenlerin modele bağımsız değişken olarak alınmalarına karar verilebilir.²³²

Genellikle kullanılan bir başka yol da, değişkenlerin grupları ayırmadaki gücüne bakmaksızın araştırma yapılan konu hakkında literatürde önemli olabileceği düşünülen tüm bağımsız değişkenler modele alınarak Adım Adım Diskriminant Analizi (Stepwise Discriminant Analysis) ile modele girecek bağımsız değişkenlere karar verilmesidir²³³. Bu yöntemle bağımsız değişkenler arasındaki çoklu doğrusal bağlantı problemi de çözülmüş olmaktadır. Yaygın kullanım alanı bulmasına rağmen, değişkenlerin seçim ve sıralanmasında kullanılan Adım Adım Diskriminant Analizi'nin bazı teorik sakıncalarının da farkında olmak gerekmektedir. Bu yöntemdeki temel problem, seçilen alt kümenin içindeki değişkenlerin her zaman

²³² Jacques Tacq, **Multivariate Analysis Techniques in Social Science Research**, Sage Publication, 1997, p. 248.

²³³ Joseph F. Hair, v.d., **a.g.e.**, p.293.

bağımlı değişkeni açıklayan en iyi değişkenler olmamasıdır. Ayrıca bağımsız değişkenler arasındaki korelasyonun yüksek olduğu durumlarda boyut indirgemek amacıyla Adım Adım yöntemi yerine Faktör Analizi'nin kullanılması daha uygundur.²³⁴

Diskriminant Analizi sonucunda elde edilen bağımsız değişkenlerin grupları ayırmadaki gücünün etkisine genellikle, 1932'de Wilks²³⁵ tarafından önerilen, modelin Wilks' Lamda istatistiği incelenerek karar verilmektedir. McKay ve Campbell²³⁶ da 1982'de yaptıkları çalışma ile Wilks' Lamda değerinin toplam grupları ayırmada endeks olarak kullanılmasının kabul edilebilir olduğunu ifade etmişlerdir. Bununla birlikte Wilks' Lamda kadar önemli olan, doğrusal bileşenlerin (kanonik değişkenlerin) tümü ile ilgili özdeğerlerin de dikkate alınması gerekmektedir. Analiz sonucunda elde edilen değişkenler kümesinde yer alan bazı önemli doğrusal bileşenlerin çok küçük özdeğerini dikkate alarak hata yapmaktan kaçınmak gerekmektedir. Grupları ayıran değişkenlerden oluşan doğrusal bileşenlerin önemi kendilerine karşılık gelen özdeğerleri tarafından yansıtılmaktadır. Özdeğeri maksimum yapacak şekilde elde edilen diskriminant fonksiyonunun, grup ayırımında önemli olan doğrusal bileşenleri içinde bulundurduğu anlamına gelmektedir.²³⁷

Modelde yer alacak değişkenlerin belirlenmesinde kullanılan Adım Adım Diskriminant Analizi'nde üç farklı yaklaşım vardır. Bunlar; İleri Doğru Seçim (Forward Selection), Geriye Doğru Seçim (Backward Selection) ile ileri ve geriye doğru eleme yöntemlerinin birleşiminden oluşan Adım Adım Seçim (Stepwise Selection) yaklaşımıdır. Bu yaklaşımlarda modele girecek veya çıkacak değişkenlere

²³⁴ A. Pedro Durate Silva, Antonie Stam, Edited by Laurence G. Grimm, Paul R. Yarnold, **a.g.e.**, p. 289.

²³⁵ S. S. Wilks, **Certain Generalizations in the Analysis of Variance**, Biometrika, Vol. 24, No. 3/4, 1932, pp. 471–494.

²³⁶ R. J. McKay, N. A. Campbell, **Variable Selection Techniques in Discriminant Analysis I. Description**, British Journal of Mathematical Statistical Psychology, 35, 1982a, pp. 1–29.

²³⁷ A. Pedro Durate Silva, Antonie Stam, Edited by Laurence G. Grimm, Paul R. Yarnold, **a.g.e.**, p. 289.

karar verirken genellikle kullanılan seçim kriterleri; Wilks' Λ , Rao's V , Mahalanobis uzaklığı ve gruplararası F oranıdır.²³⁸

Değişkenlerin diskriminant fonksiyonunda yer almaması önemsiz oldukları anlamına gelmemektedir. Değişkenler tek tek analiz edilip Wilks' Λ değeri bakıldığında en düşük değerle modelde olması beklenen değişkenin bazı durumlarda Adım Adım Diskriminant Analizi yöntemi ile modele girmediği görülmektedir. Bunun nedeni değişkenlerin kendi içlerindeki ilişkileridir. Bunu saptamak için korelasyon matrisi incelenebilir. Bu matrisdeki bazı bağımsız değişkenler arasındaki korelasyon değerlerinin büyük olması bu değerlere ait değişkenlerden ikisinin birden modelde kullanılmaması gerektiğini ifade etmekle birlikte, bazı durumlarda (bağımsız değişkenlerin bağımlı değişkenle olan korelasyon değerlerinin birbirine yakın olması gibi) hangisinin modelde kullanılmasının daha önemli olduğunu söylememektedir.²³⁹

Veri setindeki çoklu doğrusal bağlantı için Adım Adım Diskriminant Analizi'nin kullanılıp kullanılmamasına çoklu doğrusal bağlantının kaynağına bağlı olarak karar verilmelidir. Anakütleye dayalı çoklu doğrusal bağlantı, bağımsız değişkenler arasındaki ilişkinin düzenini oluşturulabilecek tüm mümkün örnek seti içinde barındırır. Bu durum için Adım Adım Diskriminant Analizi'nin kullanılması örnekten örneğe değişmeyeceği için uygundur. Diğer taraftan bağımsız değişkenler arasındaki ilişki örnekten örneğe farklılık gösterdiğinde bu yöntemin kullanılması uygun olmayacaktır.²⁴⁰

2.1.2 Diskriminant Analizinin Varsayımları

Diskriminant fonksiyonunun oluşturulmasındaki en önemli varsayımlar çoklu normallik varsayımı ve grupların eşit varyansa sahip olması varsayımlarıdır.

Veri setinin çoklu normallik varsayımını sağlamaması diskriminant fonksiyonunun tahmin edilmesinde soruna neden olabilir. Bu durumda alternatif bir yöntem olan

²³⁸ Subhash Sharma, **a.g.e.**, pp. 265–266.

²³⁹ **A.e.**, p. 272.

²⁴⁰ **A.e.**, pp. 272–273.

Lojistik Regresyon Analizi'nin kullanılması önerilmektedir. Eşit varyanslılık varsayımının sağlanamaması, başka bir ifade ile eşit olmayan varyans kovaryans matrisi ise sınıflandırma aşamasını olumsuz etkilemektedir. Eğer örnek büyüklüğü küçük ve varyans kovaryans matrisi eşit değilse tahminlerin istatistik anlamlılığını da olumsuz şekilde etkilemektedir. Örnek büyüklüğü yeterli olan gruplar arasında eşit varyanslı olma varsayımının sağlanamaması ise, gözlemlerin varyans kovaryans matris değeri daha büyük olan grupta, olması gerekenden daha fazla yer almasına neden olmaktadır ve bu sıklıkla karşılaşılan bir durumdur.

Bu iki temel varsayımın dışında veri setinin, sonuçları etkileyen bir diğer belirleyici özelliği de bağımsız değişkenler arasında çoklu doğrusal bağlantı problemi olmasıdır.²⁴¹ Bu problemi ortadan kaldırmak için genellikle modele alınacak değişkenlerin seçiminde “Adım Adım (Stepwise)” veya “En İyi Alt Model Seçimi (Best Subset Selection)” yöntemleri kullanılmaktadır.

Diskriminant Analizi'nin bir diğer varsayımı da, bağımlı ve bağımsız değişkenler arasındaki ilişkinin doğrusal olması gerekliliğini vurgulayan, doğrusallık varsayımıdır.

2.1.2.1 Çoklu Normallik Varsayımı

Bağımsız değişkenler vektörünün çoklu normalliği sağlıyor olması aşağıdaki özelliklerin de sağlandığını göstermektedir:²⁴²

- a- Bağımsız değişkenlerden oluşan doğrusal kombinasyonlar da normal dağılıma uymaktadır.
- b- Bağımsız değişkenlerden oluşan tüm mümkün alt kümeler de çoklu normalliği sağlamaktadır.
- c- Kovaryansın sıfır olması, ele alınan değişkenlerin birbirini etkilemeden (bağımsız olarak) dağıldığını göstermektedir.
- d- Bağımsız değişkenlerin koşullu dağılımı da (çoklu) normaldir.

²⁴¹ Joseph F. Hair, v.d., **a.g.e.**, pp. 290–291.

²⁴² Richard A. Johnson, Dean W. Wichern, **Applied Multivariate Statistical Analysis**, Third Edition, Prentice Hall, New Jersey, 1992, p. 133.

Çoklu normallik varsayımının test edilmesinin zorlukları nedeniyle, varsayımın testinde birkaç pratik seçenek geliştirilmiştir. Bunlardan biri, her bir değişkenin tek tek Q-Q grafiği kullanılarak normallik testinin yapılmasıdır. Eğer bu çizilen grafiklerden herhangi biri normalliği sağlamıyorsa çoklu normallikle ilgili şüpheler doğmaktadır. Öte yandan tüm değişkenler tek tek normallik varsayımını sağlasa bile bu, çoklu normalliğin göstergesi değildir. Bu nedenle, çoklu normalliğin test edilmesi gerekmektedir.²⁴³

Çeşitli çoklu normallik testlerinde ve grafik yöntemde Mahalanobis uzaklığının karesi kullanılmaktadır²⁴⁴. Mahalanobis uzaklığı, iki nokta arasındaki İstatistik uzaklığa (Standart uzaklık $\left(\frac{x_i - x_j}{s}\right)^2$) değişkenler arasındaki kovaryans ve korelasyon değerlerindeki kullanılmasıyla elde edilen uzaklıktır. İki gözlem arasındaki Mahalanobis uzaklığının karesi şu şekilde hesaplanmaktadır;²⁴⁵

$$MD_{ik}^2 = \frac{1}{1 - r^2} \left[\frac{(x_{i1} - x_{k1})^2}{s_1^2} + \frac{(x_{i2} - x_{k2})^2}{s_2^2} + \frac{2r(x_{i1} - x_{k1})(x_{i2} - x_{k2})}{s_1 s_2} \right]$$

Mahalanobis uzaklığının en önemli özelliği, değişkenler ilişkili değilse Mahalanobis uzaklığının İstatistik uzaklığa dönüşmesidir. Ayrıca değişkenlerin varyansları “1” ise Mahalanobis uzaklığı Öklid uzaklığına indirgenir. Kısaca, Öklid uzaklığı ve İstatistik uzaklık Mahalanobis uzaklığının özel durumudur.²⁴⁶

p değişkenli durum için, iki gözlem arasındaki Mahalanobis uzaklığı en basit olarak aşağıdaki gibidir;

$$MD_{ik} = (x_i - x_k)' S^{-1} (x_i - x_k)$$

Burada x , $p \times 1$ koordinat vektörü ve S , $p \times p$ kovaryans matrisidir.

Johnson ve Wichern tarafından çoklu normalliğin kontrol edilmesinde, tek değişkenli Q-Q grafiğinin benzeri olan Mahalanobis uzaklığının kullanıldığı χ^2 grafiğinin

²⁴³ A. B. Ellem, **Multivariate Analysis** (Stat501), January 10, 2008 (Çevirimiçi) <http://mcs.une.edu.au/~stat501/UnitNotes/unitnotes08.pdf>, 25 Ağustos 2011, p. 28.

²⁴⁴ M. Bilodeau, D. Brenner, **Theory of Multivariate Statistics**, Springer, 1999, p. 170.

²⁴⁵ Subhash Sharma, **a.g.e.**, p. 44.

²⁴⁶ Carl J. Huberty, **Applied Discriminant Analysis**, John Wiley & Sons INC., 1994, p. 43.

kullanılması uygundur. Bu durumda uzaklıkların karesi küçükten büyüğe sıralayarak, bu değerlere karşılık gelen beklenen değerleri χ^2 dağılımı kullanarak hesaplanır ve sıralanmış Mahalanobis uzaklık değerleri ile χ^2 dağılımı kullanılarak elde edilen beklenen değerlerinin grafiği çizilir. Bu grafik sonucunda doğrusal çizgiden sapma çoklu normallik varsayımının sağlanmadığı anlamına gelmektedir. Veri setinde n ve $n - p$ değeri 25 veya 30'dan büyükse Mahalanobis uzaklıklarının karesi χ^2 dağılıma benzer özellikler göstermektedir.²⁴⁷

Bununla birlikte Bilodeau ve Brenner²⁴⁸, tek değişkenli durumlar için örnek büyüklüğünün 25'den fazla olması χ^2 yaklaşımının kullanılması için yeterli olsa da, değişken sayısı arttığında (örneğin 4 olduğunda) örnek büyüklüğünün 100 olmasının χ^2 yaklaşımının kullanılması için yeterli olmayacağını ifade etmişlerdir. Bunun yerine Beta dağılımının kullanılması önerilmektedir. Bunun dışında, Johnson ve Wichern²⁴⁹ da küçük örnekler için χ^2 dağılımının daha genel hali olan Gamma dağılımının kullanılmasını önermekle birlikte, uygulamada χ^2 dağılımının kullanılması daha yaygındır.

Lachenbruch, Sneeringer ve Revo 1973'de yaptıkları çalışma ile çoklu normallik varsayımının sağlanmaması ile sınıflandırma oranlarının nasıl etkilendiğini incelemişlerdir. Bu çalışma ile çoklu normallik varsayımının sağlanmaması sonucunda anlamlılık testleri ve doğru sınıflandırma oranlarının güvenilirliğini kaybedeceği saptanmıştır.²⁵⁰ Çoklu normallik varsayımının sağlanmadığı durumda öncelikle uygun dönüşümler yapılarak veri setinin tekrar oluşturulmasına çalışmak gerekmektedir. Dönüşümler denendikten sonra da sorun devam ediyorsa açıklayıcı değişkenler için herhangi bir dağılım varsayımına dayanmayan Lojistik Regresyon Analizi'nin kullanılması daha uygun olacaktır.²⁵¹

²⁴⁷ Richard A. Johnson, Dean W. Wichern, **a.g.e.**, pp. 158–160.

²⁴⁸ M. Bilodeau, D. Brenner, **a.g.e.**, p. 186.

²⁴⁹ Richard A. Johnson, Dean W. Wichern, **a.g.e.**, p. 160.

²⁵⁰ Peter A. Lachenbruch, Cheryl Sneeringer, Lawrence T. Revo, "Robustness of the Linear and Quadratic Discriminant Function to Certain Types of Non-Normality", **Communications in Statistics**, 1973, Vol. 1, pp. 39–56.

²⁵¹ Subhash Sharma, **a.g.e.**, p. 8.

2.1.2.2 Eşit Varyanslılık Varsayımı

Çoklu normallik varsayımı dışında Diskriminant Analizi sonuçlarının güvenilirliği açısından eşit varyanslılık varsayımının da sağlanması gerekmektedir. Tüm grupların varyanslarının eşit olduğunun ileri sürüldüğü sıfır hipotezinin testi için, F dağılımına uyan Box's M değerinden hesaplanan F test istatistiği kullanılmaktadır.²⁵²

$$M = (n - g) \log|S| - \sum_{i=1}^g (n_i - 1) \log|S_i|$$
$$F = \begin{cases} M/b & \text{eğer } \tau > (1 - \rho)^2 \\ \frac{f_2 M}{f_1(b - M)} & \text{eğer } \tau \leq (1 - \rho)^2 \end{cases}$$

Burada g ; grup sayısı, n_i ; i . gruptaki birim sayısı, S ; ortak kareler toplam matrisi ve S_i ; i . grubun kareler toplam matrisidir. F dağılımına uyan Box's M test istatistiği için f_1 ve f_2 serbestlik dereceleri, modeldeki değişken sayısı p ile gösterilmek üzere sırası ile şu şekilde hesaplanır;²⁵³

$$f_1 = (g - 1)p(p + 1)/2$$

$$f_2 = \frac{f_1 + 2}{|\tau - (1 - \rho)^2|}$$

2.1

İkinci serbestlik derecesindeki τ ve ρ katsayıları da şu şekilde elde edilmektedir;

$$\tau = \frac{(p - 1)(p + 2)}{6(g - 1)} \left(\sum_{i=1}^g \frac{1}{(n_i - 1)^2} - \frac{1}{(n - g)^2} \right)$$

$$\rho = 1 - \frac{2p^2 + 3p - 1}{6(p + 1)(g - 1)} \left(\sum_{i=1}^g \frac{1}{(n_i - 1)} - \frac{1}{(n - g)} \right)$$

²⁵² Kazım Özdamar, Cilt 2, **a.g.e.**, s. 407.

²⁵³ SPSS, (Çevirimiçi) http://support.spss.com/ProductsExt/SPSS/Documentation/Statistics/algorithms/14.0/app14_boxs_m.pdf, 10 Eylül 2011.

F test istatistiğinin formülünde yer alan b değeri de şu şekilde hesaplanmaktadır;²⁵⁴

$$b = \begin{cases} \frac{f_1}{\rho - (f_1/f_2)} & \text{eğer } \tau > (1 - \rho)^2 \\ \frac{f_2}{\rho - (2/f_2)} & \text{eğer } \tau \leq (1 - \rho)^2 \end{cases}$$

Eğer $(1 - \rho)^2 - \tau = 0$ ise Box's M değerinden elde edilen F testi yerine *Bartlett* χ^2 test istatistiğinin kullanılması uygundur. χ^2 test istatistiği şu şekilde hesaplanabilir;²⁵⁵

$$\chi^2 = M\rho$$

χ^2 istatistiği, f_1 (Eşitlik 2.1'de yer alan) serbestlik derecesi ile kabul edilen anlamlılık düzeyindeki χ^2 tablo değeri ile karşılaştırılarak yorumlanır.

Eşit varyanslılık varsayımının sağlanmadığı durumda, doğrusal diskriminant fonksiyonunun yerine karesel diskriminant fonksiyonunun oluşturulması daha uygundur²⁵⁶.

Karesel diskriminant fonksiyonun nasıl oluşturulacağına "Diskriminant Fonksiyonunun Elde Edilmesi" başlığında (Başlık 2.1.3) değinilmiştir.

2.1.2.3 Çoklu Doğrusal Bağlantı Problemi Olmaması Varsayımı

Birçok istatistik yöntemin uygulanabilir olması için sağlanması gereken varsayımlardan biri de bağımsız değişkenler arasında ilişki olmamasıdır. Bağımsız değişkenler arasındaki ilişki en basit şekli ile iki değişken arasındaki korelasyon katsayısının gücü ile belirlenebilmektedir. Çoklu doğrusal bağlantının olması, bir bağımsız değişken diğer bir bağımsız değişken veya değişkenler tarafından tam olarak tahmin edilebiliyorsa bu değişkenlerden biri veya bazılarının gereksiz olacağı anlamına gelmektedir. Bu durumda söz konusu değişken veya değişkenlerin modelde yer almaması daha uygundur.

²⁵⁴ Kazım Özdamar, Cilt 2, **a.g.e.**, s. 407.

²⁵⁵ **A.e.**, s. 408.

²⁵⁶ Subhash Sharma, **a.g.e.**, p. 264.

Çoklu doğrusal bağlantı problemini saptamak için çeşitli yöntemler kullanılmaktadır.²⁵⁷

- Bu yöntemlerden biri, daha öncede değinildiği gibi, basit korelasyon matrisinin incelenmesidir. İki bağımsız değişken arasındaki basit korelasyon katsayısı anlamlı olması çoklu doğrusal bağlantı probleminin ortaya çıkabileceğini göstermektedir.
- Çoklu doğrusal bağlantının saptanmasında kullanılan bir diğer yöntem de modele bağımsız değişkenler ilave edildikçe belirlilik katsayısındaki değişimlerin incelenmesidir.
- Bir başka yöntem de kısmi korelasyon katsayılarının incelenmesidir. İki değişken arasındaki basit korelasyon katsayıları anlamlı olduğu halde kısmi korelasyon katsayılarının anlamsız çıkması çoklu doğrusal bağlantı problemi için bir gösterge olabilmektedir.²⁵⁸
- Çoklu doğrusal bağlantı problemi Varyans Artış Faktörleri (Variance Inflation Factors-VIF) incelenerek de saptanabilmektedir.²⁵⁹

$$VIF = \frac{1}{(1 - R^2)}$$

- Tolerans değeri $T = 1 - R_i^2$ şeklinde hesaplanmaktadır, bağımsız değişkenler arasındaki çoklu bağımlılığın miktarının ölçüsüdür.²⁶⁰ Eşitlikten de anlaşılacağı gibi küçük Tolerans değeri daha büyük VIF değeri demektir.
- Çoklu doğrusal bağlantı probleminin saptanmasında kullanılan, bir diğer yöntem de Koşul Endeks (Condition Index-CI) değerlerinin hesaplanmasıdır. CI değerleri aşağıdaki gibi hesaplanmaktadır:²⁶¹

$$CI = \sqrt{\frac{V_{max}}{V_{x_i}}}$$

Burada maksimum açıklanan varyans V_{max} , ile i . değişken tarafından açıklanan toplam varyans da V_{x_i} ile gösterilmektedir. CI değerinin 10 ile 30 arasında değer

²⁵⁷ Damodar N. Gujarati, **a.g.e.**, ss. 335–339.

²⁵⁸ Neyran Orhunbilge, **Uygulamalı Regresyon ve Korelasyon Analizi**, İstanbul, 2002, s. 241.

²⁵⁹ Damodar N. Gujarati, **a.g.e.**, s. 339.

²⁶⁰ Joseph F. Hair, v.d., **a.g.e.**, p.273.

²⁶¹ Damodar N. Gujarati, **a.g.e.**, s. 338.

alması çoklu doğrusal bağlantı probleminin orta düzeyde, 30'dan büyük olması da çok güçlü olduğu anlamına gelmektedir.

Çoklu doğrusal bağlantı probleminin saptanmasından sonra, soruna neden olan değişkenlerden hangisinin modelde kalıp hangisinin dışarıda bırakılacağına karar vermek kolay değildir. Bu aşamada çoklu doğrusal bağlantı problemini çözmek* için Adım Adım Regresyon Analizi'ni kullanmak uygun olacaktır. Bunun dışında bazı durumlarda çoklu doğrusal bağlantı problemine neden olan değişkenlere dönüşüm uygulanarak modele alınması daha uygun olabileceği gibi her zaman mümkün olmamakla birlikte örnek büyütülerek çoklu doğrusal bağlantı problemi çözülmeye çalışılabilir. Bunların dışında çoklu doğrusal bağlantı problemini çözmek için veri setine çok değişkenli istatistik yöntemlerden olan Faktör Analizi uygulanabilir.²⁶²

2.1.2.4 Doğrusallık Varsayımı

Diskriminant Analizi'nde de korelasyon katsayılarına dayanan çok değişkenli tekniklerdeki gizli varsayım olan doğrusallık varsayımının sağlanması önemlidir. Bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin doğrusal olmaması, doğrusal korelasyon katsayılarının gerçek ilişkiyi daha düşük göstermesine neden olmaktadır.²⁶³

2.1.3 Diskriminant Fonksiyonunun Elde Edilmesi

İki Kategorili Diskriminant Analizi'ndeki amaç, grupların ayırımını en iyi şekilde yapacağı düşünülen bağımsız değişkenler setinin iki grubu ayırmada yeterli olup olmadığının analizinin yapılmasıdır. Bu nedenle iki grubun en üst düzeyde ayırımını yapabilecek bağımsız değişkenlerin doğrusal kombinasyonları araştırılmaktadır. Bu doğrusal kombinasyona “diskriminant fonksiyonu” adı verilmekte ve

$$T - \bar{T} = k_1(X_1 - \bar{X}_1) + k_2(X_2 - \bar{X}_2) + \dots + k_p(X_p - \bar{X}_p)$$

* Çoklu doğrusal bağlantı problemi ortaya çıktığında çözüm yöntemleri ile ilgili ayrıntılı bilgi için Neyran Orhunbilge'nin “Uygulamalı Regresyon ve Korelasyon Analizi” kitabı sayfa 242'ye bakınız.

²⁶² Damodar N. Gujarati, **a.g.e.**, ss. 340–344.

²⁶³ Ali Sait Albayrak, **Uygulamalı Çok Değişkenli İstatistik Teknikler**, Asil Yayın Dağıtım Ltd. Şti., 2006, s. 65.

$$t = k_1x_1 + k_2x_2 + \dots + k_px_p \quad 2.2$$

şekilinde gösterilmektedir.²⁶⁴ Eşitlikte küçük harfle gösterilen t ve x_i , ortalamadan sapmaları ifade etmektedir. k_i katsayıları, diskriminant ağırlıkları (kanonik değerler) olarak adlandırılmaktadır. İki grup olduğunda sadece bir diskriminant fonksiyonu vardır.

Diskriminant fonksiyonundaki ağırlıklar, gruplararası varyansın gruplarıçi varyansa oranını veren aşağıdaki eşitlikle hesaplanan λ (eigenvalue-özdeğer) değeri²⁶⁵

$$\lambda = \frac{\text{Gruplararası Varyans}}{\text{Gruplarıçi Varyans}}$$

şeklinde yazılmaklabirlikte, yukarıdaki eşitlik Fisher tarafından önerilen ve k katsayılarının elde edilmesinde kullanılmak üzere şu şekilde ifade edilmektedir.²⁶⁶

$$\lambda = \frac{k'Bk}{k'Wk}$$

Buradan, B ; gruplararası varyans, W ; gruplarıçi varyans matrisi olmak üzere λ değerini maksimum yapacak şekilde k katsayıları elde edilmektedir.*

Bu şekilde elde edilen eşitliğe *Fisher'in Doğrusal Diskriminant Fonksiyonu* adı verilmektedir²⁶⁷.

Katsayıların hesaplanmasında kullanılan eşitlikler doğrusal olarak bağımsız olmadığında, elde edilen katsayıların oranları birbirine eşit olmak üzere sonsuz sayıda çözüm elde edilecektir. Çözüm kümesi sonsuz olduğunda, standartlaştırma

²⁶⁴ Jacques Tacq, **a.g.e.**, p. 233.

²⁶⁵ Subhash Sharma, **a.g.e.**, p. 241.

²⁶⁶ Hüseyin Tatlıdil, **a.g.e.**, s. 258.

* λ 'in k 'ya göre kısmi türevleri alınarak sıfıra eşitlendiğinde elde edilen $(W^{-1}B - \lambda I)k = 0$ denklem sisteminde (W gruplarıçi ve B gruplararası varyasyonu ifade etmektedir) önce λ değerleri (özdeğerler) hesaplanarak denklem sisteminde yerine konulup diskriminant ağırlıkları hesaplanmaktadır.

²⁶⁷ Subhash Sharma, **a.g.e.**, p. 245.

amacıyla diskriminant fonksiyonunun katsayıları normalleştirilerek kullanılmalıdır. Normalleştirme, katsayıların k uzunluğuna* bölünerek elde edilmesidir.²⁶⁸

$$k_i^* = \frac{k_i}{\sqrt{k'k}} \quad i = 1, 2, \dots, p$$

k ; $p \times 1$ boyutunda p değişkene ait diskriminant ağırlıklar vektörüdür.

Her grup için birer diskriminant fonksiyonu yazılabileceği gibi grupların normal dağılıma uyduğu varsayımına dayanan Kanonik Diskriminant Analizi'nde uygulandığı şekilde, diskriminant ağırlıkları, ortalama fark vektörleri kullanılarak sabit değer de yer aldığı ortak bir diskriminant fonksiyonu olarak da ifade edilebilmektedir. Matris gösterimi ile diskriminant ağırlıkları²⁶⁹

$$k = S^{-1}(\bar{X}_1 - \bar{X}_2) \quad 2.3$$

ile bulunurken, fonksiyondaki sabit değer,²⁷⁰

$$k_0 = \left(-\frac{1}{2}\right)\bar{X}_1'S^{-1}\bar{X}_1 + \left(-\frac{1}{2}\right)\bar{X}_2'S^{-1}\bar{X}_2 \quad 2.4$$

eşitliği ile hesaplanmaktadır.

Eşitlik 2.3 ve 2.4'deki, S ; $p \times p$ boyutlu ortak varyans kovaryans matrisi olup, iki grubun örnek büyüklüğü ve varyans kovaryans matrisleri kullanılarak şu şekilde elde edilmektedir;

$$S = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2} \quad 2.5$$

$(\bar{X}_1 - \bar{X}_2)$; $p \times p$ boyutunda p tane değişkenin grup ortalamalarının fark vektörüdür.

Dikkat edilmesi gereken önemli bir nokta, diskriminant ağırlıklarının işaretlerinin diskriminant ekseninin yönünü gösteriyor olduğudur. Bağımsız değişkenlerin

* k uzunluğu $\|k\| = \sqrt{k'k}$ şeklinde hesaplanmaktadır.

²⁶⁸ Jacques Tacq, **a.g.e.**, p. 245.

²⁶⁹ Richard A. Johnson, Dean W. Wichern, **a.g.e.**, p. 505.

²⁷⁰ Kazım Özdamar, Cilt 2, **a.g.e.**, s. 361.

birbirleriyle ilişkilerinin yönüne göre işaret değişmektedir. Örneğin değişkenler arasındaki ilişki pozitif ise diskriminant katsayıları diskriminant ekseninin yönünü beklenenin aksi şekilde etkilemektedir yani değişkenlerin işaretleri terstir.²⁷¹ Sabit terim bağımsız değişkenlerin ortalamadan farkları kullanılmadığında fonksiyonda yer alır. Diskriminant ağırlıklarının hesaplanmasında her bir değişkenin ortalamadan farkları kullanıldığında ise sabit terim yoktur. Ayrıca, değişkenin standartlaştırılmış değerleri (z_i) yer aldığında diskriminant fonksiyonu yine sabit terim içermez. Standartlaştırılmış değerlerin kullanılması grupların belirlenmesinde hangi bağımsız değişkenin daha güçlü olduğunun karşılaştırılmasına imkan verdiği için tercih edilmektedir.

Diskriminant fonksiyonu veya fonksiyonlarının ayırma güçlerinin istatistik anlamlılığı *Wilks'Λ*, *Hotelling's Trace* ve *Pillai's Kriteriya* değerleri ile değerlendirilebilmektedir. *Roy'un En Büyük Karakteristik Kökü*, sadece ilk diskriminant fonksiyonunun ayırma gücünü değerlendirmede kullanılmaktadır.²⁷² Öneren istatistikçilerin adı ile anılan yukarıda ki dört yöntem test istatistikleri değerlerini elde ederken λ değerlerini kullanmaktadırlar.*

Burada unutulmaması gereken nokta, diskriminant fonksiyonunun oluşturulmasında, Bölüm 2.1.2'de incelenen varsayımlardan olan çoklu normallik ve eşit varyanslılık varsayımlarının sağlanıyor olması koşuludur.

Verilerin normal dağılıma uyduğu ve grupların eşit varyanslılık varsayımının sağlandığı durumda diskriminant fonksiyonunu oluştururken, fonksiyonun katsayılarının (doğrusal bileşenlerinin) hesaplanmasında Eşitlik 2.5'de verilen ortak varyans kovaryans matrisi kullanılırken, verilerin normal dağılıma uyduğu ancak gruplara ait eşit varyanslılık varsayımının sağlanmadığı durumda ise, grupların

²⁷¹ Jacques Tacq, **a.g.e.**, p. 238.

²⁷² Joseph F. Hair, v.d., **a.g.e.**, p. 294.

* *Wilks'Λ*, *Hotelling's Trace*, *Pillai's Kriteriya* ve *Roy'un En Büyük Karakteristik Kökü* ile ilgili ayrıntılı bilgi için bakınız: Neyran Orhunbilge, **Çok Değişkenli İstatistik Yöntemler**, İstanbul, 2010, ss. 342–344.

varyans kovaryans matrislerinin farkları ($S_1 - S_2$) alınarak karesel diskriminant fonksiyonunun doğrusal bileşenlerinin elde edilmesi daha doğru olacaktır.²⁷³

Bununla birlikte Marks ve Dunn 1974'te sundukları çalışmada varyans kovaryans matrisinin eşit olmadığı durumda diskriminant fonksiyonunu farklı λ değerleri ve örnek büyüklükleri için karesel ve doğrusal fonksiyon denklemlerini test ederek hangisinin daha iyi sonuç verdiğini saptamışlardır. Bu çalışmanın sonucunda büyük örnek ve λ oranının büyük değeri için karesel fonksiyonun doğrusal fonksiyondan daha iyi sonuç verdiği elde edilmiştir.²⁷⁴ 1969'da Gilbert²⁷⁵ tarafından yapılan bir diğer çalışmada da çoklu normallik varsayımı sağlanırken varyans kovaryans matrisinin eşit olmadığı durum için doğrusal ve karesel sınıflandırma kuralına göre farklı bağımsız değişken sayısı ve örnek büyüklüğü için grupların sınıflandırılmasında farklı sonuçlar elde edildiği gösterilmiştir.

2.1.4 Diskriminant Fonksiyonunun Test Edilmesi ve Yorumlanması

İki grup için bağımsız değişkenlerin anlamlı şekilde farklı olup olmadığının testleri tek tek yapılabileceği gibi değişkenlerin birlikte anlamlı olup olmadığının da test edilmesi gerekmektedir. Bu bağlamda testler tek ve çok değişkenli testler olmak üzere ikiye ayrılmaktadır.*

- *Tek Değişkenli Testler*

Tek değişkenli testlerde modelde sanki tek değişken varmış gibi düşünülerek testler yapılmaktadır. Bu çalışmada *Student t* testi kullanılmakla birlikte bilgisayar paket program (SPSS) çıktısında Wilks' Lamda ve tek değişkenli *F* test sonuçları da yer almaktadır. *F* istatistik değeri, *Student t* istatistik değerinin karesine eşittir ve değişkenlerin tek tek gruplararası ve gruplarıçi varyanslarının birbirlerine

²⁷³ C. A. B. Smith, "Some Examples of Discrimination", **Annals of Eugenics**, 1947, 13, pp. 272–282.

²⁷⁴ S. Marks, O. J. Dunn, "Discriminant Functions When Covariance Matrices are Unequal", **Journal of the American Statistical Association**, 1974, 69, pp. 555–559.

²⁷⁵ Ethel S. Gilbert, "The Effect of Unequal Variance-Covariance Matrices on Fisher's Linear Discriminant Function", **Biometrics**, 1969, 25, pp. 505–515.

* Anlamlılık testlerinden tek değişkenli için *Student t* testi ve çok değişkenli için de *Hotelling T²* testi burada ele alınmıştır. Bunun dışındaki anlamlılık testleri (*Wilks'λ* ve *Rao V*) için Neyran Orhunbilge'nin **Çok Değişkenli İstatistik Yöntemler** kitabı sayfa 224-225'e bakınız.

oranlanması ile de hesaplanabilmektedir. Wilks' Lamda istatistiği ise her bir değişkenin grupları içi sapmasının toplam sapmaya oranından bulunmaktadır. Bağımsız değişkenlerin katsayılarının sıfıra eşit olduğunu ileri süren sıfır hipotezinin test sonuçlarının anlamlı olması beklenmektedir.

- *Çok Değişkenli Testler*

Modeldeki bağımsız değişkenlerin hepsinin birlikte anlamlı olup olmadığının testinin yapılması çok değişkenli testler olarak adlandırılmaktadır. *Student t* testinin karesine benzer olan ve çok değişkenli testlerde kullanılan istatistik *Hotelling T²* istatistiğidir. Bu test sadece bir değişkeni değil model için tanımlanan tüm değişkenleri içermektedir ve birden fazla değişken test edildiği için grupların ortalamaları yerine grup merkezlerinin farklarını incelemektedir. Ayrıca tek değişken incelendiğinde formülde $1/s_w^2$ olarak yer alan değer yerine gruplararası kovaryans matrisini ifade eden (C_w) değerinin tersi alınmaktadır.²⁷⁶

$$t^2 = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\frac{s_w^2}{n_1} + \frac{s_w^2}{n_2}} = \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_1 - \bar{X}_2) \frac{1}{s_w^2} (\bar{X}_1 - \bar{X}_2)$$

Hotelling T² test istatistiği

$$\left[\frac{(n - p - 1)}{p(n - 2)} \right] T^2$$

formülü ile hesaplanmakla birlikte p ve $n - p - 1$ serbestlik derecesi ile F dağılımına uymaktadır.

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} d' C_w^{-1} d$$

²⁷⁶ Jacques Tacq, **a.g.e.**, pp. 250–251.

Formülde T^2 eşitliğindeki $d' C_w^{-1} d$ kısmı Mahalanobis uzaklığını vermektedir.* *Hotelling* T^2 testi dışında bağımlı değişkenin iki kategoriden fazla olduğu durumlar için geliştirilmiş en genel çok değişkenli test olan Wilks' Lamda (bkz. sayfa 107) bağımlı değişkenin iki kategorili olduğu durumlar için de kullanılabilir.²⁷⁷

2.1.5 Gözlemlerin Gruplarının Belirlenmesi

Veri setindeki birimlerin veya yeni yapılacak gözlemlerin hangi gruba ait olduğunu belirlemek için temelde iki farklı yol izlenmektedir.

1. Gözlemlerin her bir grup ortalama vektörüne olan uzaklıkları belirlenip, bu uzaklık içinden değer olarak en küçük olan grup hangisi ise gözlem o gruba atanır.²⁷⁸
2. Gözlemin hangi sınıfta yer alacağını belirlemeye yarayan bir diğer yöntem de kriter değerinin kullanılmasıdır. En genel hali ile kriter değeri şu şekilde ifade edilmektedir:²⁷⁹

$$(\bar{X}_1 - \bar{X}_2)' S^{-1} x_0 \geq \frac{1}{2} [(\bar{X}_1 - \bar{X}_2)' S^{-1} (\bar{X}_1 - \bar{X}_2)] + \ln \left[\left(\frac{C(1|2)}{C(2|1)} \right) \left(\frac{p_2}{p_1} \right) \right]$$

Yukarıdaki eşitsizliğin sağlanması durumunda x_0 gözlemi (incelenen gözleme ait değişkenlerin değerlerinin yer aldığı vektör) 1. gruba atanmaktadır. Eşitsizliğin sağlanmadığı durumda ise gözlem 2. gruba atanmaktadır. Buradaki p_1 ve p_2 sırası ile 1. ve 2. gruba ait ön olasılık değerleridir. $C(1|2)$ ve $C(2|1)$ de sırası ile gözlemin gerçekte 2. grupta iken 1. gruba atanması ile gerçekte 1. grupta iken 2. gruba atanmasının maliyetini (cost) göstermektedir.

* *Hotelling* T^2 , *Mahalanobis* D^2 uzaklığını içermektedir: $D^2 = d' C_w^{-1} d$

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} d' C_w^{-1} d = \frac{n_1 n_2}{n_1 + n_2} D^2$$

Buradan F testi de D^2 'nin beklenen değeri tarafından şu şekilde ifade edilebilmektedir:

$$F = \frac{n - p - 1}{p(n - 2)} \frac{n_1 n_2}{n_1 + n_2} D^2$$

²⁷⁷ Jacques Tacq, **a.g.e.**, p. 348.

²⁷⁸ Kazım Özdamar, Cilt 2, **a.g.e.**, s. 381.

²⁷⁹ Richard A. Johnson, Dean W. Wichern, **a.g.e.**, p. 499,505.

Gözlemlerin gruplara ait olma ön olasılıklarının ve yanlış atama maliyetlerinin eşit olduğu durumlar için ise yukarıdaki eşitsizliğin sağ tarafındaki ifade

$$\frac{1}{2}[(\bar{X}_1 - \bar{X}_2)'S^{-1}(\bar{X}_1 - \bar{X}_2)]$$

olarak değişmektedir.

2.1.6 Diskriminant Analizinin Ayırıcı Gücünün Değerlendirilmesi

Yeni gözlemlerin iki gruptan hangisine ait olduğunun belirlenmesi Diskriminant Analizi'nin hedeflerinden bir diğeridir. Sınıflandırma genelde Diskriminant Analizi'nden bağımsız bir süreç olarak düşünülmesine karşın, birçok kaynak ve istatistik paket programında bu süreç Diskriminant Analizi'nin bir parçası olarak ele alınmaktadır ve fonksiyondan elde edilen diskriminant değerleri kullanılarak uygulanmaktadır. Diskriminant değerlerinden oluşan bu uzay yanlış sınıflandırma hatasını veya maliyetini minimize edecek şekilde kritik değer veya ayırıcı değer olarak adlandırılan diskriminant değeriyle ikiye bölünmektedir. Bu şekilde, gözlemin hangi gruba ait olduğuna karar verilmektedir.²⁸⁰

Genel uygunluğun değerlendirilmesi için sınıflandırma matrisi (classification matrix) oluşturulmaktadır. Bu matris oluşturulmadan önce her birimin bağımsız değişkenlere ait değerleri diskriminant fonksiyonunda yerine yazılarak hangi gruba ait olduğuna karar verilmektedir. Bu işlemler eğer veri seti yeterince büyükse fonksiyonu oluşturmak ve sınıfları belirlemek için kullanılmak üzere iki parçaya ayrılır ve bir grup gözlem ile fonksiyon yazıldıktan sonra kullanılmayan diğer gözlemlerle de sınıflandırma matrisi elde edilmektedir. Bu aşamadan sonra sınıflandırma matrisinin ayırma gücünün istatistik testi için *Press's Q* istatistiği kullanılmaktadır. Bu istatistik değeri, doğru sınıflandırma sayısını toplam örnek büyüklüğü ve grup sayısı ile karşılaştıran basit bir ölçümdür. Bulunan değer 1 serbestlik dereceli χ^2 dağılımından elde edilen kritik değerle karşılaştırılır. Eğer hesaplanan değer kritik değerden büyükse, sınıflandırma matrisinin istatistik olarak anlamlı ve

²⁸⁰ Ali Sait Albayrak, a.g.e., ss. 319–320.

yorumlanabilir olduğu anlamına gelmektedir. Q istatistiği şu şekilde hesaplanmaktadır:²⁸¹

$$Press's Q = \frac{[n - (n_d g)]^2}{n(g - 1)}$$

Burada n ; toplam örnek büyüklüğü, n_d ; doğru sınıflandırılan birim sayısı, g ; grup sayısını göstermektedir.

Sınıflandırma işlemi tamamlandıktan sonra yapılması gereken bir diğer işlem de yanlış sınıflandırılan birimlerin saptanması ve yanlış sınıflandırmaya neden olan durumun incelenmesidir.²⁸²

2.2 Çok Kategorili Diskriminant Analizi

Çok Kategorili Diskriminant Analizi, İki Kategorili Diskriminant Analizi'nin genelleştirilmiş şeklidir. Ayırıcı değişkenlerin seçimi ve varsayımlar ve Başlık 2.1.1 ve 2.1.2'de İki Kategorili Diskriminant Analizi için yapılan açıklamalarla aynıdır. Burada yalnızca iki analiz arasındaki farklılıklar ele alınmıştır.

Bağımlı değişkenin ikiden çok kategorisi olması durumunda kullanılan Diskriminant Analiz tekniğidir. Çok Kategorili Diskriminant Analizi'nde gözlemlerin, kategoriler arasındaki ayırma gücü en yüksek olacak biçimde belli sayıda doğrusal bağlantılar yardımıyla sınıflandırması amaçlanmaktadır. Burada belirlenmesi gereken konu ayırımı en iyi şekilde yapabilecek doğrusal fonksiyon sayısına karar vermektir. İkiden çok grup olması durumunda bulunacak maksimum diskriminant fonksiyonu sayısı g ; kategori sayısını, p ; değişken sayısını göstermek üzere $\min(g - 1, p)$ adet olacaktır.²⁸³

²⁸¹ Joseph F. Hair, v.d., **a.g.e.**, p. 303.

²⁸² **A.e.**, p. 304.

²⁸³ Subhash Sharma, **a.g.e.**, p. 251.

2.2.1 Diskriminant Fonksiyonunun Elde Edilmesi

İkiden fazla grup olması durumunda, ayırıcı fonksiyonların elde edilmesi, yine gruplararası varyans ile gruplariçi varyans oranının maksimum olacak şekilde oluşturulmasına dayanmaktadır. Buradan $\min(g - 1, p)$ olmak üzere elde edilecek özdeğerlere (λ_i) , karşılık gelen özvektörler diskriminant fonksiyonlarını oluşturmaktadır. Diskriminant fonksiyonlarıyla elde edilecek değerler arasında korelasyon olmaması, üzerinde durulması gereken önemli bir noktadır. Bu şartın sağlanması için mümkün tüm fonksiyonların denenmesi gerekmektedir.²⁸⁴

Diskriminant fonksiyonlarındaki bağımsız değişkenlerin grupları ayırma güçlerinin ya da fonksiyona katkılarının belirlenmesi özellikle yorumlama aşamasında önemli olacağından genellikle aşağıdaki şekilde standartlaştırılmaktadır²⁸⁵.

$$k_i^{*(j)} = (k_i^{(j)} W_{ii})^{1/2}$$

Burada W_{ii} ; gruplariçi varyasyon matrisinin köşegen değerleridir. Ayrıca $i = 1, 2, \dots, p$ ve $j = 1, 2, \dots, \min(g - 1, p)$ kadar değerler almaktadır.

2.2.2 Diskriminant Fonksiyonunun Anlamlılık Testleri

Gruplariçi değişimin toplam değişime oranından elde edilen *Wilks' Lamda* (Λ) değeri kullanılarak hesaplanan test istatistiği şu şekildedir;²⁸⁶

$$\chi^2 = - \left[n - 1 - \frac{(p + g)}{2} \right] \ln(\Lambda) \quad 2.6$$

ve bu değer $p(k - 1)$ serbestlik derecesi ile χ^2 dağılımına uymaktadır.

Tatsuoka, ayırıcı fonksiyonun sayısını belirlemede kullanılmak üzere, Λ oranını²⁸⁷

$$\Lambda = \prod_{i=1}^r \frac{1}{1 + \lambda_i}$$

²⁸⁴ Neyran Orhunbilge, **Çok Değişkenli İstatistik Yöntemler**, a.g.e., s. 244.

²⁸⁵ Hüseyin Tatlıdil, a.g.e., s. 267.

²⁸⁶ Kazım Özdamar, Cilt 2, a.g.e., s. 141.

²⁸⁷ Hüseyin Tatlıdil, a.g.e., s. 268.

eşitliği ile göstermiştir. Bu eşitliğin logaritması alındığında

$$\ln(\Lambda) = -[ln(1 + \lambda_1) + ln(1 + \lambda_2) + \dots + ln(1 + \lambda_r)]$$

ifadesine dönüşmekte ve Eşitlik 2.6'da yerine koyulabilir hale gelmektedir. Tüm mümkün diskriminant fonksiyonlarının ayırıcı özellik taşımadığının iddia edildiği sıfır hipotezini test etmek için kullanılan test istatistiği (Eşitlik 2.6) tekrar yazıldığında,²⁸⁸

$$\chi^2 = \left[n - 1 - \frac{(p + g)}{2} \right] \sum_{i=1}^r \ln(1 + \lambda_i) \quad 2.7$$

değeri elde edilir. p ; ayırıcı değişken sayısını, g ; grup sayısını ve r ; toplam ayırıcı fonksiyon sayısını göstermek üzere Eşitlik 2.7'deki bu değer $(p - r - 1)(g - r)$ serbestlik derecesi ile χ^2 dağılımına uymaktadır. Eğer bu test sonucunda sıfır hipotezi reddedilirse, öncelikle en büyük özdeğere sahip olan ilk diskriminant fonksiyonunun ayırıcı özellik taşıdığına karar verilir ve diğer ayırıcı fonksiyonların önemli olup olmadığının testi yapılmak üzere bu fonksiyona ait değer dışarıda bırakılarak hipotez testi tekrar kurulur. Kalan ayırıcı fonksiyonların testine aynı şekilde sıfır hipotezi kabul edilinceye kadar devam edilecektir.

2.2.3 Gözlemlerin Gruplarının Belirlenmesi

İki Kategorili Diskriminant Analizi'nde kullanılan yöntemler aynı şekilde geçerli olmakla birlikte, sadece ufak değişiklikler söz konusudur. Çok Kategorili Diskriminant Analizi'nde sınıflandırma kuralı şu şekilde ifade edilmektedir.²⁸⁹

$$-\frac{1}{2}[(x - \bar{x}_i)'S^{-1}(x - \bar{x}_i)] + \ln p_i$$

Buradaki p_i ; değeri i . guruba ait olmanın ön olasılığıdır. Eğer bu değer verilmemiş ise $p_i = n_i/n$ olarak elde edilebilir. x gözleminin her bir grup merkezinden olan

²⁸⁸ Neyran Orhunbilge, **Çok Değişkenli İstatistik Yöntemler**, a.g.e., s. 246.

²⁸⁹ Kazım Özdamar, Cilt 2, a.g.e., ss. 400–401.

farkları belirlendikten sonra gözlem bu değerler içinde sifira en yakın olan gruba atanır.

Gözlemlerin sınıflandırılmasında kullanılan bir diğer yol da ayırma fonksiyonundan yararlanılmasıdır. Burada x gözlem vektöründeki değerler ayırma fonksiyonlarında yerine konularak sınıflandırma skorları hesaplanmaktadır. Elde edilen sonuçlara göre gözlem, sıfırdan en büyük farklılığının hesaplandığı gruba atanmaktadır.

Gözlemlerin grupları belirlendikten sonra Çok Kategorili Diskriminant Analizi'nin başarısı da İki Kategorili Diskriminant Analizi'nde olduğu gibi değerlendirilmektedir.

2.3 Diskriminant Fonksiyonunun Dışsal Geçerliliği

Diskriminant Analizi sonucunda elde edilen diskriminant fonksiyonu, gözlemlerin sınıflandırılması için kullanılacaksa dışsal geçerliliğinin* (external validation) incelenmesi gerekmektedir. Diskriminant fonksiyonunun doğruluğunu referans alan dışsal geçerlilik, diğer örneklerdeki gözlemleri sınıflandırabilmelidir. Diskriminant fonksiyonunun tahmin edilmesinde kullanılan gözlemlerin aynı zamanda sınıflandırılmasından elde edilen hata oranı yanlı olmaktadır. Bu nedenle, fonksiyonun yazılmasında kullanılan gözlemler diskriminant fonksiyonunun geçerliliğini ölçmede kullanılmamalıdır. Diskriminant fonksiyonunun geçerliliğini ölçmek için sıklıkla kullanılan üç teknik vardır. Bu teknikler şu şekilde sıralanabilir:

- *Gizleme yöntemi (Holdout Method)*

Diskriminant Analizi'nin son adımında sonuçların geçerliliği için çapraz geçerlilik yöntemi de denilen Gizleme Yöntemi uygulanabilir. Bu yöntem en basit şekli ile üç adımdan oluşmaktadır. İlk olarak veri seti rastgele iki parçaya ayrılır. Gözlemlerin bir parçası kullanılarak diskriminant fonksiyonu elde edildikten sonra, fonksiyonun oluşturulmasında kullanılmayan gözlemlere ait değerler bu fonksiyonda yerine konularak doğru sınıflandırma oranları elde edilmektedir. Bazı araştırmacılar bu

* Dışsal geçerlilik (external validation), bir örneklem üzerinde elde edilen araştırma sonuçlarının daha büyük gruplara ve evrene genellenme derecesidir.

işlemi birkaç kez tekrarlayıp sınıflandırma işlemi sonucunda bulunan çıktıların ortalamalarını alarak doğru sınıflandırma oranının bulunmasının daha uygun olduğunu düşünmektedirler²⁹⁰. Ayrıca bu yöntemi kullanabilmek için veri setinin yeterince büyük olması gerekmektedir.

- *U-Yöntemi (U-Method)*

Bu yöntem Lachenbruch²⁹¹ tarafından 1967 yılında önerilmiştir. Her seferinde bir gözlem dışarıda tutarak $n - 1$ gözlemlerle diskriminant fonksiyonu tahmin edilip, fonksiyonu oluşturmak için kullanılmayan gözlemin değerleri elde edilen fonksiyonda yerine koyularak gözlemin sınıfı belirlenir ve bu işlem n defa tekrarlanarak elde tutulan n birim için sınıflama yapılır. Bu yöntem, sınıflama oranı açısından neredeyse yansız tahmin yapmakla birlikte varyans ve ortalama hata karelerinin küçük olmaması nedeniyle eleştirilmektedir.

- *Bootstrap Yöntemi*

İstatistikte çoğunlukla parametrelerin normal dağılım varsayımına dayanılarak tahmin edilmesi söz konusu iken Bootstrap Yöntemi'nde veri setinden rastgele iadeli örnekler alınarak parametrelerin, dağılım varsayımından bağımsız olarak tahmini yapılabilmektedir. Bootstrap, küçük veri setleri için de kullanılabilen bir yöntemdir.²⁹²

Bootstrap Yöntemi literatürde birkaç farklı şekilde adlandırılmaktadır. İlk olarak dağılım varsayımı olup olmadığından yola çıkılarak “Parametrik Bootstrap” ya da “Parametrik Olmayan Bootstrap” olmak üzere ikiye ayrılmıştır. İkinci olarak ise örneğin alınış şekline ve/veya parametrelerin hesaplanış şekline göre Basit Bootstrap,

²⁹⁰ Joseph F. Hair, v.d., **a.g.e.**, p. 311.

²⁹¹ P. A. Lachenbruch, “An Almost Unbiased Method of Obtaining Confidence Intervals for the Probability of Misclassification in Discriminant Analysis”, **Biometrics**, Vol: 23, No. 4, 1967, pp. 639–645.

²⁹² R. Wehrens, H. Putter, L. M. C. Buydens, “The Bootstrap: A Tutorial”, **Chemometrics and Intelligent Laboratory Systems**, Vol: 54, 2000, pp. 35–52.

Çift Bootstrap, Ağırlıklı Bootstrap, Tekrarlamalı Bootstrap gibi farklı isimler almaktadır.²⁹³

En genel şekli ile veri kümesinde yerine koyma yöntemi ile örnekler seçilerek öğrenme kümesi oluşturulur. n örnekten oluşan veri kümesinden yerine konularak yine n örnek seçilir. Bu küme öğrenme kümesi olarak ele alınmaktadır. Öğrenme kümesinde yer almayan örnekler sına kümesi olarak kullanılır. n örnekten oluşan veri kümesinde bir örneğin seçilmeme olasılığı $1 - \frac{1}{n}$ ve sına kümesinde yer alma olasılığı $\lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right)^n = e^{-1} = 0,368$ 'dir. Diğer bir ifadeyle, öğrenme kümesi veri kümesindeki örneklerin %63,2'sinden oluşmaktadır. Buradan modelin başarısını sadece sına kümesi kullanarak belirlemek kötümser bir yaklaşımdır. Model başarısı hem öğrenme kümesinin hem de sına kümesinin başarısı ile değerlendirilmelidir. Bunun için $hata = 0,632hata_{sınama} + 0,368hata_{öğrenme}$ değeri hesaplanır ve işlem birkaç kez tekrarlanıp hataların ortalaması bulunarak diskriminant fonksiyonunun dışsal geçerliliği test edilmektedir.²⁹⁴

Diskriminant fonksiyonunun dışsal geçerliliğini ölçmek için kullanılan Gizleme, U ve Bootstrap yöntemleri arasından son dönemde, bilgisayar paket programların gelişimiyle de birlikte en yaygın kullanılan Bootstrap yöntemidir.

²⁹³ M. H. Aslan, F. Yıldırım, "Endüstri Mühendisliği Yaşam Süresi Analizlerinde Yeniden Örnekleme Yöntemlerinin Kullanılması", **2. Mühendislik ve Teknoloji Sempozyumu, 30 Nisan- 1 Mayıs 2009**, Çankaya Üniversitesi, s. 169.

²⁹⁴ Ş. G. Öğüdücü, "Veri Madenciliği Farklı Sınıflandırma Yöntemleri", (Çevirimiçi) <http://web.itu.edu.tr/~sgunduz/courses/verimaden>, 15 Ağustos 2011.

ÜÇÜNCÜ BÖLÜM: İŞLETME PERFORMANSINI ETKİLEYEBİLECEK FAKTÖRLERİN ARAŞTIRILMASI

Genel bir tanımlama ile işletme performansı, faaliyetlerinin kalitesi olarak ifade edilebilir. İşletmelerin performansı, yöneticilerin maliyet kar dengesini en iyi şekilde yönetmesiyle ilgilidir²⁹⁵. İşletmelerin performanslarının değerlendirilmesinde çok farklı kriterler kullanılmaktadır. Performansı etkileyen faktörlerin en iyi şekilde saptanması, işletmenin belirlediği finansal ve faaliyetlerine ilişkin hedeflerine ulaşmak için kaynaklarını en etkin şekilde kullanması açısından önemlidir.

Bu çalışmada işletmelerin performanslarını etkileyebilecek değişkenler çeşitli performans göstergeleri kombinasyonlarıyla analiz edilerek işletmelerin belirlediği hedefler doğrultusunda hangi değişkenlerin önemli olduğu saptanmaya çalışılmıştır.

3.1 İşletme Performans Kriterleri

Sosyal ve yardım amaçlı kuruluşlar dışında işletmelerin amacı mali dönemlerinin sonunda elde ettikleri karı en yüksek düzeye taşımaktır (maksimize etmektir).²⁹⁶ İşletmelerin “karlılık oranları”nın yüksek olması, işletmelerin ulaşması gereken önemli hedeflerden biri olmakla birlikte işletmelerin sürekliliğini sağlamak için yeterli değildir. Karın maksimize edilmesi, işletmelerin faaliyetleri ya da finansal kararları ile bağlantılı olarak gerçekleşebilir. İşletmelerin karlarını kısa vadeli yatırımlarla (hisse senedi, hazine bonusu vb.) yükseltebilseler bile bu durum işletmelerin, faaliyetlerinden elde edecekleri nakit girişlerini olumsuz yönde etkileyerek taşıdıkları riski arttırabilir. Riskin artması ise diğer bir performans göstergesi olan “büyüme”yi etkileyecektir. Sadece büyüme odaklı bir hedef belirlemek ise işletmelerin kontrolsüz borçlanmasına neden olabilir.

²⁹⁵ Anil M. Pandya, Narender V. Rao, “Deversification and Firm Performance an Emprical Evaluation”, **Journal of Finacial and Strategic Decisions**, V.II. N. 2, 1998, p. 67.

²⁹⁶ N. Bayyurt, “İşletmelerde Performans Değerlendirmenin Önemi ve Performans Göstergeleri Arasındaki İlişkiler”, **Sosyal Siyaset Konferansları Dergisi**, Sayı 53, 2007/2 Prof. Dr. Haşmet Başar’a Armağan Özel Sayı, s. 580.

Sanayi işletmelerinde, “verimlilik” de yine bir performans göstergesi olarak ele alınmaktadır. Kaynakların verimli bir şekilde kullanılması sonucunda işletmelerin maliyetlerinde önemli düşüşler meydana gelerek kapasite kullanım oranında ve toplam karında artışlar gerçekleşmektedir. Verimlilik artışı ise kapasite kullanım oranlarını yükseltmekte ve birim maliyetlerde önemli düşüşler elde edilmesine neden olabilmektedir. Sanayi işletmelerinin verimli üretim yapmaları ile ekonomide; sürekli, sağlıklı ve dengeli gelişme gerçekleşebilecek, enflasyonla mücadele, ihracatı artırma ve refahı yükseltme de mümkün olabilecektir.²⁹⁷ İşletmelerin verimlilik hedefi ekonomiye ve işletmelere birçok fayda sağlamasına rağmen tek başına yeterli bir başarı göstergesi değildir.

İşletmeler için farklı karlılık, büyüme, verimlilik vb. hedefler mümkünse de işletmelerin tek bir hedefe yönelmeleri işletmeler için yetersiz ve hatta riskli olmaktadır. Bu nedenle işletmeler performans göstergesi olarak kabul edilen bu hedeflerin hepsini birlikte düzeltmeyi amaçlamalıdır.²⁹⁸

3.1.1 Karlılık Oranları

İşletmenin faaliyetleri ile sonucunda elde etmiş olduğu karların karşılaştırıldığı bu oranlar faaliyetlerin ne kadar etkin olduğunu ortaya koymaktadır. Karlılık oranları işletme sahipleri açısından, yabancı kaynakların sahiplerine ödenen anapara ve faizlerin işletmenin faaliyetleri sonucunda elde ettiği kardan mı yoksa düşük karlılık oranı nedeniyle öz sermayeden mi karşılandığını ortaya koyması sebebiyle önemlidir. Diğer taraftan bu oranlar işletmenin piyasadaki başarısı açısından yöneticiler, personel, kamu kuruluşları, kredi verenler, potansiyel ortaklıklar tarafından da dikkatle izlenilmelidirler.²⁹⁹

Bilançonun farklı kalemlerinin kullanılmasıyla elde edilen çeşitli karlılık oranı hesaplamaları mevcuttur. Literatürde sık kullanılan karlılık oranlarına aşağıda değinilmiştir.

²⁹⁷ M. Tekin, “Sanayi İşletmelerinde Verimlilik ve Önemi”, **Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi**, Sayı: 1, 1992, s. 170.

²⁹⁸ N. Bayyurt, **a.g.e.**, s. 581.

²⁹⁹ M. Tükenmez, vd., **Finansal Yönetim**, Vizyon Eğitim ve Danışmanlık Ltd. Şti., İzmir 1999, ss. 413-414.

Öz Sermaye Karlılığı:

Öz sermaye karlılığı farklı kar kalemleri ile hesaplanabilmektedir.

$$\text{Öz Sermaye Karlılığı} = \frac{\text{Net Dönem Karı}}{\text{Öz Sermaye}}$$

oranın sonucu, işletme sahipleri açısından öz sermayenin alternatif kullanım alanlarının değerlendirilmesi nedeniyle önem taşımaktadır. Alternatif kullanım alanlarının kar getirisi bu orandan daha yüksek ise öz sermaye sahipleri alternatif alanlara yöneleceklerdir.³⁰⁰

Ekonomik Karlılık Oranı:

$$\text{Ekonomik Karlılık} = \frac{(\text{Vergi Öncesi Kar} + \text{Finansal Giderler})}{\text{Toplam Varlıklar}}$$

olarak hesaplanan bu oran toplam kaynakların ne ölçüde karlı kullanıldığını göstermekte³⁰¹ ve işletmeye yatırılan fonların getirisini ölçmekte kullanılmaktadır.

Aktif Karlılık:

$$\text{Aktif Karlılık} = \frac{\text{Net Dönem Karı}}{\text{Toplam Varlıklar}}$$

olarak ifade edilen bu oran, varlıkların ne ölçüde verimli kullanıldığını göstermektedir ve işletmelerin performans göstergelerinden olan çeşitli “karlılık oranı” hesaplamaları içinde önemli bir yer almaktadır.

Faaliyet Kar Marjı:

$$\text{Faaliyet Kar Marjı} = \frac{\text{Faaliyet Karı}}{\text{Satış Gelirleri}}$$

şeklinde hesaplanan faaliyet kar marjının yüksek olması, işletmenin ana faaliyet konusunda başarılı olduğunu ifade etmektedir. Ancak oranın, faaliyet dışı gider ve

³⁰⁰ M. Tükenmez, vd., **a.g.e.**, s. 418.

³⁰¹ **A.e.**, s. 419.

zararlar ile vergiyi karşıladıktan sonra yeterli ölçüde dönem net karı bırakacak büyüklükte olması önem taşımaktadır.³⁰²

Brüt Kar Marjı:

İşletmeler açısından

$$\text{Brüt Kar Marjı} = \frac{\text{Brüt Kar}}{\text{Satış Gelirleri}}$$

şeklinde hesaplanan bu oranın yeterliliği, aynı sektördeki işletmelerin oranları ile karşılaştırılarak ve işletmenin geçmiş yıllardaki oranları dikkate alınarak belirlenebilmektedir³⁰³. İşletmelerin brüt kar marjının değişmesinde hangi etkenlerin geçerli olduğu saptanmalıdır³⁰⁴. Satışlar ve satılan ürün maliyeti bu oranın yüksek çıkmasında önemli bir nedendir.

Net Kar Marjı:

Aşağıda gösterildiği gibi hesaplanan

$$\text{Net Kar Marjı} = \frac{\text{Net Dönem Karı}}{\text{Satış Gelirleri}}$$

'nın yüksek çıkması işletmenin satışlarını yüksek karla gerçekleştirdiğini göstermektedir. Oranın yeterliliği için belirli bir standart yoktur ancak aynı sektördeki diğer işletmelere ve geçmiş yıllardaki oranlara göre değerlendirilebilmektedir.

Vergi Öncesi Kar Marjı:

$$\text{Vergi Öncesi Kar Marjı} = \frac{\text{Vergi Öncesi Kar}}{\text{Satış Gelirleri}}$$

³⁰² N. Kurnaz, "Finansal (Mali) Tablolar Analizi", (Çevrimiçi) <http://www.niyazikurnaz.net>, 12 Temmuz 2011.

³⁰³ M. Tükenmez, vd., **a.g.e.**, s. 415.

³⁰⁴ Ö. Akgüç, **Finansal Yönetim**, Avcıol Basım, Yenilenmiş 7. Baskı, İstanbul 1998, s. 71.

olarak hesaplanan bu gösterge de net kar marjı ile aynı şekilde yorumlanmaktadır. İki oranın da yüksek çıkması beklenmektedir.

Faiz Karşılama Oranı:

İşletmenin faaliyetleri sonucunda elde ettiği fonlar aracılığı ile ödemek zorunda olduğu faizleri kaç defa kazandığını ölçmek amacıyla kullanılmaktadır³⁰⁵. İşletmenin faiz giderlerini karşılayamaması, işletmeyi mali açıdan zor durumda bırakarak iflasına neden olabilmektedir.³⁰⁶

$$\text{Faiz Karşılama Oranı} = \frac{\text{Faaliyet Karı}}{\text{Finansal Giderler}}$$

3.1.2 Verimlilik

Verimlilik, üretilen miktarla bu üretimde kullanılan herhangi bir kaynak arasındaki oranı göstermektedir. Üretim faktörlerinin her biri için verimlilik, üretim faktörlerinin sağladığı artışla ifade edilebilmektedir. İşletme performans göstergelerinden olan “verimlilik” için de karlılık oranının da olduğu gibi farklı yaklaşımlar kullanılmaktadır. Bu yaklaşımlardan emek (işgücü) verimliliği belirli bir dönemde mevcut işgücünün üretim miktarında sağladığı artışın dönem başındaki toplam üretime oranlanmasıyla hesaplanmaktadır. Makine verimliliği ise işletmede belirli bir dönemde mevcut makinelerle yapılan üretimdeki artışın dönem başındaki toplam üretime oranlanmasıyla elde edilmektedir. Diğer bir yaklaşımda sermaye verimliliği yani rantabilite (karlılık), mevcut sermayede belirli bir dönemdeki artışın dönembaşındaki sermaye miktarına oranlanması sonucu elde edilmektedir.³⁰⁷

Verimliliğin artırılması, özellikle sanayi kesiminde ülkenin gelişimini sürdürebilmesi açısından önem taşımaktadır. Verimlilikteki düşüş, enflasyon seviyesi ve istihdamı olumsuz yönde etkilediği gibi şirket düzeyinde de karlılığı

³⁰⁵ M. Tükenmez, vd., **a.g.e.**, s. 425.

³⁰⁶ N. Akdoğan, N. Tenker, **Finansal Tablolar ve Mali Analiz Teknikleri**, 4. Baskı İstanbul, 1994, s. 452.

³⁰⁷ M. Tekin, **a.g.e.**, s. 170.

azaltmaktadır. Bu sebeple özellikle son yıllarda makro ve mikro düzeyde verimlilik artışını sürekli kılmak önemli olmuştur.³⁰⁸

Kaynakların verimli bir şekilde kullanılması sonucunda işletmelerin kapasite kullanım oranında ve toplam karında önemli artışlar sağlanabilmektedir.³⁰⁹

Emek verimliliği, sermaye verimliliği, hammadde verimliliği gibi farklı girdilerin verimliliği gibi kısmi verimlilikleri hesaplamak mümkündür. Bunun dışında verimlilik ölçüsü olarak çok faktörlü verimlilik ve toplam faktörlü verimlilik oranları da kullanılabilir.* Ancak hesaplamadaki zorluklar nedeniyle en çok kullanılan kısmi verimlilik oranları olup bu oranlar içinde de emek verimliliğidir ve aksi belirtilmedikçe verimlilik ifadesiyle emek verimliliğinden bahsedilmektedir. Bu çalışmada işletme performans göstergelerinden olan verimlilik için emek kullanılmış ve oran,

$$\text{Emek Verimliliği} = \frac{\text{Katma Değer}}{\text{Çalışan Sayısı}}$$

olarak hesaplanabilmiştir.

3.1.3 Büyüme Oranları

İşletmelerin büyüme oranları; satışlardaki artış, vergi öncesi kar (dönem karı), öz sermaye varlık (aktif) toplamı ve hisse başına kar olarak hesaplanabilmektedir.

Satışlardaki Artış:

Satışlardaki artış cari fiyatlarla hesaplanmaktadır.³¹⁰

$$\Delta S = \frac{S_t - S_{t-1}}{S_{t-1}} * 100$$

³⁰⁸ A. Tezeren, **İmalat Sanayinde Verimliliği Etkileyen Faktörler**, 319 Nolu Milli Produktivite Merkezi Yayınları, Ankara, 1985.

³⁰⁹ M. Tekin, **a.e.**, ss. 171–172.

* Ayrıntılı bilgi için bkz. Bilge Acar Bolat, **Firma Performansını Etkileyen Faktörlerin Yapısal Eşitlik Modeli ile İncelenmesi**, Basılmış Doktora Tezi, İstanbul, 2010, ss. 32–39.

³¹⁰ Ö. Akgüç, **Finansal Yönetim**, **a.g.e.**, ss. 80–81.

Δs ; net satışlardaki % artış, S_t ; t . dönemdeki cari fiyatlarla net satış tutarı, S_{t-1} ; bir önceki dönemdeki net satış tutarını belirtmektedir.

Fiyat artışlarının hızlı olduğu yani fiyat hareketlerinin istikrarlı olmadığı ekonomilerde, cari fiyatlara göre hesaplanan artış hızlarının, fiyat artış hızı dikkate alınarak düzeltilmesi gerekmektedir. Bu amaçla cari fiyatlar yerine reel fiyatlarla hesaplanan artışlar kullanılmaktadır.

$$\Delta s_r = \frac{\%100 + \% \Delta s}{\%100 + \% \Delta F} - 1$$

Δs_r ; bir önceki döneme göre satışlardaki reel olarak artış oranını, ΔF ; bir önceki yıla göre fiyatlardaki ortalama artış yüzdesini ifade etmektedir.

Dönem Karında Artış:

Cari fiyatlarla dönem karındaki veya vergi öncesi kardaki artış oranı aşağıdaki gibi hesaplanmaktadır.³¹¹

$$\Delta k = \frac{K_t - K_{t-1}}{K_{t-1}} * 100$$

Δk ; cari fiyatlarla dönem karındaki % artış, K_t ; cari fiyatlarla dönem karını, K_{t-1} ; bir önceki dönem karını ifade etmektedir.

Enflasyonun etkisinden arındırılmak amacıyla reel kar artış hızı (Δk_r) hesaplanmaktadır.

$$\Delta k_r = \frac{\%100 + \% \Delta k}{\%100 + \% \Delta F} - 1$$

Öz Sermaye Artışı:

$$\Delta \ddot{o} = \frac{\ddot{o}_t - \ddot{o}_{t-1}}{\ddot{o}_{t-1}} * 100$$

³¹¹ Ö. Akgüç, **Finansal Yönetim**, a.g.e., ss. 81–82.

$\Delta\ddot{o}$; öz sermaye artışı, \ddot{o}_t ; cari dönem öz sermaye tutarı, \ddot{o}_{t-1} ; bir önceki dönem öz sermaye değeridir.

Bir işletmenin öz sermayesi, bir akışı değil belirli bir tarih itibariyle stoku, başka bir ifadeyle birikmeli bir tutarı gösterdiğinden sabit fiyatlarla veya reel olarak artışı hesaplamak için tüm bilanço kalemlerinin enflasyona göre düzeltilmesi gerekmektedir.³¹²

Varlık (Aktif) Artışı:

Varlık toplamındaki artış cari fiyatlarla

$$\Delta v = \frac{v_t - v_{t-1}}{v_{t-1}} * 100$$

olarak hesaplanmaktadır.

v_t ; cari dönem varlık tutarı, v_{t-1} ise bir önceki dönem varlık tutarını ifade etmektedir. “*Varlıklarda reel artışı hesaplayabilmek için her bir varlık kaleminin hangi tarihte işletmenin varlık tutarına dahil edildiğinin bilinmesi ve o tarih esas alınarak fiyat artışlarına göre düzeltme yapılması gerekmektedir.*”³¹³

İşletme bazlı incelendiğinde büyüme kavramının ekonomik büyüme ve buna bağlı olarak istihdam, verimlilik ve gelişen teknoloji ile ilişkili olduğu görülmektedir. İşletmelerin büyümesini etkileyen sadece kendi iç dinamikleri ve faaliyet alanları değil aynı zamanda faaliyet gösterdiği ülkenin belirlediği ekonomi stratejileridir. Bu bağlamda ülkenin büyüme stratejileri işletmelerin büyümesini etkileyebilmesine rağmen, işletmelerin büyüme konusu özellikle iktisat teorisinde hakettiği şekilde incelenmemektedir. Mikro ekonomik analizlerde işletmelerin faaliyet gösterdikleri endüstriyel kolların dinamiklerine ilişkin bilgiler yetersiz kalmaktadır.³¹⁴

³¹² Ö. Akgüç, **Finansal Yönetim**, a.g.e., s. 83.

³¹³ **A.e.**, s. 83.

³¹⁴ M. Karagöz, H. Demirgil, “Firma Büyüme Performansını Etkileyen Faktörler”, **Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi**, 2009, C. 14, S.2, s. 48.

İşletmelerin büyümesini etkileyen faktörlerin belirlenmesi ve daha iyi yorumlanması işletme politikalarının oluşturulmasında önemli etkenlerden biridir. Ülkelerin belirledikleri ekonomik stratejiler işletmelerin bu ekonomik faaliyetleri dikkate almasını ve bu stratejiler yönünde hareket etmelerini gerektirmektedir. Türkiye gibi ihracata dayalı olarak büyüyen ülkelerde,³¹⁵ ihracatın ithalatı karşılama oranı yıllar itibariyle aşağıdaki Tablo 3.1’de görülebilir. Yıllar itibariyle bu oranın yükselmesi Türkiye’nin ihracat yoluyla elde edilen gelirin ithalatı karşıladığı anlamına gelmektedir. İhracatın ithalatı finanse etmesinin derecesi ihracatın ithalatı karşılama oranıyla açıklanır.³¹⁶

Tablo 3.1 Yıllara göre İhracatın İthalatı Karşılama ve İhracat Değişim Oranları

Dönem	Yıllara göre İhracatın İthalatı Karşılama Oranı (%)	Yıllara göre İhracat (Değişim Oranı) (%)	Dönem	Yıllara göre İhracatın İthalatı Karşılama Oranı (%)	Yıllara göre İhracat (Değişim Oranı) (%)
2010	61.4	11.5	2003	68.1	31.0
2009	61.4	11.5	2002	69.9	15.1
2008	65.4	23.1	2001	75.7	12.8
2007	63.1	25.4	2000	50.9	4.5
2006	61.3	16.4	1999	65.4	-1.4
2005	62.9	16.3	1998	58.7	2.7
2004	64.8	33.7			

Kaynak: <http://www.tuik.gov.tr/Gosterge.do?metod=IlgiliGosterge&id=3489>

İhracat yolu ile artan gelirler ve Türkiye’de işletmeler arasındaki rekabet nedeniyle satış hacimlerinde ihracatın payının önemini arttırmıştır. İşletmelerin ihracat hacimlerindeki artış büyüme stratejilerine önem vermelerini zorunlu kılmaktadır. Bu

³¹⁵ S. Çiftçioğlu, C. Karaaslan, H. Demir, “Saving, Investment, Exports and Economic Growth in Turkey”, **Review of Social, Economic & Business**, Vol: 3–4, ss. 20–30.

C. Bilgin, A. Şahbaz, “Türkiye’de Büyüme ve İhracat Arasındaki Nedensellik İlişkisi”, **Gaziantep Üniversitesi Sosyal Bilimler Dergisi**, 2009, (8).1, ss. 177–198.

³¹⁶ (Çevirimiçi) <http://www.havvatunc.com/ticaret-acigi-buyuyor-ihracatin-ithalati-karsima-orani-kuculuyor-1107.html>, 10 Eylül 2011.

artan dış gelirlerle beraber işletmelerin döviz kuru rejimleri, enflasyon gibi ekonomik politikalardan etkilenme riskini arttırabilmektedir.

İşletmelerin performanslarını arttıran, yenilikçi faaliyet ve ürünün geliştirilmesine yönelik düzenli Ar-Ge faaliyetleridir. Bir işletmenin hızlı büyümesi yenilikçi faaliyetlerinin sağladığı olumlu katkıların bir sonucu olarak değerlendirilebilmektedir. Piyasadaki mevcut işletmelerin pazar paylarını alabilen hızlı büyüyen yeni işletmeler ürün ve hizmetlerin daha etkin yürütülmesi için yeni yöntemlerin kullanılmasında daha başarılı olabilmektedirler.³¹⁷

3.1.4 Piyasa Performans Oranları

Hisse Başına Düşen Kar:

Bu oran işletme sahip ve ortaklarının yaptıkları yatırımlar karşılığında yeterli gelir elde edip etmediklerini ifade etmektedir ve vergi sonrası net kardan, imtiyaz veren kurucu hisseler, yönetim kurulu ve çalışanlara ödenen temettü çıkarıldıktan sonra, kalan net karın ortalama hisse senedi sayısına bölünmesi ile bulunmaktadır.³¹⁸

$$\text{Hisse Başına Düşen Kar} = \frac{\text{Net Kar}}{\text{Hisse Senedi Sayısı}}$$

Fiyat-Kazanç Oranı:

Bu oran hisse senedi fiyatının hisse başına kar değerine oranıdır.³¹⁹

$$\text{Fiyat - Kazanç} = \frac{\text{Piyasa Değeri}}{\text{Net Kar}}$$

Fiyat-Kazanç oranı yatırımcılar arasında en çok kullanılan değerlendirme yöntemidir. Bu oranın fazla kullanılmasının temel sebepleri şunlardır.³²⁰

³¹⁷ M. Karagöz, H. Demirgil, a.g.e., ss. 61–65.

³¹⁸ **Sermaye Piyasası ve Borsa Temel Bilgiler Kılavuzu**, İstanbul Menkul Kıymetler Borsası Yayınları, 17. Basım, Temmuz 2002, s.141.

³¹⁹ A.e., s. 141.

³²⁰ M. B. Karan, **Yatırım Analizi ve Portföy Yönetimi**, Gazi Kitabevi, Ankara, 2001, s. 356.

- Bu Fiyat-Kazanç oranı kullanıldığında, geleneksel yöntemlerde* varsayımı ya da hesabı yapılan risk, büyüme oranı ve temmettö ödeme oranı gibi bir takım verileri kullanmaya gerek kalmamaktadır.
- Yatırımcılar açısından hesaplanması kolaydır.
- Risk ve büyüme gibi verileri yansıtmaktadır.

Piyasa Deęeri/Defter Deęeri Oranı (PD/DD):

Bu oran bir iřletmenin hisse senedi fiyatının hisse başına öz sermaye deęerine bölünmesiyle elde edilmektedir.³²¹

$$PD/DD \text{ Oranı} = \frac{\text{Piyasa Deęeri}}{\text{Öz Sermaye}}$$

Düşük öz sermaye karlılığı ve yüksek PD/DD oranına sahip senetler pahalı olarak kabul edilirken, yüksek öz sermaye karlılığına ve düşük PD/DD oranına sahip senetler ucuz olarak kabul edilmektedir.³²²

3.2 İřletme Performansını Etkileyebilecek Faktörler

3.2.1 Finansal Oranlar

Bir iřletmenin mali durumu deęerlendirilirken iřletmenin mali tablolarında görünen deęerlerden ziyade bilanço ve gelir tablosunda yer alan kalemler arasındaki ilişkiler daha anlamlıdır. Bu nedenle finansal analizlerde oranlardan geniş ölçüde yararlanılmaktadır. Oranlar incelenirken piyasadaki aynı sektörde yer alan dięer iřletmelerle ve geçmiş yıllarla karşılařtırmalı olarak yorumlanmalıdır.

Finansal oranlar, iřletmelerin yalnız geçmiş ve cari mali durumunu deęerlendirmek açısından deęil, planlama ve denetim iřlevini yerine getirmede ve řirket ortaklarının, pay senetlerine yatırım yapmak isteyen birikim sahiplerinin ve finansman

* Geleneksel yöntemler: řirketlerin mali tabloları kullanılarak elde edilen oranla iřletme performansının deęerlendirilmesi yöntemidir.

³²¹ M. B. Karan, **a.g.e.**, ss. 359–361.

³²² **A.e.**, s. 361.

kurumlarının, işletmenin mali durumunu ve gücünü nasıl gördüklerini, değerlendirdiklerini anlamak için de finans yöneticisine yarar sağlamaktadır.³²³

Farklı şekillerde gruplandırılabilen finansal oranlar bu çalışmada işletme faaliyetlerinin değerlendirilmesindeki kullanım biçimine göre ele alınmıştır ve Likidite Oranları, Finansal Yapı Oranları, Faaliyet Oranları ve Karlılık Oranlarını kapsamaktadır.³²⁴ Karlılık Oranları işletme performans göstergeleri başlığı altında ele alınmıştır.

3.2.1.1 Likidite Oranları

Likidite oranları, işletmenin kısa süreli borçlarını ödeme gücünü ölçmek, işletme sermayesinin yeterli olup olmadığını saptayabilmek için kullanılmaktadır. Kısa sürede borçlarını ödeyebilecek yeterli likiditeye sahip olmanın, özellikle ekonomik bunalım dönemlerinde şirketin faaliyetini sürdürmesi açısından büyük önemi vardır.³²⁵ Literatürde en çok kullanılan ve tartışılan likidite oranlarının kullanıldığı bu çalışmadaki oranlar şu şekildedir;

Cari Oran:

Bu oran işletmenin kısa vadeli borç ödeme gücü hakkında bilgi vermektedir.³²⁶

$$Cari\ Oran = \frac{Dönen\ Varlıklar}{Kısa\ vadeli\ Yükümlülükler}$$

Cari oran hesaplanmasındaki amaç, işletmenin kısa vadeli borçlarını ödeme gücünü ölçmek ve net işletme sermayesinin yeterli olup olmadığını saptayabilmektir. Net işletme sermayesi, dönen varlıklar ile kısa vadeli yükümlülükler arasındaki mutlak farktır; başka bir ifadeyle işletmenin tüm kısa vadeli borçlarını ödedikten sonra kalan işletme sermayesi tutarıdır.³²⁷

³²³ Ö. Akgüç, **Finansal Yönetim, a.g.e., s. 20.**

³²⁴ M. Tükenmez, vd., **a.g.e., 384.**

³²⁵ Ö. Akgüç, **Finansal Yönetim, a.g.e., s. 23.**

³²⁶ **Sermaye Piyasası ve Borsa Temel Bilgiler Kılavuzu, s.139.**

³²⁷ Ö. Akgüç, **Finansal Yönetim, a.g.e., s. 24.**

Asit-test Oranı (Likidite Oranı):

$$\text{Asit - test Oranı} = \frac{(\text{Dönen Varlıklar} - \text{Stoklar})}{\text{Kısa vadeli Yükümlülükler}}$$

eşitliği ile hesaplanan bu oran, kısa vadede nakde dönüşebilecek dönen varlıkların, kısa vadeli borçları karşılama gücünü göstermektedir.³²⁸ Likidite oranının payında paraya çevrilmesi zaman alabilecek olan değerlere yer verilmediğinden, bu oran cari orana kıyasla daha duyarlı bir ölçü olarak kabul edilmektedir.³²⁹

Batılı finans kurumlarında bu oranın 2 olması genel kabul gördüğü halde, yüksek enflasyonlu ve kıt fon kaynaklarına sahip ülkelerde endüstri ve sektörlerin değişik özellikleri göz önünde bulundurulduğunda cari oranın 1,5 olması genellikle yeterli olarak kabul edilmektedir.³³⁰

Çalışmada kullanılan veriler incelendiğinde veri setini oluşturan işletmelerden yüksek likidite oranına sahip sektörlerin taş, toprak ve gıda sektörü olduğu gözlenmiştir. Bu oranın yüksek olması işletmelerin faaliyet alanları ile doğrudan ilişkilidir.

Nakit Oranı:

Nakit oran, işletmenin satışlarının durması ve alacaklarını tahsil edememesi durumunda, kısa süreli borçlarını karşılayabilme gücünü göstermektedir. Bu oran hazır değer oranı olarak da bilinmektedir.³³¹

$$\text{Nakit Oranı} = \frac{(\text{Nakit ve nakit benzerleri} + \text{Finansal yatırımlar})}{\text{Kısa vadeli Yükümlülükler}}$$

Stokların Toplam Varlıklara Oranı:

$$\text{Stokların/Toplam Varlıklara Oranı} = \frac{\text{Stoklar}}{\text{Toplam Varlıklar}}$$

³²⁸ Sermaye Piyasası ve Borsa Temel Bilgiler Kılavuzu, s. 139.

³²⁹ Ö. Akgüç, **Finansal Yönetim, a.g.e.**, s. 28.

³³⁰ (Çevirimiçi) <http://www.bilgaz.net/dosyalar/OranAnalizi.pdf>, 10 Ağustos 2011.

³³¹ Ö. Akgüç, **Finansal Yönetim, a.g.e.**, s. 30.

Toplam varlıkların içinde stok oranını gösteren bu oran, stokların varlık yapısındaki durumunu saptamak için kullanılmaktadır. Bu oran işletmelerin faaliyetlerine ilişkin bilgi sağlamaktadır. Bu oranın yüksek olması işletmelerin yüksek stok hacmi ile çalıştığı ve stoklarını eritme konusunda sıkıntı yaşadığı şeklinde yorumlanabilmektedir.

Kısa vadeli (K.Vd.) Ticari Alacakların Toplam varlıklara Oranı:

$$K.Vd.Tic.Alacakların/Toplam Varlıklara Oranı = \frac{K.Vd.Ticari Alacaklar}{Toplam Varlıklar}$$

Bu oranın yüksek olması işletmelerin alacak politikalarına ilişkin risk taşıdıklarını göstermektedir.

3.2.1.2 Finansal Yapı Analizinde Kullanılan Oranlar

İşletmenin varlıklarını hangi kaynaklardan elde ettiğini, kaynakların yapısı ile dağılımını ve uzun vadedeki borçlarını ödeme gücünü analiz eden oranlar bu başlık altında toplanmaktadır. İşletmenin dönen ve duran varlıklarının yabancı kaynak veya özkaynaklar aracılığı ile finanse edilmesindeki denge; dolayısıyla özkaynak, yabancı kaynak dengesinin sağlanması, yabancı kaynak finansmanındaki anapara ve faiz ödemelerinde işletmenin güçlü bulunması anlamına gelmektedir.³³²

Finansal Kaldıraç Oranı:

$$Finansal Kaldıraç Oranı = \frac{(Kısa vadeli Yük. + Uzun vadeli Yük.)}{Toplam Varlıklar}$$

İşletmenin kaynakları arasında işletmeye sabit yük getiren kaynakların varlığı, finansal kaldıraçın nedenini oluşturmaktadır.³³³ Finansal kaldıraç oranının yüksek çıkması öz sermayenin yetersiz olduğu anlamına geldiği için kredi verenlerin

³³² M. Tükenmez, vd., **a.g.e.**, s. 393.

³³³ Ö. Akgüç, **Finansal Yönetim, a.g.e.**, s. 32.

alacaklarının tahsili riskli duruma gelebilmektedir. Yani işletmenin borçlarını ödeyememe riskinin yüksek olduğu ortaya çıkmaktadır.³³⁴

Kısa Vadeli Yabancı Kaynakların (K.Vd. Yük.) Toplam varlıklara Oranı:

$$K.Vd.Yük./Toplam Varlıklara Oranı = \frac{Kısa vadeli Yük.}{Toplam Varlıklar}$$

Eşitliği ile hesaplanan bu oran, işletmenin iktisadi varlıklarının yüzde kaçının kısa süreli yabancı kaynaklarla finanse edildiğini ortaya koymaktadır. Bu oranın yüksek olması, iktisadi varlıkların büyük bir bölümünün kısa süreli borçlarla finanse edildiğini göstermektedir. Bir işletme gelecekte ödeme güçlükleri ile karşılaşmak istemiyorsa, kısa vadeli borçlarla duran varlıklarını sürekli olarak finanse etmekten kaçınmalıdır.³³⁵

Uzun Vadeli Yabancı Kaynakların (U.Vd. Yük.) Devamlı Sermayeye Oranı:

$$U.Vd.Yük./Devamlı Sermayeye Oranı = \frac{Uzun Vadeli Yük.}{Uzun Vadeli Yük. + Öz Sermaye}$$

Bu oranın yüksek (%25'den fazla) olması yabancı kaynakların oranının yüksek olduğuna ya da işletmelerde büyük oranda uzun vadeli yabancı kaynak kullanıldığına işaret etmektedir.³³⁶ Uzun süreli yabancı kaynaklarla varlıklarını finanse eden işletmeler, özellikle durgunluk dönemlerinde borç taksitlerini ödemedi büyük güçlüklerle karşılaşabilirler.³³⁷

Duran Varlıkların Öz Sermayeye Oranı:

$$Duran Varlıkların/Öz Sermayeye Oranı = \frac{Duran Varlıklar}{Öz Sermaye}$$

³³⁴ M. Tükenmez, vd., **a.g.e.**, s. 394.

³³⁵ Ö. Akgüç, **Mali Tablolar Analizi**, Avcıol Basım-Yayın, İstanbul, Genişletilmiş 9. Bası, 1995, ss. 363–364.

³³⁶ N. Kurnaz, (Çevirimiçi) www.niyazikurnaz.net, 12 Temmuz 2011.

³³⁷ Ö. Akgüç, **Mali Tablolar Analizi**, s. 364.

Duran varlıkların öz sermayeye bölünmesiyle hesaplanan bu oranın düşük olması durumunda maddi duran varlıkların öz kaynaklarla karşılanabildiği kabul edilmektedir.³³⁸

Dönen Varlıkların Toplam Varlıklara Oranı:

$$\text{Dönen Varlıkların/Toplam Varlıklara Oranı} = \frac{\text{Dönen Varlıklar}}{\text{Toplam Varlıklar}}$$

Dönen varlıkların toplam varlıklara bölünmesiyle hesaplanan bu oranın yüksek olmaması (%50'nin altında), kaynakların çoğunluğunun duran varlıklara bağlandığı anlamına gelmektedir. Bu durumda işletmenin kaynak sıkıntısına düşmesi muhtemeldir.³³⁹

3.2.1.3 Faaliyet Oranları

Bu oranlar, işletmenin faaliyetlerinde kullandığı varlıklarını etkin bir şekilde kullanıp kullanmadığını ölçmektedir. Bu çalışmada literatürde incelenen ve hesaplanması mümkün bütün faaliyet oranları ele alınmıştır. Çalışmada kullanılan faaliyet oranları şu şekildedir;

Hazır Değerler Devir Hızı:

$$\text{Hazır Değerler Devir Hızı} = \frac{\text{Satış Gelirleri}}{\text{Hazır Değerler}}$$

Bu oranda Hazır Değer, (*Nakit ve Nakit Ben. + Finansal Yatırımlar*) toplamına eşittir ve hazır değerlerin kaç katı satış yapıldığını göstermektedir.

Hazır Değer Devir Hızı işletmede bulundurulması gerekli nakit miktarını belirlemektedir, ancak bu konuda belirli bir standart yoktur. Bu nedenle aynı sektördeki işletmelerle ve geçmiş yıllarla karşılaştırılarak bir yoruma ve sonuca gidilmeye çalışılması gerekmektedir.³⁴⁰

³³⁸ N. Kurnaz, **a.g.e.**, 12 Temmuz 2011.

³³⁹ **A.e.**, N. Kurnaz.

³⁴⁰ (Çevirimiçi) <http://www.bacegrup.com/3-4-oran-analizi.ntml>, 10 Temmuz 2011.

Stok Devir Hızı:

$$\text{Stok Devir Hızı} = \frac{\text{Satışların Maliyeti}}{\text{Ortalama Stok}}$$

eşitliği ile ifade edilen bu orandaki Ortalama Stok;

$$\text{Ortalama Stok} = (\text{bir dönem önceki Stoklar} + \text{o döneme ait Stoklar})/2$$

olarak hesaplanmaktadır. İşletmeler düşük fiyat seviyesinde satış hacimlerini arttırarak stok devir hızını yüksek tutabilmektedirler. Bu durum rekabet avantajını arttırmaktadır.³⁴¹ Yüksek devir hızı işletmelerin iyi bir stok kontrol sistemi ve etkin bir pazarlama stratejisi olduğunun göstergesi olarak değerlendirilebilir.

İşletmelerin stok devir hızlarının yüksek olması, stokların daha optimal düzeyde tutulduğunu ve kullanıldığını gösterir. Böyle bir durumda işletmeler daha az işletme sermayesi ile daha fazla kar elde etme olanağına sahiptirler; fakat işletmelerin stok devir hızlarının yüksek olması, yetersiz stoklarla çalışmasından da ileri geliyor olabilir.³⁴²

Stok devir hızı stokların verimli kullanımı ve satışlar yoluyla stokların alacaklara dönüşüm hızı hakkında bilgi vermektedir. Stok devir hızının düşmesi nakit akışlarının yavaşlamasına neden olduğundan işletme sermayesi ihtiyacını arttırmaktadır. Stok devir hızının artması nakit akışını arttırdığından ve stok maliyetini azalttığından işletme karlılığı ve verimliliğini arttırmaktadır.

Bir işletme üretimini (faaliyetini) sürdürebilmek amacıyla gerekli üretim girdilerini (hammadde), üretim sürecinde işlemleri tamamlanmamış mamulleri (yarı mamul), üretimini tamamlayıp satış için hazır tuttuğu ürünleri (mamul) stok olarak tutar. Stok devir hızı analizinde amaç stok olarak tutulan bu varlıkların işletme tarafından ne kadar hızla üretim içinde tüketildiği ve satışa hazır hale getirildiği görebilmektir. Bu şekilde stokların belli bir dönem içinde kaç kere yenilendiği ortaya çıkar.³⁴³

³⁴¹ N. Kurnaz, **a.g.e.**, 12 Temmuz 2011.

³⁴² (Çevirimiçi) <http://www.bilgaz.net/dosyalar/OranAnalizi.pdf>, 10 Ağustos 2011.

³⁴³ **A.e.**, 10 Ağustos 2011.

Stok devir hızının düşmesi nakit akışlarının yavaşlamasına neden olduğundan işletme sermayesi ihtiyacını arttırmaktadır. Stok devir hızının artması nakit akışını arttırdığından ve stok maliyetini azalttığından işletme karlılığı ve verimliliğini arttırmaktadır. Stok devir hızının düşük seviyede olması işletmenin satışlarında sorunlar yaşadığı ya da uygun stok politikaları kullanmadığı anlamına gelmektedir.

Alacak Devir Hızı:

Alacakların devir hızı bir işletmenin alacaklarını tahsil edebilirliğini yani likiditesini gösteren iyi bir ölçektir.³⁴⁴

$$\text{Alacak Devir Hızı} = \frac{\text{Satış Gelirleri}}{\text{Ortalama Ticari Alacaklar}}$$

Yukarıdaki eşitlikte paydada yer alan Ortalama Ticari Alacaklar

$$= (\text{bir dönem önceki Ticari Alacaklar} + \text{o döneme ait Ticari Alacaklar})/2$$

olarak hesaplanmaktadır. Alacak devir hızının artması, işletmenin alacaklarını vadesinde tahsil edebildiğini ve vadelerinin kısaldığını ifade etmektedir.³⁴⁵ Bu oranının yüksek olması alacakların vadeli alacaklar için tahsilat süresinin kısa olduğu anlamına gelmektedir. Yıllar itibariyle azalan alacak devir hızı işletmelerin nakit seviyesini de yükselmesine neden olmaktadır. Alacak devir hızının yüksek olduğu en önemli sektörlerden biri perakende sektörüdür.

Bir işletme alacaklarını hızlı bir şekilde (devir hızının yüksek olması durumu) tahsil edebiliyorsa, likiditesi yüksek kabul edilebilir ve işletme bu sayede hem nakit sıkıntısı içine girmez, hem de alacaklarının değeri fazla erimeden bunları daha iktisadi alanlarda kullanabilir.³⁴⁶

Dönen Varlık Devir Hızı:

$$\text{Dönen Varlık Devir Hızı} = \frac{\text{Satış Gelirleri}}{\text{Dönen Varlıklar}}$$

³⁴⁴ Ö. Akgüç, **Mali Tablolar Analizi**, s. 380.

³⁴⁵ Ö. Akgüç, **Finansal Yönetim**, s.44.

³⁴⁶ (Çevirimiçi) <http://www.bilgaz.net/dosyalar/OranAnalizi.pdf>

eşitliğiyle hesaplanan bu oran işletme sermayesi olarak da adlandırılmaktadır. Kısa vadeli borçlarla öz sermaye arasında denge olup olmadığı bu oranla ölçülmektedir. Kısa süreli borçlar uzun süreli borçlardan daha büyük boyutlarda dalgalanma gösterdiği gibi daha kısa aralıklarla yenilenmesi veya ödenmesi gerekmektedir. Bu nedenle bu oranın yüksekliği fonlama açısından daha büyük risk göstergesidir.³⁴⁷

Duran Varlık Devir Hızı:

$$\text{Duran Varlık Devir Hızı} = \frac{\text{Satış Gelirleri}}{\text{Duran Varlıklar}}$$

Bu oranın düşük çıkması duran varlıkların fazlalığına işaret ederken, kaynakların verimsiz kullanıldığının belirtisi olarak da görülmektedir. Oranın yüksek çıkması ise duran değerlere aşırı bir yatırım yapılmadığının ve kaynakların verimli kullanılmadığının göstergesi olarak kabul edilmektedir.³⁴⁸

Öz Sermaye Devir Hızı:

Öz sermayenin verimliliğini ölçen oran

$$\text{Öz Sermaye Devir Hızı} = \frac{\text{Satış Gelirleri}}{\text{Öz Sermaye}}$$

olarak hesaplanmaktadır. Oranın yüksek çıkması, öz sermayenin verimli olarak kullanıldığı, düşük çıkması da atıl öz sermaye bulunduğu anlamına gelmektedir. Oranın normalin çok üstünde çıkması ise net satışların, öz sermaye yanında yabancı kaynak kullanımı ile sağlandığının bir göstergesi olarak kabul edilmektedir. Oranın belli bir standardı bulunmamaktadır.³⁴⁹

Toplam Varlık Devir Hızı:

$$\text{Toplam Varlık Devir Hızı} = \frac{\text{Satış Gelirleri}}{\text{Toplam Varlıklar}}$$

³⁴⁷ Ö. Akgüç, **Mali Tablolar Analizi**, s. 386.

³⁴⁸ N. Kurnaz, (Çevirimiçi) www.niyazikurnaz.net, 12 Temmuz 2011.

³⁴⁹ M. Tükenmez, vd., **a.g.e.**, s. 413.

Bu oran bir işletmede teknoloji kullanımının bir göstergesi, varlık kullanımında bir etkinlik ölçüsü ya da işletmenin risk derecesinin göstergesi olarak yorumlanabilmektedir. Varlık devir hızının yüksek olduğu endüstri kollarında risk göreceli olarak daha düşük olmaktadır.³⁵⁰ Bunun yanı sıra düşük oran, işletmenin aktiflerinin atıl kapasiteyle çalıştığı bir göstergesi olarak kabul edilmektedir. Oranın düşük olması kredi verenler açısından işletme varlıkları içinde duran varlıkların büyük bir ağırlığa sahip olduğu anlamını taşımaktadır. Bununla birlikte duran varlıklarda sürekli gelir elde edilmesi oranın yükselmesine neden olacaktır.³⁵¹

Varlık devir hızı bir işletmenin karlılığını belirleyen önemli etmenlerden biridir. Bir işletme diğer koşullar aynı kalmak üzere varlık devir hızını arttırmak yoluyla öz sermaye karlılığını yükseltebilmektedir.³⁵²

3.2.2 Finansal Olmayan Değişkenler

Sermaye Yoğunluğu:

Sermaye yoğunluğu, emek üretkenliğini belirleyen önemli bir göstergedir. Sermaye yoğunluğu değerlendirilirken en çok kullanılan oranlar çalışan sayısına göre duran varlık, toplam varlık ve makine teçhizatla elde edilen değerlerdir.³⁵³ İşletmelerin makine-teçhizatla ilgili verilerine ulaşmadaki zorluklar nedeniyle bu çalışmada aşağıdaki iki oran kullanılmıştır.

$$Sermaye\ Yoğunluğu_1 = \frac{Duran\ Varlık}{Çalışan\ Sayısı}$$

$$Sermaye\ Yoğunluğu_2 = \frac{Toplam\ Varlık}{Çalışan\ Sayısı}$$

İktisatta sermayenin emeğe oranı sermaye yoğunluğu olarak ifade edilmektedir. Ekonomik büyümenin temel belirleyicileri; sermaye birikimi, teknolojik gelişme ve

³⁵⁰ Ö. Akgüç, **Finansal Yönetim**, ss.57–58.

³⁵¹ **A.e.**, s.57.

³⁵² **A.e.**, s. 58.

³⁵³ Z. Akal, **İmalatçı Kamu Kuruluşlarında İşletmelerarası Toplam Performans, Verimlilik, Karlılık ve Maliyet Karşılaştırması**, Ankara, MPM: 538, 1994, s. 53.

istihdam artışıdır. İmalat sanayinde 1980’li yıllardan sonra üretimin sürecinde gerileme yaşanmıştır. Sermaye birikimi oranındaki ciddi yavaşlama bu oluşumun esas sebebinin oluşturmaktadır. 2000–2003 yıllarında sermaye stoku yeniden ilerleme kaydetmiş ve 1985’den bu yana düşmekte olan performans yeniden pozitif bir konuma geçmiştir. Son yıllarda yaşanan emek verimliliğinin artışında, özellikle özel sektör yatırımlarında kendini gösteren sermaye birikimi gelişiminin payı olabileceği öne sürülmektedir. Ancak, bunun ne ölçüde “sürdürülebilir” olduğu tartışılmaktadır. Bu noktada büyüme stratejilerinin özellikle teknoloji ve verimlilik bağlantılı olması önerilmektedir.³⁵⁴

Yabancı Sermaye Oranı:

İşletmelerin öz sermayesinin yabancı oranını gösteren bu değişken işletmelere doğrudan yapılmış yabancı firmaların ortaklığı ile ilgili bilgi vermektedir.

Hızlı bir büyüme politikası benimseyen Türkiye’nin tasarruf, dış ticaret açığı gibi kalkınmasını engelleyen problemlerden kendisini koruyabilmesi için yabancı sermaye politikası önemli bir ekonomik stratejidir.³⁵⁵

İşletmelerin büyümek ve yeni teknolojileri en hızlı şekilde takip edebilmeleri için ihtiyaç duydukları gerekli sermaye nedeniyle yabancı sermaye yatırımlarının önemi artmıştır. Gelişen ekonomisiyle Türkiye, 2002 yılından itibaren uluslararası sermaye ve yatırımcıların ilgi odağı olmaktadır.

Tablo 3.2 incelendiğinde 2007 yılında toplam 22 milyar dolar değerinde uluslararası sermayenin ülkeye girdiği görülmektedir. Küresel ekonomik krizin etkili olduğu 2009 yılında ise ülkeye giren uluslararası sermaye çok ciddi anlamda düşüş göstermiştir. Türkiye’nin iç kaynaklarının yetersiz ve dış borçlanma maliyetlerinin çok yüksek olması sebebi ile doğrudan yabancı yatırımlar teşvik edilmektedir.

³⁵⁴ H. Suiçmez, Türkiye’de Ekonomik Büyüme ve Verimlilik Artış Performansı Işığında Nasıl bir Kalkınma Politikası Benimsenmeli, **Türkiye İşveren Sendikaları Konfederasyonu**, Ocak, 2007. (Çevirimiçi) http://www.tisk.org.tr/isveren_sayfa.asp?yazi_id=1618&id=84, 11 Haziran 2011

³⁵⁵ Açık Gazete, Yabancı Sermaye Nedir? Hazırlayan Faruk Eskioğlu, (Çevirimiçi) <http://www.acikgazete.com/ozel-dosyalar/2009/03/24/yabanci-sermaye-nedir.htm?aid=27995>, 01 Temmuz 2011

Doğrudan yabancı yatırımların getireceği teknoloji, know-how, modern işletme ve pazarlama yöntemleri işletmelerin büyüme stratejilerinde önemli hale gelmektedir.

Tablo 3.2 Yabancı Sermaye ve Ekonomik Büyüme

Yıllar	Doğrudan Yabancı Sermaye (Fiili giriş) (milyon \$)	Ekonomik Büyüme (%)
2002	1042	6.2
2003	1693	5.3
2004	2779	9.4
2005	10010	8.4
2006	20223	6.9
2007	22023	4.7
2008	18148	0.7
2009	7610	-4.7

Kaynak. Hazine Müsteşarlığı

Yenilik:

Araştırma Geliştirme (Ar-Ge) harcamalarının işletmelerin satış gelirleri içindeki payını gösteren yenilikle ilgili oran işletmelerin yenilik düzeyinin göstergesidir.³⁵⁶

$$Yenilik = \frac{Ar - Ge \text{ Harcamaları}}{Satış \text{ Gelirleri}}$$

Son yıllarda işletmeler rekabet avantajı sağlayabilmek için ürün maliyetlerini düşürmeye çalışmaktadırlar. Bu durum işletmelerin bilgi ve iletişim teknolojilerini kullanmalarını ve takip etmelerini zorunlu kılmaktadır. Uzun dönemde sürekli büyüme sağlayabilmek için işletmeler hatta ülkeler, teknolojik yenilikleri takip ederek verimlilik düzeylerini arttırmak zorundadırlar.

Ar-Ge destek programlarının uygulanmasının en önemli nedenlerinden biri, teknolojik yenilik sonucu üretkenliğin ve rekabet gücünün artırılmasıdır. Yapılan bir

³⁵⁶ Z. Griliches, "Productivity, R&D and Basic Research at the Firm Level in the 1970's", **The American Economic Review**, 1986, pp. 141–153.

anket çalışması sonucunda Türkiye'deki işletmelerin sadece yaklaşık %23'ü Ar-Ge yatırımı yapmaktadır.³⁵⁷

Üretilen yeni teknolojilerle yeni ürünler ve yeni üretim yöntemleri oluşturulmaktadır. Ürün yeniliği, yeni ürünler için piyasalar geliştirmek veya iyileştirilmiş ürünlere olan talebi arttırarak üretim düzeyinde bir artışa yol açabilir. Bu nedenle işletme ve sektör düzeyinde, ürün yeniliklerinin işgücü talebini ve istihdamı arttırma eğiliminde olması, buna paralel olarak da kişilerin ve ülkenin refah düzeyinin sürekli artış göstermesi beklenmektedir.³⁵⁸ Bu nedenle uzun dönemde büyüme ve refah için anahtar kavram teknolojik değişim ve yeniliklerdir.³⁵⁹

İhracat:

İhracat, işletmelerin yurt dışına yapmış oldukları satışları ifade etmektedir. İhracatın işletmelere sağladığı faydalar; alternatif pazar yaratarak pazar payının büyümesine olanak sağlaması ve böylece pazar payının büyümesi ile kapasite kullanım oranını arttırabilmesi, işletmenin iç pazarda rekabet gücünü arttırması ve son olarak da işletmenin iç pazarda oluşabilecek ekonomik durgunluklardan daha az etkilenmesini sağlaması olarak sıralanabilir. İşletmeler ihracat yaparak bir takım riskler de taşımaktadırlar bu risklerden bazıları; işletmenin ihracat yaptığı ülkenin yaşadığı istirsazlıkların ihracat gelirlerinde problemlere neden olabilmesi, döviz kuru dalgalanmalarının olumsuz etkisi ve herhangi bir sorun karşısında ticari ve hukuki anlaşmazlıkların çözümünün zaman kaybına neden olması şeklinde ifade edilebilir.³⁶⁰

“1966'da Vernon tarafından ileri sürülen büyüme dayalı ihracat hipotezine göre, ülkelerin sahip oldukları büyüme oranlarının, ihracatları üzerinde pozitif yönlü bir

³⁵⁷ “Yatırım Ortamı Değerlendirmesi Krizden Özel Sektör Öncülüğünde Büyüme”. (Çevirimiçi) <http://www.hazine.gov.tr>, 07 Temmuz 2011.

³⁵⁸ Bölüm-10, “Teknolojik Yenilik ve Ekonomik Performans”, (Çevirimiçi) <http://www.inovasyon.org/pdf/blm10.pdf>, 01 Temmuz 2011.

³⁵⁹ H. Suiçmez, **a.g.e.**

³⁶⁰ Bölüm-1, “Niçin İhracat”, **100 Soruda Dış Ticaret, İhracat Geliştirme Etüd Merkezi**, Ekim, 2006, (Çevirimiçi) http://www.dkib.org.tr/dosya/pratik_100soru.pdf, 10 Temmuz 2011.

etkiye yol açtığı ve dolayısıyla ülkenin ihracatında önemli ölçüde bir genişlemeye neden olduğu savunulmaktadır.”³⁶¹

Ekonomik büyüme ve ihracat arasındaki bu karşılıklı ilişki nedeniyle Türkiye’de özellikle büyümeye önem veren işletmeler ihracata yönelmişlerdir.

Büyüyen işletmelerin Ar-Ge harcamalarını özellikle son yıllarda arttırarak daha yeni teknolojiye dayalı ve kaliteli ürünler ürettikleri de görülmektedir.³⁶²

Pazar Payı:

Pazar payı, sektördeki toplam satışlar içerisindeki işletme satışlarının payını ifade etmektedir. Bu gösterge işletmenin faaliyet sektörü içindeki yerini ve rekabet durumunu belirlemede, başka bir ifadeyle işletme faaliyetinin başarısını ortaya çıkartmaktadır. Ayrıca bu oran işletmenin ürün çeşitliğinin rakibine kıyasla genişliğini ölçmektedir.³⁶³

$$\text{Pazar payı} = \frac{\text{İşletme Satışları}}{\text{Sektördeki Satışlar}}$$

PIMS (Profit Impact of Market Strategies_Pazarlama Stratejilerinin Karlılığa Etkisi) proje araştırması* sonuçları aktif karlılığa etki eden 37 faktörü analiz etmiştir. Bu faktörler içerisinde yer alan pazar payına ilişkin PIMS’in elde ettiği sonuçlar kısaca şu şekilde özetlenebilir;³⁶⁴

- Pazar payının karlılık üzerinde gerçekten büyük etkisi vardır. Pazar payı artarken karlılık süratle yükselmektedir.
- Büyük pazar payına sahip işletmelerin ortalama sermaye devir hızları yüksek ve pazarlama masraflarının satışlara oranı düşüktür. Bu nedenle pazarlama

³⁶¹ C. Bilgin, A. Şahbaz, **a.g.e.**, ss. 177–198.

³⁶² “Yatırım Ortamı Değerlendirmesi Krizden Özel Sektör Öncülüğünde Büyümeye”, **a.g.e.**, 07 Temmuz 2011.

³⁶³ Ö. Dinçer, **Stratejik Yönetim ve İşletme Politikası**, İstanbul, 1992, s. 116.

* Stratejik planlama ile işletmenin karlılığı arasındaki ilişkiyi araştırmak üzere 620 değişik konuda faaliyet gösteren 57 ana işletmede yapılan bir araştırmadır. Projenin araştırma sorumluluğunu Harvard İşletme Fakültesi Profesörlerinden Robert D. Buzzel yapmıştır.

³⁶⁴ (Çevirimiçi) <http://notoku.com/pims-analizi/> , 12 Temmuz 2011.

masraflarının satılan ürün birimi başına payı düşmekte ve bu durum karlılığı arttırmaktadır.

- Yüksek pazar payına sahip işletmelerin ürünlerinin kalitesi de yüksektir.
- Pazar payı zayıf olan işletmelerde Ar-Ge masraflarının satışlara oranı arttığı takdirde karlılık %50'den fazla azalmaktadır. Pazar payı yüksek olan işletmelerde ise karlılık oranı %18 artmaktadır.

Büyüklik:

İşletmenin büyüklüğünü belirlemede kantitatif ve kalitatif olmak üzere çeşitli kriterlerden yararlanılmaktadır. Kantitatif büyüklük ölçüleri olarak; yıllık satışlar, yıllık karlar, varlıklar, öz sermaye, yatırım toplamı, yaratılan katma değer, çalışanların sayısı, harcanan enerji miktarı ve Ar-Ge için harcanan para miktarı kullanılabilir. Kalitatif büyüklük ölçüleri ise; sermayedarların sayısı, yönetim biçim, bölgeye yönelik olup- olmama, endüstri dalındaki nisbi durum ve hukuki şekil olarak sıralanabilir.³⁶⁵

³⁶⁵ İ. Mucuk, **Modern İşletmecilik**, 7. Baskı, Türkmen Kitapevi, İstanbul, 1997, ss. 98–99.

DÖRDÜNCÜ BÖLÜM: İŞLETME PERFORMANSININ TAHMİNİNE YÖNELİK İMALAT SANAYİ SEKTÖRÜ UYGULAMASI

4.1 Araştırmanın Amacı ve Konusu

İşletmelerin performanslarının sürekliliğinin sağlanmasının önemi özellikle yatırımcılar, ortaklar ve işletme yöneticileri açısından her geçen gün artmaktadır. İşletmelerin performanslarını birden çok ve farklı kriter ile ölçmeleri ve değerlendirmeleri mümkündür. Hem işletmelerin yatırımcıları ve ortaklarının kararlarında, hem de ülke ekonomilerinin verimliliğinin ve etkinliğinin belirlenmesinde işletmelerin performanslarının değerlendirilmesi için işletme performanslarını etkileyen faktörlerin en iyi şekilde belirlenmesinin önemi günümüze kadar birçok araştırmacının dikkatini çekmektedir.

Bu çalışmada İstanbul Menkul Kıymetler Borsası (İMKB) imalat ve sanayi endeksinde bulunan işletmelerin, finansal ve finansal olmayan açıklayıcı değişkenlerinin işletmelerin geleneksel performans göstergeleri (karlılık, verimlilik, büyüme ve piyasa) üzerinde etkisinin araştırılması hedeflenmiştir. Günümüze kadar yapılan çalışmalar incelendiğinde performans göstergelerinin tek tek ele alındığı ve performansı etkileyen faktörlerin kurulan modeller için ayrı ayrı belirlendiği gözlenmiştir. Bu çalışmada ise farklı olarak, performans göstergelerinin tüm mümkün kombinasyonları denenmiş ve anlamlı olan modeller araştırmanın kapsamında incelenmiştir. İşletme performans göstergelerinin birarada ele alınması hedeflendiği için kategorik olarak tanımlanması gerekmektedir.

4.2 Araştırmanın Veri Seti

Araştırmada kullanılan veri seti 2007 yılında İMKB’de İmalat Sanayi Sektöründe faaliyet gösteren işletmelerden oluşmaktadır. 2008 yılında gerçekleşen ve Türkiye’yi de etkileyen global kriz nedeniyle işletmelerin finansal ve finansal olmayan göstergelerden özellikle “Stok Devir Hızı” ve “Alacak Devir Hızı” ile Tablo 3.2’deki

“Yabancı Sermaye ve Ekonomik Büyüme” değerleri 2008 ve 2009 yıllarında, yaşanan krizin etkilerini taşımaktadır. Ayrıca çalışmada kullanılacak bazı finansal olmayan oranlar için İstanbul Sanayi Odası’nın (İSO) her sene Eylül ayında yayımladığı özel sayıdan yararlanılması nedeniyle 2010 yılı verileri incelenememiştir. Bu nedenlerden dolayı çalışmada, işletme performans kriterlerini etkileyebilecek uygun faktörlerin saptanması amacıyla ekonomik istikrarın olduğu 2007 yılı verilerinin kullanılması uygun görülmüştür.

İşletmelere ait finansal değişkenlerin ve piyasa performans oranlarının hesaplanmasında İMKB’nin internet üzerinden yayınladığı yıllık raporlar, mali tablolar ve bu tablolara ait dipnotlardan yararlanılmıştır. Ayrıca finansal olmayan oranların belirlenmesinde İSO’nun 2007 yılında yayımladığı özel sayıdan elde edilen değerler kullanılmıştır.

Çalışmada performans göstergelerinin olası tüm kombinasyonları için incelenen işletme sayıları farklılık göstermektedir. İşletmelerin sayısına ilişkin bilgiler “Verilerin Analize Hazırlanması” başlığında (4.4) ayrıntılı olarak belirtilmiştir. Araştırma verilerinin analizinde SPSS’in son sürümlerinden olan PASW Statistics 18 ve STATA 9.2 paket programları kullanılmıştır.

4.3 Araştırma Değişkenlerinin Seçimi

Çalışmanın amacı olan işletme performansı (bağımlı değişken) belirlenirken karlılık, büyüme, piyasa ve verimlilik, verilerinin her biri birden çok değişkenle ifade edilmiştir. Bu değişkenlerden performans kriterlerini en iyi temsil edenleri saptanarak, kullanılacak oranları belirlemek amacıyla tek tek karlılık, büyüme ve piyasa performans göstergeleri için oranlar arasındaki ilişkiyi ortaya çıkarmaya yönelik bir ön analiz yapılmıştır. Bu amaçla önce “Veri setindeki değişkenlerin sıraları arasında uygunluk yoktur” sıfır hipotezi ile kurulan Kendall Uygunluk Katsayısı Testi kullanılmış ve test değerlerine göre her üç performans göstergesi değişkenlerinin de sıraları arasında uygunluk olduğu görülmüştür (H_0 red). Sıfır hipotezinin rededilmesinden sonra gerçek değerler yerine değişkenlerin sıralarının

kullanıldığı Spearman Korelasyon değerleri ile değişkenler arasındaki ilişkinin gücü ortaya konmuştur.

Karlılık, büyüme ve piyasa kriterleri için uygulanan Kendall Uygunluk Katsayısı Testi ve Spearman Sıra Korelasyon ön analiz sonuçları sırası ile EK.1, EK.2 ve EK.3’de verilmiştir.

Ön analiz sonuçlarına göre karlılık kriteri açısından Başlık 3.1.1’de tanımlanan tüm karlılık oranlarının birbirleriyle olan ilişkisi yüksek çıkmıştır. Ancak yapılan çalışmalarda çoğunlukla tercih edilmesi sebebiyle bu çalışmada da “Aktif Karlılık” oranı seçilmiştir.

Büyüme kriteri açısından, Başlık 3.1.3’te değinilen büyüme oranlarının birbirleriyle ilişkilerinin çok güçlü olmadığı görülmüştür. Bu nedenle, kullanılacak değişkenin belirlenmesi için işletmelerin sınıflamada diğer performans kriterleri ile birlikte değişkenlerin hepsi tek tek gruplanmış ve işletme sayısı baz alındığında en iyi sınıflandırmayı veren büyüme performans göstergesinin “Satışlardaki Reel Artış” ve “Varlık Artış” oranları olduğuna karar verilmiştir.

Piyasa performans kriteri açısından Başlık 3.1.4’te tanımlanan değişkenlerin birbirleri ile olan ilişkilerinin yüksek olmadığı görülmüştür. Bu durumda, büyüme kriterinde olduğu gibi tüm mümkün piyasa performans göstergeleri için karşılaştırma yapılmış ve en uygun gruplandırma sonucuna göre bağımsız değişkenin belirlenmesinde “Fiyat-Kazanç” oranının kullanılmasına karar verilmiştir. Yatırımcıların şirket analizinde performans göstergesi olarak en çok başvurduğu değişkenin Fiyat-Kazanç oranı olması da bu değişkenin tercih edilmesinde etkili olmuştur.

Bir diğer performans kriter değeri olan verimlilik için ise; karşılaştırılabilirlik, yorumlanabilirlik ve güvenilirlik hususlarında en az sorun işgücü verilerinde olduğundan, özellikle imalat sanayi sektöründe en yaygın kullanılan göstergenin literatürde “Emek Verimliliği” olduğu gözlemlenmiştir. Bu nedenle çalışmada da emek verimliliği oranı seçilmiştir.

Çalışmada kullanılan bağımsız değişkenler ise finansal oranlar (Tablo 4.1) ve finansal olmayan oranlar (Tablo 4.2) olmak üzere grupta toplanmıştır.

Tablo 4.1 Bağımsız Değişken Olarak Kullanılan Finansal Oranlar

Oran	Formüller	Modeldeki Karşılığı
Likidite Oranları		
Cari Oran	Dönen Varlık / Kısa vadeli Yükümlülükler	Likidite1
Asit-Test Oranı	(Dönen Varlık-Stoklar) / Kısa vadeli Yükümlülükler	Likidite2
Nakit Oranı	(Nakit ve Nakit Benzerleri+Finansal Yatırımlar) / Kısa Vadeli Yükümlülükler	Likidite3
Stokların Toplam Varlıklara oranı	Stoklar / Toplam Varlıklar	Likidite4
Kısa vadeli Ticari Alacakların Toplam Varlıklara Oranı	Kısa vadeli Ticari Alacaklar / Toplam Varlıklar	Likidite5
Finansal Yapı Oranları		
Finansal Kaldıraç Oranı	(Kısa Vadeli Yük.+ Uzun Vadeli Yük.) / Toplam Varlıklar	FinYap1
Kısa Vadeli Yabancı Kaynakların Toplam Varlıklara Oranı	Kısa Vadeli Yük. / Toplam Varlıklar	FinYap2
Uzun Vadeli Yabancı Kaynakların Devamlı Sermayeye Oranı	Uzun Vadeli Yük. / (Uzun Vadeli Yük. + Öz Sermaye)	FinYap3
Duran Varlıkların Öz Sermayeye Oranı	Duran Varlıklar / Öz Sermaye	FinYap4
Dönen Varlıkların Toplam Varlıklara Oranı	Dönen Varlıklar / Toplam Varlıklar	FinYap5
Faaliyet Oranları		
Hazır Değerler Devir Hızı	Satış Gelirleri / Hazır Değerler	Faaliyet1
Stok Devir Hızı	Satışların Maliyeti / Ortalama Stok	Faaliyet2
Alacak Devir Hızı	Satışların Gelirleri / Ortalama Ticari Alacaklar	Faaliyet3
Dönen Varlık Devir Hızı	Satış Gelirleri / Dönen Varlıklar	Faaliyet4
Duran Varlık Devir Hızı	Satış Gelirleri / Duran Varlıklar	Faaliyet5
Öz Sermaye Devir Hızı	Satış Gelirleri / Öz Sermaye	Faaliyet6
Toplam Varlık Devir Hızı	Satış Gelirleri / Toplam Varlıklar	Faaliyet7

Kaynak: İMKB

Tablo 4.2 Bağımsız Değişken Olarak Kullanılan Finansal Olmayan Oranlar

Oran	Formülleri	Modeldeki Karşılıkları
Sermaye Yoğunluğu-1	Duran Varlık / Çalışan Sayısı	SermayeYog1
Sermaye Yoğunluğu-2	Toplam Varlık / Çalışan Sayısı	SermayeYog2
Yenilik	Ar-Ge Harcamaları / Satış Gelirleri	Yenilik
Pazar Payı	İşletme Satışları / Sektördeki Satışlar	PazarPay
Yabancı Sermaye Oranı*		YabSermaye
İhracat*		İhracat
Büyüklik*		Buyukluk

Kaynak: İMKB, İSO

4.4 Verilerin Analize Hazırlanması

Araştırmanın veri setinde yer alan işletmelerin her biri belirlenen dört performans kriterlerinin olası tüm dörtlü, üçlü ve ikili kombinasyonları için küçükten büyüğe doğru sıralanmıştır. Bu işlem sonucunda tüm kriterlere göre küçük değer alan işletmeleri “0” büyük değer alan işletmelere ise “1” grubu atayarak iki ayrı sınıf oluşturulmuştur. Kategorize edilemeyen işletmeler ise analiz dışında bırakılmıştır. Bu şekilde oluşturulan kategorik bağımlı değişkeni açıklayabilecek faktörleri saptamak üzere 24 bağımsız değişken, bahsedilen tüm kombinasyonların modellenmesinde kullanılmıştır. Dörtlü ve üçlü performans göstergesi kombinasyonlarından Karlılık-Verimlilik-Büyüme kombinasyonu dışındaki göstergeler için işletmeler sınıflandırılmamıştır. Olası tüm ikili performanslar incelendiğinde ise Verimlilik-Piyasa ve Büyüme-Piyasa göstergelerini açıklayabilecek bağımsız değişkenler elde edilemediği için modelleri kurulamamıştır. Sorun yaşanan kombinasyonlar çıkarıldığında kalan beş

* Finansal olmayan oranlardan yabancı sermaye oranı ile ihracat değişkeni ISO dergisinden doğrudan alınmıştır. Büyüklik değişkeni ise yine aynı dergiden alınan şirketlerin katma değerlerine göre yüksek olanlara “1” düşük olanlara “0” değeri atanarak kategorik veriye dönüştürülmüştür.

kombinasyonla elde edilen modeller çalışmada ayrıntılı olarak incelenmiştir. Oluşturulan modeller ile her bir model için işletme sayıları şu şekildedir:

Model 1: Karlılık-Verimlilik-Büyüme göstergelerine göre (79 işletme için),

Model 2: Karlılık-Verimlilik göstergelerine göre (104 işletme için),

Model 3: Karlılık-Büyüme göstergelerine göre (95 işletme için),

Model 4: Karlılık-Piyasa göstergelerine göre (78 işletme için),

Model 5: Verimlilik-Büyüme göstergelerine göre (88 işletme için) performans göstergelerini etkileyen faktörler belirlenmiştir.

Kurulan modellerde karlılık için aktif karlılığı, verimlilik için emek verimliliği, piyasa için de fiyat-kazanç oranı kullanılmıştır. Büyüme değişkeni olarak Model 1’de varlık artış oranı, Model 3 ve Model 5’te ise satışlardaki reel artış oranı benimsenmiştir.

4.5 Finansal Performans Tahmin Modellerinin Uygulanması

Her bir modelin açıklayıcı değişkenlerinin saptanması için hem Diskriminant Analizi hem de Lojistik Regresyon Analizi uygulanmıştır. Diskriminant Analizi için PASW Statistics 18 paket programı kullanılmıştır. Lojistik Regresyon Analizi için ise “Kesirli Polinomlar”, “Roc Eğrisi” ve modelin oluşturulmasında kullanılan “En İyi Model Seçim” sonuçlarının elde edilebilmesi amacıyla PASW Statistics 18 ile birlikte Stata 9.2 istatistik paket programı da kullanılmıştır.

4.5.1 Model 1: Performans Göstergesi Olarak Karlılık Verimlilik Büyüme

Bu modelde performans göstergesi olarak karlılık, verimlilik ve büyüme için seçilen “aktif karlılığı”, “emek verimliliği” ile “varlık artış” değerlerini birlikte açıklayan faktörlerin saptanması amaçlanmıştır.

Model 1 oluşturulurken, bağımsız değişkenler arasındaki yüksek korelasyon sebebiyle ortaya çıkabilecek çoklu doğrusal bağlantı sorunun önüne geçebilmek amacıyla Adım Adım (Stepwise) Regresyon Yöntemi hem Diskriminant hem de Lojistik Regresyon Analizi'nde uygulanmıştır.

Diskriminant Analizi için çoklu doğrusal bağlantı varsayımı haricinde eşit varyanslılık ve çoklu normallik varsayımlarının testleri yapılarak modellerin grupları tahmin etmedeki başarısı değerlendirilmiştir.

Lojistik Regresyon Analizi için ise Adım Adım Regresyon Yöntemi uygulandıktan sonra bulunan modelden elde edilen lojit değerleri ile modeldeki bağımsız değişkenlerin doğrusallıkları, Kesirli Polinomlar (Fractional Polynomial) yöntemi ile kontrol edildikten sonra, bulunan model ve parametrelerin uygunluk test sonuçları incelenmiş ve klasik yöntem ve ROC eğrisi çizilerek modelin sınıfları tahmin etmedeki başarısı değerlendirilmiştir.

4.5.1.1 Diskriminant Analizi Sonuçları

Model 1 için öncelikle analizin sağlaması gereken eşit varyanslılık, çoklu normallik ve bağımsız değişkenler arasında ilişki olmaması varsayımlarının sağlanıp sağlanmadığının kontrolleri yapılmıştır.

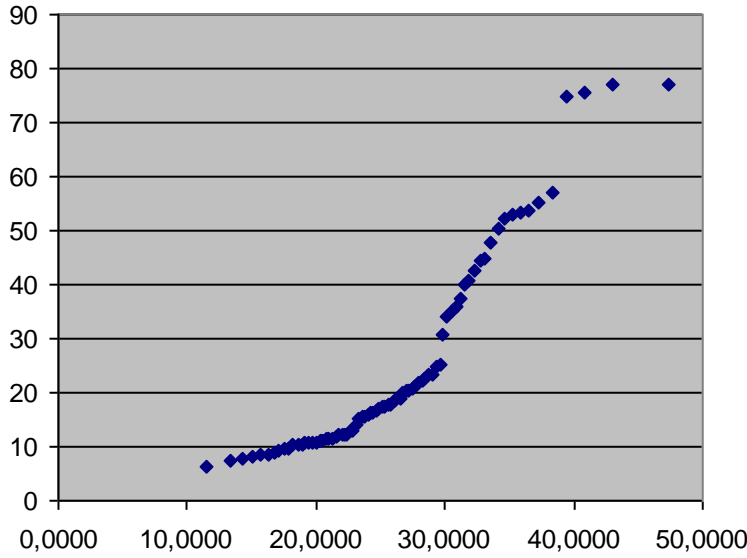
Diskriminant Analizi ile yapılan testler ve sonuçlarının yorumlanmasında kullanılan değerler Tablo 4.3'te özetlenmiştir (Paket program sonuçları için bkz. EK. 4).

Eşit varyanslılık için iki grubun varyans-kovaryans matrislerinin eşit olduğunu ileri süren sıfır hipotezinin Box's M test sonucuna göre reddedildiği görülmektedir (sig. 0,000). Bu durumda eşit varyanslılık varsayımı sağlanmamaktadır.

Tablo 4.3 Model 1 için Diskriminant Analizi Testleri ve Sonuçları

Box's M		Eigenvalues			Wilks' Lambda	
Değeri	Sig.	Değeri	Varyans (%)	Kanonik Korelasyon	Değeri	Sig.
1835,091	0,000	0,874	100	0,683	0,534	0,000

Veri setinden hesaplanan Mahalanobis uzaklığının sıralı değerleri ile bunların χ^2 değerlerinden elde edilen serpilme diyagramı çizildiğinde çoklu normallik varsayımının da sağlanmadığı görülmektedir.



Şekil 4.1 Model 1 için Serpilme Diyagramı

Çoklu doğrusal bağlantı sorunu, analiz uygulanırken Adım Adım Regresyon Yöntemi kullanılarak aşılmaq istenmiş ve Tablo 4.4'ten standartlaştırılmış kanonik diskriminant fonksiyonu şu şekilde elde edilmiştir.

$$T = 0,548 \text{ Likidite5} + 0,634 \text{ FinYap2} + 1,768 \text{ SermayeYog1} \\ - 2,000 \text{ SermayeYog2} - 0,376 \text{ YabSermaye} - 0,530 \text{ Buyukluk}$$

Diskriminant Modeli'ne giren değişkenler incelendiğinde sermaye yoğunluğu için hesaplanan iki oranın (SermayeYog1 ve SermayeYog2) birbiriyle yüksek korelasyona sahip olması nedeniyle (bkz. EK. 4) Adım Adım Regresyon Yönteminin çoklu doğrusal bağlantı problemini çözmede etkin olamadığı görülmüştür.

Tablo 4.4 Model 1 için Diskriminant Modeli ve Doğru Sınıflandırma Sonuçları

Standartlaştırılmamış ve Standartlaştırılmış Kanonik Diskriminant Fonksiyon Katsayıları			Sınıflandırma Sonuçları			
Değişkenler	β	β^*	Sınıf	0	1	Doğru Yüzdesi
Likidite5	4,206	0,548	0	23	9	71,9
FinYap2	3,786	0,634	1	6	41	87,2
SermayeYog1	0,000	1,768	Toplam			81,0
SermayeYog2	0,000	-2,000				
YabSermaye	-0,015	-0,376				
Buyukluk	-1,444	-0,530				
(Sabit)	-0,871					

Tablo 4.4 ile verilen Diskriminant modeli sonuçları, bağımsız değişkenlerin sınıfları ayırmadaki önemini ortaya koyan “Grup Ortalamalarının Eşitliği Testi” ile de, modele giren değişkenlerin grup ortalamalarının farklı olması nedeniyle desteklenmektedir (bkz. EK. 4).

Diskriminant Analizi sonuçlarına göre model, bağımlı değişkendeki varyansın yaklaşık olarak 0,466’sını ($0,683^2$) açıklayabilmektedir. Wilks’ Lamda değeri ile de diskriminant skorlarındaki toplam varyansın gruplar arasındaki farklar tarafından açıklanamayan kısmının 0,534 olduğu görülmektedir (Tablo 4.3). Bu değer ayrıca, fonksiyonun iyi bir ayırım yapıp yapmadığının göstergesi olan özdeğer (eigenvalue) istatistiğinin anlamlılığının testinde kullanılmaktadır. Wilks’ Lamda değeri Tablo 4.3’de görüldüğü üzere anlamlıdır (sig. 0.000).

Model 1 için yapılan analiz sonucunda Diskriminant Fonksiyonu’nun grupları doğru tahmin etme oranı % 81 olarak bulunmuş olmasına rağmen, varsayımlarının sağlanmaması nedeniyle işletme performans kriterlerini etkileyen faktörlerin saptanması için daha az varsayımın sağlanmasının yeterli olduğu Lojistik Regresyon Analizi’nin uygulanmasına karar verilmiştir.

4.5.1.2 Lojistik Regresyon Analizi Sonuçları

Modelin ve parametrelerin uygunluk testine geçmeden önce, modelde bulunan bağımsız değişkenlerin modelin lojit değerleriyle doğrusallık kontrolünü yapmak amacıyla Kesikli Polinomlar Yöntemi uygulanmıştır (STATA paket programında kullanılan kodlar ve sonuçlar için bkz. EK. 5). Bu yöntemin sonuçlarına, göre değişkenlerin modelin lojit değerleri ile doğrusal olduğu ve modele değişkenlerin kendi değerlerinin girmesinin 0,01 anlamlılık düzeyinde uygun olduğu görülmüştür.

Modelin uygunluğunun sınanması* için uygulanan Hosmer-Lemeshow Testi sonucuna göre 0,961 anlamlılık düzeyinde sıfır hipotezi red edilememektedir. Başka bir ifade ile Model 1, veri setini iyi temsil etmektedir.

Modeldeki parametrelerin uygunluğunun sınanması için ise parametrelerin birlikte anlamlılığını test eden Omnibus Testi uygulanmıştır. Bu testin sonucuna göre modeldeki parametrelerin anlamlı olduğu sonucuna varılmıştır (sig. 0,000).

Tablo 4.5 Model 1 için Test ve Doğru Sınıflandırma Sonuçları

Test Sonuçları				Sınıflandırma Sonuçları			
				Sınıf	0	1	Doğru Yüzdesi
Testler	χ^2	<i>df</i>	<i>Sig.</i>	0	25	7	78,1
Hosmer-Lemeshow Testi	2,523	8	0,961	1	5	42	89,4
Omnibus Testi	56,842	6	0,000	Toplam			84,8

$exp(\beta)$ değerlerinin yorumlanmasında sağladığı kolaylık nedeniyle standartlaştırılmış bağımsız değişken değerleriyle model tekrar kurulmuş ve parametreler %0,1 ile %1,5 düzeyinde anlamlı bulunmuştur.

* H_0 : Teorik model veriyi iyi temsil etmektedir.

H_1 : Teorik model veriyi iyi temsil etmemektedir.

Tablo 4.6 Model 1’deki Açıklayıcı Değişkenler ve Katsayıları

Modele Giren Değişkenler				Standartlaştırılmış Değişkenler için Değerler	
Değişkenler	β	<i>Wald</i>	<i>Sig.</i>	β^*	$\exp(\beta^*)$
FinYap2	-9,933	8,914	0,003	-1,832	0,160
SermayeYog2	0,000	11,026	0,001	2,697	14,841
Buyukluk	3,737	7,230	0,007	1,436	4,205
YabSermaye	0,042	6,612	0,010	1,078	2,938
Likidite5	-10,854	5,899	0,015	-1,464	0,231
FinYap5	8,508	6,508	0,011	1,603	4,969
(Sabit)	-1,689	7,469	0,006	1,265	3,544

Model 1’i (Karlılık-Verimlilik-Büyüme) etkileyen bağımlı değişkenlerden en önemlisinin sermaye yoğunluğu (SermayeYog2) olduğu Tablo 4.6’dan açıkça görülmektedir. Modeldeki diğer değişkenlerin sabit kaldığı varsayımıyla sermaye yoğunluğu değişkenindeki bir birimlik değişim; işletmenin, performansı yüksek işletmeler olarak tanımlanan sınıfa (“1” sınıfı) atanması olasılığının, performansı düşük işletmeler olarak tanımlanan sınıfa (“0” sınıfı) atanması olasılığına oranının $\exp(\beta)=14,841$ kat etkilendiği şeklinde yorumlanmaktadır. Kukla değişken “büyüklük” dahil, modeldeki diğer değişkenlerin katsayılarının üstel değerleri de aynı şekilde yorumlanabilmektedir.

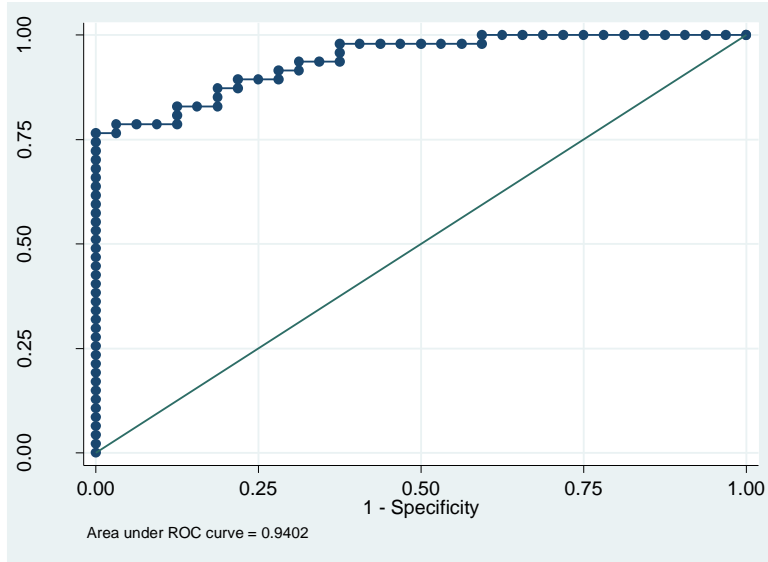
Modelin grupları doğru sınıflandırma yüzdesi Tablo 4.5’den de görüldüğü gibi yanlış sınıflandırma maliyetlerinin birbirine eşit olduğu varsayımıyla, kesim değeri (cut value) “0,50” olarak belirlenmiştir. Analiz sonucunda düşük performanslı işletmeleri doğru sınıflandırma oranı olan duyarlılık (sensitivity) değeri %78,1, yüksek performanslı işletmeleri doğru sınıflandırma oranı olan özgünlük (specificity) değeri %89,4, toplamda grupları doğru sınıflandırma oranı olan doğruluk (accuracy) değeri ise % 84,8 olarak elde edilmiştir.

Sınıflandırma modellerinin değerlendirilmesi ve kıyaslanması için önerilen bir diğer yöntem de ROC eğrisidir (Receiver Operating Characteristic Curves).

Farklı kesim değerleri hesaplanarak oluşturulan ROC eğrisi; modelin sınıflandırma gücünün belirlenmesine, çeşitli modellerin etkinliğinin kıyaslanmasına, en uygun kesim değerinin belirlenmesine ve sonuçların kalitesinin izlenmesine olanak sağlamaktadır.³⁶⁶

ROC eğrilerinin oluşturulduğu koordinat sisteminin ordinatında, testin gerçek pozitif oranı (duyarlılık), apsisinde ise yanlış pozitif oranı (1-özgünlük) yer almaktadır. Model ne kadar iyi ise eğrinin o kadar yukarıya (yüksek duyarlılık bölgesine) ve sola (düşük yanlış pozitif oran (1-özgünlük) bölgesine) doğru kayması beklenmektedir.

ROC eğrisi altında kalan alan, aynı örnek üzerinde geliştirilmiş modellerin karşılaştırılmasında önemli bir performans ölçüsü olmanın yanı sıra modelin doğru tahmin olasılığını da göstermektedir³⁶⁷ ve 0,5 ile 1,0 arasında değer almaktadır. Genel kural olarak, elde edilen değer 0,5 ise ($ROC = 0,5$) model sınıfları ayırmada başarılı değildir (Para atarak gruplara karar vermekle aynı yöntemdir). $0,7 \leq ROC < 0,8$ aralığında modelin grupları kabul edilebilir düzeyde, $0,8 \leq RC < 0,9$ aralığında mükemmel, $ROC \geq 0,9$ olduğunda ise olağanüstü ayırdığı kabul edilmektedir.



Şekil 4.2 Model 1 için ROC Eğrisi

³⁶⁶ A. Dirican, “Tanı Testi Performanslarının Değerlendirilmesi ve Kıyaslanması”, **Cerrahpaşa Tıp Dergisi**, Cilt 32, Sayı 1, 2001, ss. 25–30.

³⁶⁷ N. A. Obuchowski, “ROC Analysis”, **American Journal of Roentgenology**, 184, 2005, pp. 364–372.

ROC eğrisi altındaki alan 0,9402 olarak bulunduğundan, modelin grupları olağanüstü ayırdığı yorumu yapılabilir.

Sonuç olarak, işletmelerin Karlılık-Verimlilik-Büyüme’den oluşan performans göstergelerini etkileyen değişkenlerin saptanması amacıyla, kategorik bağımlı değişken analizlerinden olan Lojistik Regresyon Analizi’nin kullanılmasının uygunluğu açıkça görülmektedir.

4.5.2 Model 2: Performans Göstergesi Olarak Karlılık ve Verimlilik

İşletme performans göstergelerinden olan karlılık ve verimlilik kriterleri için seçilen “aktif karlılığı” ve “emek verimliliği” değerlerini birlikte etkileyen değişkenlerin saptanması amacıyla kurulan Model 2, sırasıyla Diskriminant ve Lojistik Regresyon Analizi kullanılmıştır.

4.5.2.1 Diskriminant Analizi Sonuçları

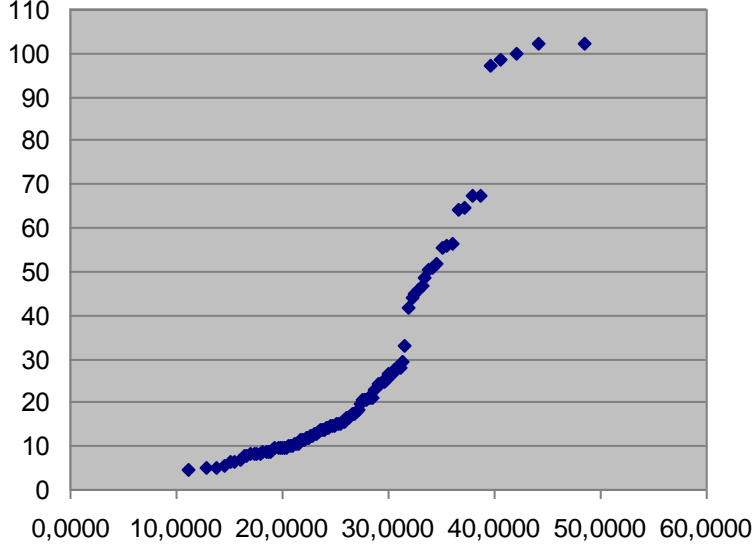
Model 2 oluşturulurken birçok bağımsız değişken arasında güçlü ilişki olduğu korelasyon matrisine bakılarak saptanmıştır. Bunun sonucunda diskriminant fonksiyonuna girecek değişkenlerinin de Adım Adım Regresyon Yöntemi ile seçilmesine karar verilmiştir.

Eşit varyanslılık varsayımı için grupların varyansları birbirine eşittir şeklindeki sıfır hipotezinin, Box’s M test sonucuna göre reddildiği görülmektedir (sig. 0,000). Diğer bir ifadeyle, eşit varyanslılık varsayımı sağlanmamaktadır.

Tablo 4.7 Model 2 için Diskriminant Analizi Testleri ve Sonuçları

Box’s M		Eigenvalues			Wilks’ Lambda	
Değeri	Sig.	Değeri	Varyans (%)	Kanonik Korelasyon	Değeri	Sig.
2341,607	0,000	0,777	100,00	0,661	0,563	0,000

Çoklu normallik varsayımının da sağlanmadığı Şekil 4.3'ten görülmektedir.



Şekil 4.3 Model 2 için Serpilme Diyagramı

Adım Adım Regresyon Yöntemi sonucunda elde edilen diskriminant fonksiyonu Tablo 4.8’de verilen standartlaştırılmış kanonik değerler kullanılarak aşağıdaki gibi yazılmaktadır.

$$T = -0,809 \text{FinYap1} + 0,444 \text{SermayeYog2} + 0,557 \text{Buyukluk}$$

Model sonuçları, bağımsız değişkenlerin sınıfları ayırmadaki önemini ortaya koyan “Grup Ortalamalarının Eşitliği Testi” ile de, modele giren değişkenlerin grup ortalamalarının farklı olması nedeniyle desteklenmektedir.

Tablo 4.8 Model 2 için Diskriminant Modeli ve Doğru Sınıflandırma Sonuçları

Standartlaştırılmış ve Standartlaştırılmamış Kanonik Diskriminant Fonksiyon Katsayıları			Sınıflandırma Sonuçları			
Değişkenler	β	β^*	Sınıf	0	1	DoğruYüzdesi
FinYap1	- 4,629	-0,809	0	37	10	78,7
SermayeYog2	0,000	0,444	1	9	48	84,2
Buyukluk	1,592	0,557	Toplam			81,7
(Sabit)	1,276					

Diskriminant Analizi sonuçlarına göre model, bağımlı değişken performansındaki varyansın yaklaşık olarak 0,437'sini ($0,661^2$) açıklayabilmektedir. Wilks' Lamda değeri ile de diskriminant skorlarındaki toplam varyansın gruplar arasındaki farklar tarafından açıklanamayan kısmının 0,563 olduğu görülmektedir (Tablo 4.7). Bu değer ayrıca, fonksiyonun iyi bir ayırım yapıp yapmadığının göstergesi olan özdeğer (eigenvalues) istatistiğinin anlamlılığında kullanılan Wilks' Lamda değeri Tablo 4.7'den görüldüğü üzere anlamlıdır (sig. 0.000).

Model 2 için yapılan analiz sonucunda diskriminant fonksiyonunun grupları doğru tahmin etme oranı % 81,7 olarak bulunmuş olmasına rağmen, çoklu normallik ve eşit varyanslılık varsayımlarının sağlanmaması nedeniyle işletme performans kriterlerini etkileyen faktörlerin saptanması için Lojistik Regresyon Analizi'nin uygulanmasına karar verilmiştir.

4.5.2.2 Lojistik Regresyon Analizi Sonuçları

Modelin ve parametrelerin uygunluk testine geçmeden, önce modelde bulunan bağımsız değişkenlerin modelin lojit değerleriyle doğrusallık kontrolünü yapmak amacıyla Kesikli Polinomlar Yöntemi uygulanmıştır. Bu yöntemin sonuçlarına göre, değişkenlerin modelin lojit değerleri ile doğrusal olduğu ve modele değişkenlerin kendi değerlerinin girmesinin 0,01 anlamlılık düzeyinde uygun olduğu görülmüştür.

Modelin uygunluğu için kullanılan Hosmer-Lemeshow testi sonucuna göre modelin verileri iyi temsil ettiğinin iddia edildiği sıfır hipotezi 0.597 anlamlılık düzeyinde kabul edilmektedir.

Modeldeki parametrelerin birlikte anlamlılığının sınanması için Omnibus Testi uygulanmış ve elde edilen sonuçlar incelendiğinde modeldeki parametrelerin anlamlı olduğuna karar verilmiştir (sig. 0,000).

Tablo 4.9 Model 2 için Test ve Doğru Sınıflandırma Sonuçları

Test Sonuçları				Sınıflandırma Sonuçları			
				Sınıf	0	1	Doğru Yüzdesi
Testler	χ^2	<i>df</i>	<i>Sig.</i>	0	39	8	83,0
Hosmer-Lemeshow Testi	6,446	8	0,597	1	6	51	89,5
Omnibus Testi	78,873	4	0,000	Toplam			86,5

Adım Adım Regresyon Yöntemiyle kurulan Lojistik Regresyon Modeli'ndeki bağımsız değişkenler Tablo 4.10'da görülmektedir. $exp(\beta)$ değerlerinin yorumlanmasında sağladığı kolaylık nedeniyle standartlaştırılmış bağımsız değişken değerleriyle model tekrar kurulmuş ve FinYap1, SermayeYog2 ve Buyukluk değişkenlerinin anlamlı olduğuna karar verilmiştir (sig. 0,000) ayrıca YabSermaye değişkeni de %3,8 düzeyinde anlamlı bulunmuştur.

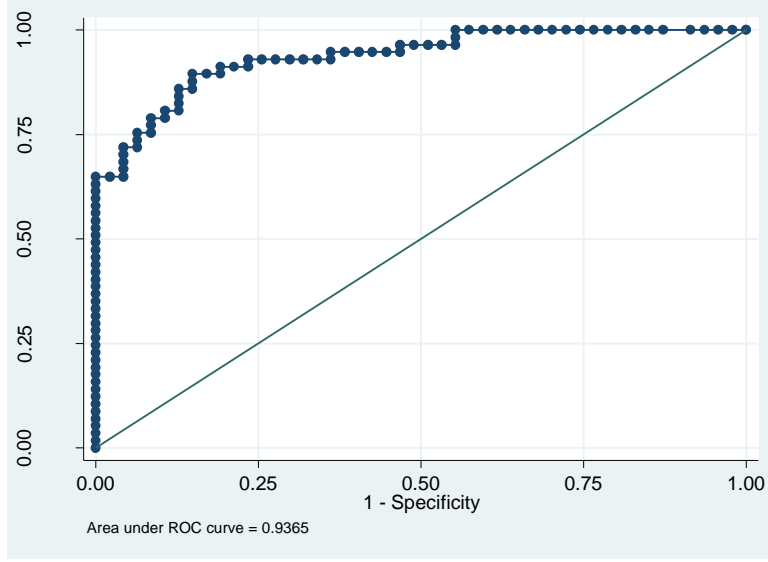
Tablo 4.10 Model 2'deki Açıklayıcı Değişkenler ve Katsayıları

Modele Giren Değişkenler				Standartlaştırılmış Değişkenler için Değerler	
Değişkenler	β	<i>Wald</i>	<i>Sig.</i>	β^*	$exp(\beta^*)$
FinYap1	-9,273	16,915	0,000	-1,874	0,154
SermayeYog2	0,000	15,117	0,000	3,585	36,043
Buyukluk	5,397	13,436	0,000	2,006	7,430
YabSermaye	0,023	4,306	0,038	0,632	1,881
(Sabit)	1,090	1,678	0,195	1,274	3,574

İşletmeler için performans göstergelerinden olan karlılık-verimlilik kriterleri birlikte ele alındığında Model 2'nin Model 1'deki değişkenlere oldukça benzer olduğu görülmektedir. İki modelin değişkenleri arasındaki fark, Model 1'de likidite oranlarından beşincisinin yer alması ve Model 1'de ikinci ve beşinci finansal yapı oranları yer alırken Model 2'de finansal yapı oranlarından sadece birincisinin modelde bulunmasıdır.

Modelin grupları doğru sınıflandırma oranı Tablo 4.9'dan da görülebileceği üzere %86,5 olarak bulunmuştur. Sınıflandırma modellerinin performansının

değerlendirilmesi için önerilen bir yöntem olan ROC eğrisi altındaki alan 0,9365 olarak bulunmuştur. Bu sonuçla modelin grup ayrımını mükemmel yaptığı söylenebilmektedir.



Şekil 4.4 Model 2 için ROC Eğrisi

Sonuç olarak, işletmelerin Karlılık-Verimlilik'ten oluşan performans göstergelerini etkileyen değişkenlerin tespit edilmesi için Lojistik Regresyon Analizinin kullanılmasının uygunluğu açıkça görülmektedir.

4.5.3 Model 3: Performans Göstergesi Olarak Karlılık ve Büyüme

İşletme performans kriterlerinden karlılık ve büyüme piyasa göstergeleri değerlerini etkileyen faktörlerin saptanması amacıyla kurulan Model 3 için Diskriminant ve Lojistik Regresyon Analiz sonuçları aşağıda verilmiştir. Bu modelde karlılık ve piyasa göstergeleri için seçilen “aktif karlılığı” ve “satışlardaki reel artış oranı” değerlerini birlikte açıklayan değişkenlerin belirlenmesi amaçlanmıştır.

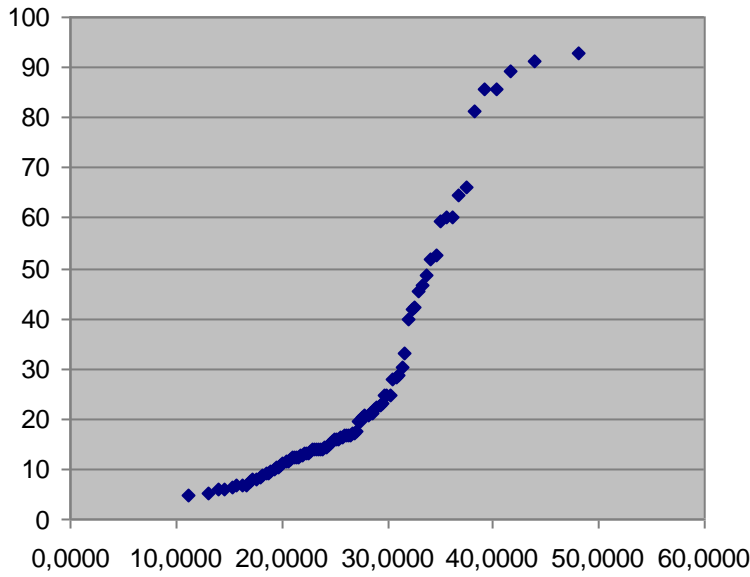
4.5.3.1 Diskriminant Analizi Sonuçları

Model 3 için uygulanan Diskriminant Analizi sonuçlarına göre, grupların varyanslarının birbirine eşit olduğunu ileri süren sıfır hipotezi Box's M testine göre reddedilmekte (sig. 0,000) ve eşit varyanslılık varsayımı bu modelde de sağlanmamaktadır.

Tablo 4.11 Model 3 için Diskriminant Analizi Testleri ve Sonuçları

Box's M		Eigenvalues			Wilks' Lambda	
Değeri	Sig.	Değeri	Varyans (%)	Kanonik Korelasyon	Değeri	Sig.
1868,295	0,000	0,563	100,00	0,600	0,640	0,000

Ayrıca, çoklu normallik varsayımının da sağlanmadığı çizilen serpilme diyagramından görülmektedir (Şekil 4.5).



Şekil 4.5 Model 3 için Serpilme Diyagramı

Adım Adım Regresyon Yöntemi sonucunda elde edilen standartlaştırılmış diskriminant fonksiyonu Tablo 4.12’de verilen değerler kullanılarak aşağıdaki gibi yazılmaktadır.

$$T = 0,444 Likidite5 + 0,945FinYap2 + 0,528Faaliyet1 - 1,060Faaliyet7 - 0,627 YabSermaye$$

Model sonuçları, bağımsız değişkenlerin sınıfları ayırmadaki anlamlılığına bakan “Grup Ortalamalarının Eşitliği Testi” ile desteklenmektedir.

Tablo 4.12 Model 3 için Diskriminant Modeli ve Doğru Sınıflandırma Sonuçları

Standartlaştırılmış ve Standartlaştırılmamış Kanonik Diskriminant Fonksiyon Katsayıları			Sınıflandırma Sonuçları			
Değişkenler	β	β^*	Sınıf	0	1	Doğru Yüzdesi
Likidite5	3,906	0,444	0	28	13	68,3
FinYap2	5,526	0,945	1	4	50	92,6
Faaliyet1	0,001	0,528	Toplam			81,7
Faaliyet7	-1,072	-1,060				
YabSermaye	-0,023	-0,627				
(Sabit)	-1.030					

Diskriminant Analizi sonuçlarına göre model, bağımlı değişkendeki varyansın yaklaşık olarak 0,36'sini ($0,600^2$) açıklayabilmektedir. Wilks' Lamda değeri ile de diskriminant skorlarındaki toplam varyansın gruplar arasındaki farklar tarafından açıklanamayan kısmının 0,640 olduğu görülmektedir (Tablo 4.11). Aynı zamanda fonksiyonun iyi bir ayırım yapıp yapmadığının göstergesi olan özdeğer (eigenvalues) istatistiğinin anlamlılığında kullanılan Wilks' Lamda değeri Tablo 4.11'den görüldüğü üzere anlamlıdır (sig. 0.000).

Model 3 için yapılan analiz sonucunda diskriminant fonksiyonunun grupları doğru tahmin etme oranı %82,1 olarak bulunmuş olmasına rağmen, varsayımlarının sağlanmaması nedeniyle işletme performans kriterlerini etkileyen faktörlerin belirlenmesi için Lojistik Regresyon Analizi'nin uygulanmasına karar verilmiştir.

4.5.3.2 Lojistik Regresyon Analizi Sonuçları

Modelin ve parametrelerin uygunluk testinden önce gerçekleştirilen Kesikli Polinomlar Yöntemi sonuçlarına göre, değişkenlerin modelin lojit değerleri ile doğrusal olduğu ve modele değişkenlerin kendi değerlerinin girmesinin 0,01 anlamlılık düzeyinde uygun olduğu görülmüştür.

Modelin uygunluğu için kullanılan Hosmer-Lemeshow test sonucuna göre, 0,064 anlamlılık düzeyinde teorik model veri setini iyi temsil ettiğini ileri süren sıfır hipotezinin anlamlı olduğu kabul edilmektedir.

Model parametrelerinin birlikte anlamlılığını sınyan Omnibus test sonucu incelendiğinde ise, modeldeki bağımsız deęişkenlere ait parametrelerin anlamlı olduęu söylenebilmektedir.

Tablo 4.13 Model 3 için Test ve Doğru Sınıflandırma Sonuçları

Test Sonuçları				Sınıflandırma Sonuçları			
				Sınıf	0	1	Doğru Yüzdesi
Testler	χ^2	<i>df</i>	<i>Sig.</i>	0	30	11	73,1
Hosmer-Lemeshow Testi	14,759	8	0,064	1	5	49	90,7
Omnibus Testi	44,385	6	0,000	Toplam			83,2

Çoklu doğrusal bağlantı sorunu ile karşılaşmamak için modele girecek deęişkenlerin seçimi yine Adım Adım Regresyon Yöntemi ile yapılmıştır. Modele giren deęişkenler Tablo 4.14'te verilmiştir.

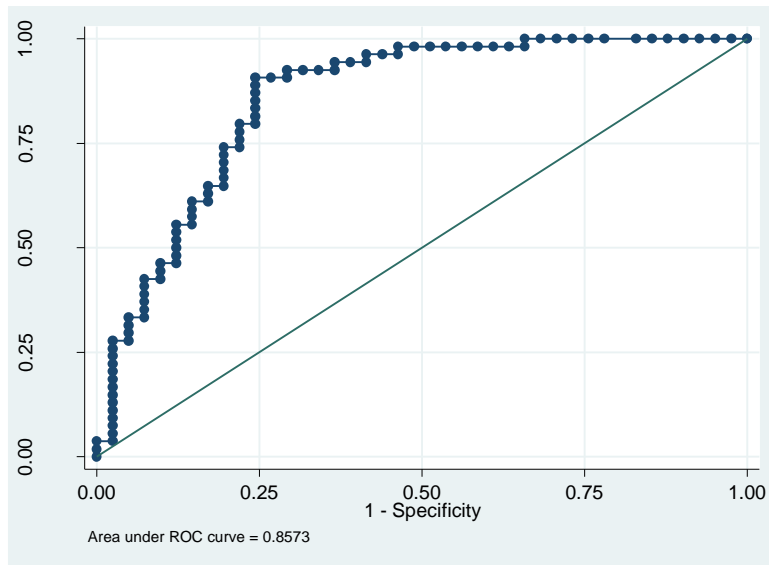
Yorumlamadaki kolaylığı nedeniyle, standartlaştırılmış bağımsız deęişken deęerleri ile model tekrar elde edilmiştir. Tablo 4.14'ten görüldüğü gibi Model 3 için FinYap2 ve Faaliyet7 oranları önemli birer açıklayıcı deęişkendir. Deęişkenlerin tek tek anlamlılıkları test edildiğinde Faaliyet1'in dışındaki deęişkenlerin %0,1 ile %2,8 düzeyinde modelde olması anlamlı bulunmuştur. Faaliyet1'in de anlamlılık düzeyinin 0,057 deęeri ile modelde bulunduğunda grupları doğru atama oranını arttırdığı ve bu deęişkenin işletmelerin nakit politikalarına ilişkin deęerlendirme yapılmasına imkan vermesi nedeniyle modelde kalması uygun bulunmuştur.

Tablo 4.14 Model 3'deki Açıklayıcı Değişkenler ve Katsayıları

Modele Giren Değişkenler				Standartlaştırılmış Değişkenler için Değerler	
Değişkenler	β	Wald	Sig.	β^*	$\exp(\beta^*)$
FinYap2	-9,099	15,028	0,000	-1,693	0,184
Faaliyet7	2,272	11,301	0,001	2,255	9,534
YabSermaye	0,039	5,178	0,023	1,033	2,810
Likidite5	-6,762	4,831	0,028	-0,781	0,458
Faaliyet1	-0,003	3,614	0,057	-1,663	0,190
(Sabit)	1,558	0,468	0,036	0,209	1,233

Modelin grupları doğru sınıflandırma yüzdesi Tablo 4.13'ten de görüldüğü gibi yanlış sınıflandırma maliyetlerinin birbirine eşit olduğu varsayımıyla, kesim değeri (cut value) "0,50" olarak belirlenmiştir. Analiz sonucunda düşük performanslı işletmeleri doğru sınıflandırma oranı olan duyarlılık (sensitivity) değeri %73,1, yüksek performanslı işletmeleri doğru sınıflandırma oranı olan özgünlük (specificity) değeri %90,7, toplamda grupları doğru sınıflandırma oranı olan doğruluk (accuracy) değeri ise % 83,2 olarak elde edilmiştir.

ROC eğrisi çizildiğinde de eğrinin altında kalan alanın 0,8573 değeriyle $0,8 \leq ROC < 0,9$ aralığında yer aldığı ve sonuç olarak modelin, grupları mükemmel ayırdığı görülmektedir.

**Şekil 4.6 Model 3 için ROC Eğrisi**

Sonuç olarak, işletmelerin Karlılık-Büyüme'den oluşan performans göstergelerini etkileyen değişkenlerin tespit edilmesi amacıyla Lojistik Regresyon Analizi kullanılmasının uygunluğu açıkça görülmektedir.

4.5.4 Model 4: Performans Göstergesi Olarak Karlılık ve Piyasa

Bu modelde karlılık ve piyasa performans göstergeleri için seçilen “aktif karlılığı” ile “fiyat-kazanç” değerlerini birlikte açıklayan faktörlerin saptanması amaçlanmıştır.

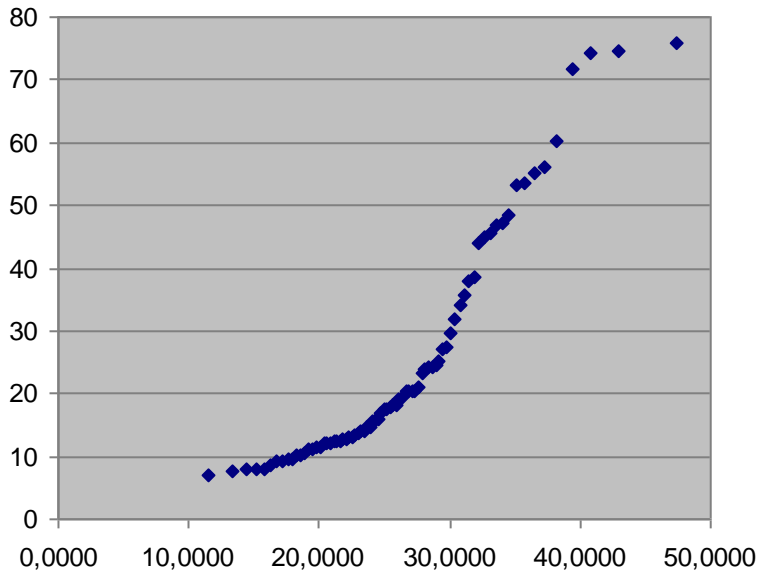
4.5.4.1 Diskriminant Analizi Sonuçları

Model 4 için uygulanan Diskriminant Analizi sonuçlarına bakıldığında grupların varyansların birbirine eşit olduğunu ileri süren sıfır hipotezi Box's M testine göre reddedilmekte (sig. 0,000) ve yine eşit varyanslılık varsayımı sağlanmamaktadır.

Tablo 4.15 Model 4 için Diskriminant Analizi Testleri ve Sonuçları

Box's M		Eigenvalues			Wilks' Lambda	
Değeri	Sig.	Değeri	Varyans (%)	Kanonik Korelasyon	Değeri	Sig.
1792,673	0,000	0,617	100,00	0,618	0,600	0,000

Ayrıca, çoklu normallik varsayımının da sağlanmadığı çizilen serpilme diyagramından görülmektedir (Şekil 4.7).



Şekil 4.7 Model 4 için Serpilme Diyagramı

Adım Adım Regresyon Yöntemi sonucunda elde edilen standartlaştırılmış diskriminant fonksiyonu Tablo 4.16’da verilen değerler kullanılarak aşağıdaki gibi yazılmaktadır.

$$T = -0,913FinYap1 + 0,505PazarPay + 0,448Buyukluk$$

Model sonuçları, bağımsız değişkenlerin sınıfları ayırmadaki önemini ortaya koyan “Grup Ortalamalarının Eşitliği Testi” ile desteklenmektedir.

Tablo 4.16 Model 4 için Diskriminant Modeli ve Doğru Sınıflandırma Sonuçları

Standartlaştırılmış ve Standartlaştırılmamış Kanonik Diskriminant Fonksiyon Katsayıları			Sınıflandırma Sonuçları			
Değişkenler	β	β^*	Sınıf	0	1	DoğruYüzdesi
FinYap1	-5,240	-0,913	0	21	10	67,6
PazarPay	6,939	0,505	1	6	41	87,2
Buyukluk	1,213	0,448	Toplam			79,5
(Sabit)	1,747					

Diskriminant Analizi sonuçlarına göre model, bağımlı değişkendeki varyansın yaklaşık olarak 0,38’ini ($0,618^2$) açıklayabilmektedir. Wilks’ Lamda değeri ile de diskriminant skorlarındaki toplam varyansın gruplar arasındaki farklar tarafından açıklanamayan kısmının 0,600 olduğu görülmektedir (Tablo 4.15). Aynı zamanda fonksiyonun iyi bir ayırım yapıp yapmadığının göstergesi olan ve özdeğer (eigenvalues) istatistiğinin anlamlılığında kullanılan Wilks’ Lamda değerinin anlamlı olduğu görülmektedir (sig. 0.000).

Model 4 için yapılan analiz sonucunda diskriminant fonksiyonunun grupları doğru sınıflandırma oranı %79,5 olarak bulunmuş olmasına rağmen, varsayımlarının sağlanmaması nedeniyle işletme performans kriterlerini etkileyen faktörlerin tespiti için Lojistik Regresyon Analizi’nin uygulanmasına karar verilmiştir.

4.5.4.2 Lojistik Regresyon Analizi Sonuçları

Karlılık-Piyasa kriterlerine göre işletme performansını etkileyen faktörler olarak Adım Adım Regresyon Yöntemi sonucunda modele giren değişkenler; FinYap1,

PazarPayı ve Likidite1 oranlarıdır (bkz. Tablo 4.18). Bu değişkenler ile modelden hesaplanan lojit değerlerinin doğrusallık kontrolleri yapıldığında, modele giren üç değişkenin de 0,01 anlamlılık düzeyinde doğrusal olduğu görülmektedir.

Modelin uygunluğu için kullanılan Hosmer-Lemeshow testi sonucuna göre modelin veri setini iyi temsil ettiğini ileri süren sıfır hipotezi 0,530 anlamlılık düzeyinde kabul edilmiştir. Ayrıca Omnibus test sonucuna bakılarak, modeldeki üç parametrenin birlikte modelde bulunmasının anlamlı olduğuna karar verilmiştir.

Tablo 4.17 Model 4 için Test ve Doğru Sınıflandırma Sonuçları

Test Sonuçları				Sınıflandırma Sonuçları			
				Sınıf	0	1	Doğru Yüzdesi
Testler	χ^2	<i>df</i>	<i>Sig.</i>	0	23	8	74,2
Hosmer-Lemeshow Testi	7,061	8	0,530	1	6	41	87,2
Omnibus Testi	44,422	3	0,000	Toplam			82,1

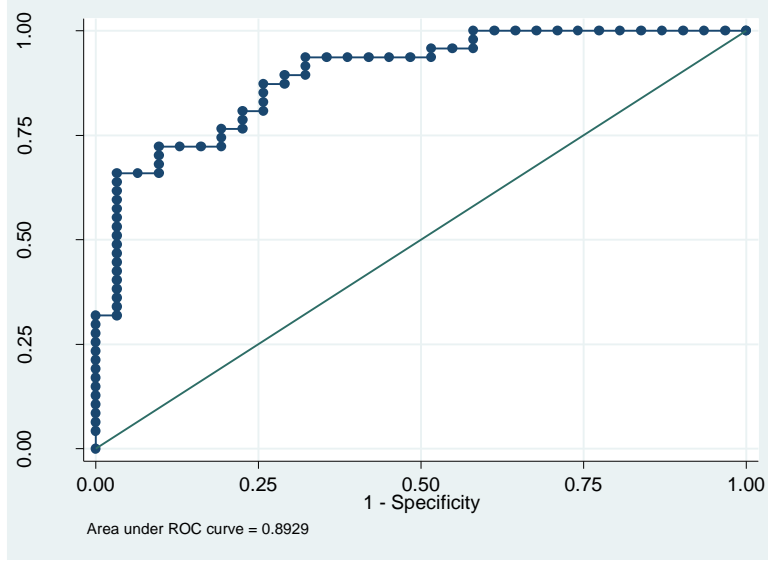
Model kurulurken kullanılan Adım Adım Regresyon Yöntemi'ne göre, modele giren değişkenler ve yorumlamadaki kolaylığı ile kullanılan standartlaştırılmış değişkenler için değerleri Tablo 4.18'de verilmiştir ve FinYap1 değişkenini 0,000 anlamlılık düzeyi, PazarPay değişkenininin %0,5 ve Likidite1 değişkenininin de %4,9 düzeyinde modelde kalması anlamlı bulunmuştur.

Tablo 4.18 Model 4'teki Açıklayıcı Değişkenler ve Katsayıları

Modele Giren Değişkenler				Standartlaştırılmış Değişkenler için Değerler	
Değişkenler	β	<i>Wald</i>	<i>Sig.</i>	β^*	$\exp(\beta^*)$
FinYap1	-12,687	15,356	0,000	-2,504	0,082
PazarPay	65,391	7,948	0,005	4,902	134,604
Likidite1	-0,583	3,888	0,049	-0,931	0,394
(Sabit)	5,536	9,755	0,002	1,478	4,382

Modelin grupları doğru sınıflandırma yüzdesi Tablo 4.17'de görüldüğü gibi % 82,1 olarak elde edilmiştir. ROC eğrisi altında kalan alan ise 0,8929 olarak bulunmuştur

(Şekil 4.8). Bu sonuçla modelin grup ayırımı mükemmel yaptığı söylenebilmektedir.



Şekil 4.8 Model 4 için ROC Eğrisi

Sonuç olarak, işletmelerin Karlılık-Piyasa performans göstergelerini birlikte açıklayan değişkenlerin belirlenmesi için Lojistik Regresyon Analizi sonuçlarının kullanılmasının uygunluğu açıkça görülmektedir.

4.5.5 Model 5: Performans Göstergesi Olarak Verimlilik ve Büyüme

Bu modelde verimlik ve büyüme performans göstergeleri için seçilen “emek verimliliği” ile “satışlardaki reel artış” değerlerini birlikte açıklayan faktörlerin saptanması amaçlanmıştır. Diskriminant ve Lojistik Regresyon Analiz sonuçları aşağıda verilmiştir.

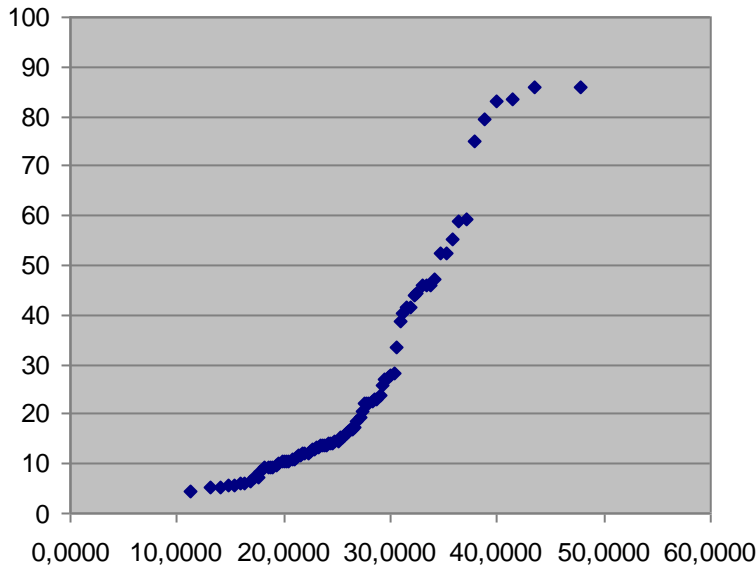
4.5.5.1 Diskriminant Analizi Sonuçları

Model 5 için uygulanan Diskriminant Analizi sonuçları incelendiğinde, grupların varyansların birbirine eşit olduğunu ileri süren sıfır hipotezi Box's M testine göre reddedilmekte (sig. 0,000) ve eşit varyanslılık varsayımının bu modelde de sağlanmadığı görülmüştür.

Tablo 4.19 Model 5 için Diskriminant Analizi Testleri ve Sonuçları

Box's M		Eigenvalues			Wilks' Lambda	
Değeri	Sig.	Değeri	Varyans (%)	Kanonik Korelasyon	Değeri	Sig.
1966,907	0,000	0,502	100,00	0,578	0,666	0,000

Ayrıca, çoklu normallik varsayımının da sağlanmadığı Şekil 4.9'da verilen serpilme diyagramından görülmektedir.



Şekil 4.9 Model 5 için Serpilme Diyagramı

Adım Adım Regresyon Yöntemi sonucunda elde edilen standartlaştırılmış diskriminant fonksiyonu Tablo 4.20'de verilen değerler kullanılarak aşağıdaki gibi yazılmaktadır.

$$T = -0,700 \text{ FinYap2} + 0,386 \text{ Faaliyet4} + 0,444 \text{ SermayeYg1} \\ + 0,712 \text{ YabSermaye}$$

Model sonuçları, bağımsız değişkenlerin sınıfları ayırmadaki anlamlılığına bakan "Grup Ortalamalarının Eşitliği Testi" ile desteklenmektedir.

Tablo 4.20 Model 5 için Diskriminant Modeli ve Doğru Sınıflandırma Sonuçları

Standartlaştırılmış ve Standartlaştırılmamış Kanonik Diskriminant Fonksiyon Katsayıları			Sınıflandırma Sonuçları			
Değişkenler	β	β^*	Sınıf	0	1	Doğru Yüzdesi
FinYap2	-3,504	-0,700	0	30	11	73,2
Faaliyet4	0,398	0,386	1	8	39	83,0
SermayeYog1	0,000	0,444	Toplam			78,4
YabSermaye (Sabit)	0,028	0,712				
	-0,445					

Diskriminant Analizi sonuçlarına göre model, bağımlı değişkendeki varyansın yaklaşık olarak 0,334'ünü ($0,578^2$) açıklayabilmektedir. Wilks' Lamda değeri ile de diskriminant skorlarındaki toplam varyansın gruplar arasındaki farklar tarafından açıklanamayan kısmının 0,666 olduğu görülmektedir (Tablo 4.19). Aynı zamanda fonksiyonun iyi bir ayırım yapıp yapmadığının göstergesi olan ve özdeğer (eigenvalues) istatistiğinin anlamlılığında da kullanılan Wilks' Lamda değeri anlamlıdır (sig. 0.000).

Model 5 için yapılan analiz ile diskriminant fonksiyonunun grupları doğru tahmin etme oranı %78,4 olarak bulunmuş olmasına rağmen, varsayımlarının sağlanmaması nedeniyle işletme performans kriterlerini etkileyen faktörlerin saptanması için Lojistik Regresyon Analizi'nin uygulanmasına karar verilmiştir.

4.5.5.2 Lojistik Regresyon Analizi Sonuçları

Verimlilik-Büyüme kriterlerine göre işletme performansı değerlendirilmek istendiğinde, bu performans göstergelerini etkileyen açıklayıcı değişkenler Tablo 4.22'de yer almaktadır.

Mode giren değişkenlerle bu modelden hesaplanan lojit değerleri arasında doğrusallık kontrolü sonucunda değişkenlerin modelde dönüşüm yapılmadan kullanılmasının uygun olduğuna karar verilmiştir.

Modelin uygunluğu için Hosmer-Lemeshow testi sonucu incelendiğinde, modelin veri setini iyi temsil ettiği söylenebilmektedir (sig. 0.444). Ayrıca modeldeki

parametrelerin birlikte anlamlılığını test eden Omnibus testi sonucuna bakılarak parametrelerin birlikte modelde olmasının anlamlı olduğuna karar verilmiştir (sig.0.000).

Tablo 4.21 Model 5 için Test ve Doğru Sınıflandırma Sonuçları

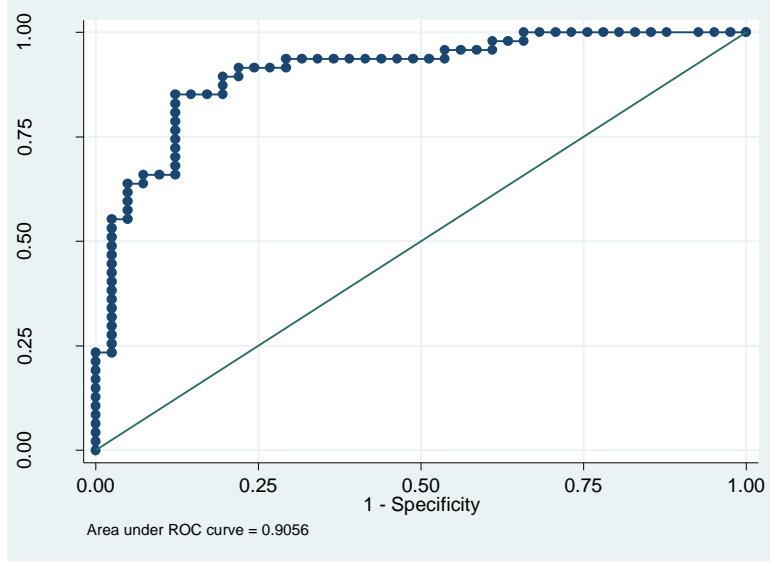
Test Sonuçları				Sınıflandırma Sonuçları			
				Sınıf	0	1	Doğru Yüzdesi
Testler	χ^2	<i>df</i>	<i>Sig.</i>	0	33	8	80,5
Hosmer-Lemeshow Testi	7,890	8	0,444	1	6	41	87,2
Omnibus Testi	53,593	5	0,000	Toplam			84,1

Adım Adım Regresyon Yöntemi'yle kurulan Lojistik Regresyon Modeli'ndeki bağımsız değişkenler Tablo 4.22'de görülmektedir. $exp(\beta)$ değerlerinin yorumlanmasında sağladığı kolaylık nedeniyle, standartlaştırılmış bağımsız değişken değerleriyle model tekrar kurulmuş ve parametrelerin %0,1 ile %2,4 düzeyinde anlamlı oldukları saptanmıştır.

Tablo 4.22 Model 5'deki Açıklayıcı Değişkenler ve Katsayıları

Modele Giren Değişkenler				Standartlaştırılmış Değişkenler için Değerler	
Değişkenler	β	<i>Wald</i>	<i>Sig.</i>	β^*	$exp(\beta^*)$
YabSermaye	0,056	10,737	0,001	1,476	4,376
Faaliyet7	2,729	8,875	0,003	1,257	3,516
FinYap2	-7,337	8,344	0,004	-1,570	0,208
SermayeYog2	0,000	7,838	0,005	3,232	25,338
Likidite5	-7,240	5,066	0,024	-0,868	0,420
(Sabit)	-1,205	1,189	0,275	0,522	1,685

Tablo 4.21'den de görülebileceği üzere, modelin sınıfları doğru tahmin etme oranı %84,1 olarak bulunmuştur. Bu orana alternatif bir yaklaşım olan ROC eğrisi değerine bakıldığında ise eğrinin altında kalan alanın 0,9056 olduğu ve modelin sınıfları olağanüstü ayırdığı sonucuna varılabilmektedir.



Şekil 4.10 Model 5 için ROC Eğrisi

Sonuç olarak, işletmelerin Verimlilik-Büyüme'den oluşan performans göstergelerini etkileyen değişkenlerin belirlenmesi amacıyla, kategorik bağımlı değişken analizlerinden olan Lojistik Regresyon Analizi'nin kullanılmasının uygunluğu açıkça görülmektedir.

SONUÇ

Çalışmanın sonuç bölümünde işletmenin finansal performansını (karlılık, verimlilik, büyüme ve piyasa) etkileyen faktörler uygulama bölümünde belirlenen kriterler dikkate alınarak elde edilen modeller çerçevesinde değerlendirilmiştir. Sonuçlar yorumlanırken verilerin elde edildiği 2007 yılına ait ekonomik koşullar da dikkate alınmıştır. Modeller incelendiğinde Lojistik Regresyon Analizi sonuçları şu şekilde elde edilmiştir:

- **Model 1’de (Karlılık-Verimlilik-Büyüme)** yer alan değişkenler: SermayeYog2, FinYap2(-), Büyüklük, YabSermaye, FinYap5 ve Likidite5(-).
Grupları doğru sınıflandırma oranı: % 84,8.
- **Model 2’de (Karlılık-Verimlilik)** yer alan değişkenler: FinYap1(-), SermayeYog2, Büyüklük, YabSermaye.
Grupları doğru sınıflandırma oranı: % 86,5.
- **Model 3’de (Karlılık-Büyüme)** yer alan değişkenler: FinYap2(-), Faaliyet7, Likidite5(-) ve Faaliyet1.
Grupları doğru sınıflandırma oranı: % 83,2.
- **Model 4’te (Karlılık-Piyasa)** yer alan değişkenler: FinYap1(-), PazarPay ve Likidite1(-).
Grupları doğru sınıflandırma oranı: % 82,1.
- **Model 5’de (Verimlilik-Büyüme)** yer alan değişkenler: YabSermaye, Faaliyet7, FinYap2(-), SermayeYog2 ve Likidite5(-).
Grupları doğru sınıflandırma oranı: % 84,1.

Çalışmada Model 4 (Karlılık-Piyasa) dışında diğer tüm modeller performansı etkileyen faktör olarak “yabancı sermaye”yi belirlenmiştir. Bu değişkenin 2007 yılında açıklayıcı bir değişken olmasındaki nedenin, Türkiye’ye gelen yabancı sermaye (Bkz. Tablo 3.2) hacmindeki artış olduğu söylenebilir. İşletmelerin performanslarının değerlendirilmesinde faaliyet gösterdikleri ülkelerin ekonomik politikalarının ve yatırım kararlarının etkisini görmek mümkündür. Türkiye’nin 2007 yılında ekonomik istikrar yaşaması ve AB uyum sürecinde doğrudan yabancı

sermaye yatırımları* çekmek için yaptığı yasal düzenlemelerin etkisiyle yüksek seviyede gerçekleşen yabancı sermaye akışı, performans kriterlerini etkileyen faktörlerde önemli bir açıklayıcı değişken olarak görülmektedir. Bununla birlikte aynı dönem için yabancı sermayenin hem ekonomik gelişmelere hem de işletmelerin performansına olan etkisinin önemi de dikkat edilmesi gereken bir durumdur.

Rekabet koşullarının her geçen gün daha da zorlaşması işletmelerin rakipleri karşısında daha esnek, yenilikçi ve farklılık yaratmaları gerekliliğini ortaya çıkartmaktadır. Böylece işletmeler maliyetlerini düşürerek kaliteli ürün üretme yarışına girmişlerdir. Sermaye yoğunluğu işletme verimliliğini doğrudan etkileyen bir değişken olarak analiz sonuçlarında yer almıştır. Model 1 (Karlılık-Verimlilik-Büyüme), Model 2 (Karlılık-Verimlilik) ve Model 5 (Verimlilik-Büyüme) incelendiğinde “sermaye yoğunluğu” değişkeninin işletmelerin performansı değerlendirilirken önemli bir kriter olduğu sonucuna varılmıştır. Bu değişkenin modellerin içinde yer alması, performans göstergelerinden özellikle “verimlilik” değişkeninin açıklanmasında önemli bir faktör olmasından kaynaklanabilir. İşletme verimliliği ve sermaye yoğunluğu arasındaki doğrudan ve yüksek ilişkinin işletmelerin 2007 yılında finansal yapılarının güçlenmesiyle (yabancı sermaye yatırımlarıyla birlikte) fazla fonları Ar-Ge faaliyetleri için değerlendirmelerinden kaynaklandığı sonucuna varmak mümkündür.

2007 yılının en önemli ekonomik faaliyetlerinden olan ihracattaki büyüme (Bkz. Tablo 3.1), işletmelerin performanslarını özellikle karlılık ve verimlilik açısından etkilediği görülmüştür. İşletmelerin artan ihracatları satış hacimlerini yükseltmiş ve bu durum pazar payında olumlu etki yaratmıştır. İhracat verimliliğinin artmasına yol açabildiği gibi, verimliliğin artması da ihracatın gelişmesinde etkili olabilmektedir. Dış piyasa ile olan rekabet işletmelerin düşük fiyat avantajı elde edebilmek için maliyetlerini azaltma konusunda baskı oluşturmakta ve uluslararası pazarlara olan

* “Doğrudan yabancı sermaye yatırımları uluslar arası sermaye akışı demektir: ya yabancı bir ülkede bir firma kurulur, ya da yabancı ülkedeki bir işletmeye sermaye yatırılarak ana firma büyür.” Suna Oksay, “Çokuluslu Şirketler Teorileri Çerçevesinde, Yabancı Sermaye Yatırımlarının İncelenerek, Değerlendirilmesi”, Çevirimiçi <http://www.econturk.org/Turkiyeekonomisi/oksay3.pdf>, 10 Eylül 2011.

açılımla da işletmeler yeni teknolojileri kullanmaya zorlamaktadır. Bu durum da doğrudan verimliliği arttırıcı bir faktör olarak ortaya çıkmaktadır.³⁶⁸

İhracat genişlemesi faktör verimliliğini yükseltmekte, teknolojik yeniliklerin adaptasyonu ve kaynakların daha etkin kullanımını sağlamaktadır. Ayrıca, yabancı rekabetin getireceği avantajlar ve uluslararası piyasalara açılımın doğurduğu ölçek etkisinin getirisi gibi unsurlar dolayısı ile ekonomik büyümede artış gözlemlenmektedir.³⁶⁹

Çalışmada yer alan “büyüklük” değişkeni de (kukla değişken), Model 1 (Karlılık-Verimlilik-Büyüme) ve Model 2 (Karlılık-Verimlilik) için işletmeleri performansa göre sınıflamada modele girmiştir. İşletmelerin karlılık, verimlilik ve büyüme gibi performans kriterleri incelenirken yeni teknoloji yatırımlarının ve bu yatırımlar için işletmelerin katlandığı Ar-Ge harcamalarına hak ettiği önemin verilmesi gerekmektedir. Yeni teknolojilere yapılan yatırımlar işletmelerin üretim maliyetlerini düşürdüğünden karlılığın artmasında önemli bir faktör haline gelmektedir. Araştırmalar satışlardaki artış veya karlılık gibi performans göstergeleri ile Ar-Ge harcamaları arasında güçlü ilişkiler olduğunu ortaya koymaktadır³⁷⁰. Yapılan çalışmalarda pazar payları yüksek olan işletmelerin karlılık düzeyinin de yüksek olduğu sonucuna varılmıştır³⁷¹. Büyüklük değişkeni işletmelerin ekonomiye yaptığı katkı olarak ele alındığında performanslarını etkileyen hatta arttırıcı bir değişken olarak öne çıkmaktadır.

İşletmelerin finansal performanslarını etkileyen faktörlerin araştırılmasına yönelik geniş bir literatür bulunmaktadır. Yapılan araştırmalar incelendikçe birçok çalışmada birbiri ile tutarlı sonuçlar elde edilirken bazılarının ise tam tersi sonuçlara ulaştığı

³⁶⁸ Bayfuss, 1987, p.175.

³⁶⁹ Demetrios Moschos, “Export Expansion, Growth and the Level of Economic Development: An Empirical Analysis”, **Journal of Development Economics**, Vol: 30 Issue 1, 1989, p. 93.

Epaminondas Panas, George Wamvoukas, “Further Evidence on The Export-Led Growth Hypothesis”, **Applied Economics Letters**, 9, 2002, p.731.

³⁷⁰ Graham K. Morbey, Robert M. Reithner, “How Research and Development Affects Sales Growth, Productivity and Profitability”, **Research-Technology Management**, Vol: 33, 3, 1990, pp. 11–15.

Henry G. Grabowski, Dennis C. Mueller, “Industrial Research and Development, Intangible Capital Stocks, and Firm Profit Rates”, **The Bell Journal of Economics**, Vol: 9, No. 2, 1978, pp. 328–343.

³⁷¹ David M. Szymanski, Sundar G. Bharadwaj, P. Rajan Varadarjan, “An Analysis of the Market Share-Profitability Relationship”, **The Journal of Marketing**, Vol: 57, No. 3, 1993, pp. 1–18.

görülmektedir. Bu çalışmalar özellikle işletmelerin sermaye yapısına (borç ve öz sermaye) ilişkin yapılan araştırmalardan oluşmaktadır. İşletmelerin borçlanma stratejilerinin performanslarını ne kadar etkilediğini ve borçlanma yöntemlerinden hangilerinin performansı daha çok arttırdığı yönünde yapılan araştırmalar finans literatüründe geniş yer bulan sermaye yapısı teorileriyle* de desteklenmektedir. Araştırmacıların mikro düzeyde işletmelerin performanslarını artırıcı yönde neler yapabilecekleri üzerinde bu kadar çok durmalarının en önemli nedeni, işletmelerin performanslarının iyi olmasının ülkelerin ekonomik etkinliğinin ölçülmesinde önemli göstergelerden biri olmasından kaynaklanmaktadır. Kayo ve Kimura (2011)³⁷² yaptıkları çalışmada ülkelerin ekonomik özelliklerinin işletmelerin borçlanma yapılarını etkilediğini belirtmişlerdir. Bunun yanı sıra işletmelerin borçlanma yapılarının ağırlıklı olarak kendi içsel özelliklerinden kaynaklandığının da altını çizmişlerdir.

İşletmelerin performanslarını ölçmek için farklı finansal oranlardan yararlanılmaktadır. Çalışmalarda genel olarak finansal performans ölçütü olarak aktif karlılığı, net karlılık, öz sermaye karlılığı gibi oranların kullanıldığı gözlenmiştir. Bu çalışmada işletmelerin performansının belirlenmesinde hem finansal hem de finansal olmayan ölçütlerden yararlanılarak çeşitli kombinasyonlarla modeller oluşturulmuş ve değişkenler incelenmiştir.

Yapılan çalışmalar incelendiğinde elde edilen sonuçların ülkelerin ekonomik gelişmişlik düzeyi, araştırılan dönem, işletmelerin faaliyet kolları (özellikle hizmet, üretim) gibi farklılıklar göstermektedir. Türkiye gibi gelişmekte olan ülkelerde faaliyet gösteren işletmelerin performansı değerlendirilirken de benzer kategoride olan ülkeler üzerine yapılan çalışmalar ele alınmıştır. Ekonomik gelişmişlik düzey farklılığı nedeniyle gelişmiş ülkeler ve gelişmekte olan ülkelerle yapılan araştırmaların sonuçlarında farklılıklar olduğu belirlenmiştir.

* Sermaye yapısı teorileri hisse senetleri fiyatlarının ve sermaye maliyetlerinin sermaye yapısı tarafından nasıl etkilendiğini ortaya koymak için geliştirilmiştir. Finans literatüründe kullanılan ve optimal sermaye yapısını açıklamaya yönelik temel teoriler dört kısımda incelenmektedir. Dengeleme, Hiyerarşi, Vekalet ve Açıklama Teorileri. (<http://eidergisi.istanbul.edu.tr/sayi11/ueis11m3.pdf>)

³⁷² Eduardo K. Kayo, Herbert Kimura, "Hierarchical Determinants of Capital Structure", **Journal of Banking & Finance**, 35, 2011, pp. 358–371.

Bütün modeller ayrı ayrı incelendiğinde finansal yapı oranlarından Kısa Vadeli Yük./Toplam Varlıklar ve Toplam Yük./Toplam Varlık oranlarının anlamlı ve negatif yönlü bir ilişkiye sahip olduğu görülmektedir. Bu durum işletmelerin uzun vadeli ve öz sermayeyle finansman yolunu tercih ettiği anlamına gelmektedir. Finansal yapısı güçlü olan işletmelerin iyi performans göstermesi beklenmektedir. 2007 yılı itibariyle de işletmelerin, özellikle öz sermayenin varlık finansmanında payının arttığı görülmektedir.³⁷³ Yapılan düzenlemelerle Türkiye’de yatırım fırsatlarını değerlendirmek isteyen yabancı yatırımcıların sağladığı fonların işletmelerin finansal yapılarının güçlenmesine katkı da bulunduğu sonucuna varılabilir.

İşletmelerin borç oranı, karlılık ya da işletme riski yüksek olduğu zaman düşük ya da negatif yönlüdür. İşletmelerin maddi olmayan duran varlıkları ile düşük borç oranı birbirleri ile ilişkilidir. Büyüme fırsatlarına sahip şirketlerin düşük borç oranına sahip olduğu görülmektedir³⁷⁴. Çalışmadaki veri seti incelendiğinde de büyüme oranının yüksek olduğu işletmelerin düşük borç oranına sahip olduğu gözlenmiştir. Aşağıda özetlenmiş olan, finansal yapı ile işletme performansı arasındaki ilişkiyi inceleyen çalışmalarda birbirinden farklı sonuçlar elde edilse de genel olarak ilişkinin negatif yönlü olduğu sonucuna varılmıştır.

Ebaid, Ibrahim El-Sayed (2009)³⁷⁵ Mısır için yaptıkları çalışmada işletme performansını öz sermaye karlılığı, aktif karlılığı ve brüt kar gibi değişkenlerle ölçerken çoklu regresyon yöntemi kullanarak sermaye yapısı ile işletme performansı arasında herhangi bir ilişki belirleyememişlerdir. Booth ve diğerleri (2001)³⁷⁶ ve Wiwattanakantang (1999)³⁷⁷ gelişmekte olan ülkeler için yaptıkları çalışmalarda

³⁷³ **İstanbul Sanayi Odası Dergisi**, “Türkiye’nin 500 Büyük Sanayi Kuruluşu 2007, Özel Sayı, Sayı 509, Ağustos 2008, s. 165.

³⁷⁴ Stewart C. Myers, “Capital Structure”, **Journal of Economic Perspectives**, Vol:15, Number 2, Spring 2001, pp. 81–102.

³⁷⁵ Ibrahim El-Sayed Ebaid, “The Impact of Capital-Structure Choice on Firm Performance: Empirical Evidence from Egypt”, **Journal of Risk Finance**, Vol: 10, Issue 5, 2009, pp. 477–487.

³⁷⁶ Laurence Booth, Varouj Aivazian, Aslı Demircug-Kunt, Vojislav Maksimovic, “Capital Structures in Developing Countries”, **The Journal of Finance**, The Journal of Finance, Vol: LVI, No. 1, February 2001, pp. 87–130.

³⁷⁷ Yupana Wiwattanakantang, “An Empirical Study on the Determinants of the Capital Structure of Thai Firms”, **Pacific-Basin Finance Journal**, Vol: 7, 1999, pp. 371–403.

finansal yapı (kaldıraç) oranları ile karlılık arasında negatif yönlü bir ilişki olduğu sonucuna varmışlardır.

Friend ve Lang (1988)³⁷⁸, Titman ve Wessel (1988)³⁷⁹, Kester (1986)³⁸⁰ ve Myers (2001)³⁸¹ gelişmiş ülkeler için finansal yapı oranları ile karlılık arasında negatif yönlü bir ilişki olduğu sonucuna varmışlardır. Long ve Malitz (1985)³⁸² ile Wald (1999)³⁸³ ise bu ilişkinin pozitif yönlü olduğunu saptayarak literatürde yapılan çalışmaların farklı sonuçlar elde ettiğini göstermişlerdir.

Adekunle ve Sunday (2010)³⁸⁴ Nijerya için 30 firma ile yaptıkları çalışmada borç oranının karlılıkla negatif yönlü ilişkisi olduğunu belirlemişlerdir. Chhibber ve Majumdar (1997)³⁸⁵ ile Rao, M-Yahyaee ve Syed (2007)³⁸⁶ ise yaptıkları çalışmalarda finansal kaldıraç ile işletme performansı arasında negatif bir ilişki elde etmişlerdir. Bu çalışmalarda ayrıca finansal performansın likidite üzerindeki etkisinin de negatif olduğu vurgulanmıştır.

Finansal yapı oranlarının performansa etkisi incelendikten sonra performansı etkileyen bir diğer değişken olan likidite oranları da çalışmada Model 3 (Karlılık-Büyüme), Model 4 (Karlılık-Piyasa) ve Model 5'e (Verimlilik-Büyüme) ters işaretli olarak girmiştir. Eriotis (2007)³⁸⁷ yapmış olduğu çalışmada işletmenin borçlanma

³⁷⁸ Irwin Friend, Larry H. P. Lang, "An Empirical Test of the Impact of Managerial Self-Interest on Corporate Capital Structure", **The Journal of Finance**, Vol: WLIII, No. 2, June 1988, pp. 271–281.

³⁷⁹ Sheridan Titman, Roberto Wessels, "The Determinants of Capital Structure Choice", **The Journal of Finance**, Vol: XLIII, No. 1, March 1988, pp. 1–19.

³⁸⁰ W. Carl Kester, "Capital and Ownership Structure: A Comparison of United States and Japanese Manufacturing Corporations", **Financial Management in Japan**, Spring 1986, pp. 5–16.

³⁸¹ Stewart C. Myers, **a.g.e.**, p. 83.

³⁸² Michael S. Long, Ilen B. Malitz, "Investment Patterns and Financial Leverage", **National Bureau of Economic Research**, Ed. Benjamin M. Friedman, Chapter 9, 1985, pp. 325–352.

³⁸³ John K. Wald, "How Firm Characteristics Affect Capital Structure: An International Comparison", **The Journal of Financial Research**, Vol: XXII, No. 2, Summer 1999, pp. 161–187.

³⁸⁴ Adekunle A. Onaolapo, Sunday O. Kajola, "Capital Structure and Firm Performance: Evidence from Nigeria", **European Journal of Economics, Finance and Administrative Sciences**, Issue 25, 2010, pp. 70–82.

³⁸⁵ Pradeep K. Chhibber, Sumit K. Majumdar, "Foreign Ownership and Profitability: Property Rights, Strategic Control and Corporate Performance in Indian Industry", **The William Davidson Institute**, Working Paper Number 64, April 1997, pp. 1–39.

³⁸⁶ Narendar V. Rao, Khamis Hamed Mohamed Al-Yahyaee, Lateef A. M. Syed, "Capital Structure and Financial Performance: Evidence from Oman", **Indian Journal of Economics and Business**, June 2007.

³⁸⁷ Nikolaos Eriotis, "How Firm Characteristics Affect Capital Structure: An Empirical Study", **Managerial Finance**, Vol: 33 No. 5, 2007, pp. 321–331.

oranı ile büyüme oranı, likidite oranı ve faiz karşılama oranları arasında negatif bir ilişki olduğunu belirtirken işletmelerin büyüklüğünün borçlanma oranı ile pozitif ilişkisi olduğunu ortaya koymuştur.

İşletme yöneticilerinin amaçlarında birisi varlık devir hızını (toplam varlık devir hızını) arttırmaktır. Bu değer artırılmasında toplam varlıklara göre satış gelirlerinin maksimizasyonu önemlidir. Varlık devir hızı (finansal verimlilik) arttırılmaksızın elde edilen kar artışı olumlu bir gelişme olarak değerlendirilemez. Bunun nedeni karın hesaplanması sırasında stok değerlendirme yöntemleri, finansal tabloların hazırlanması sırasında genel kabul görmüş muhasebe ilkelerine tam olarak uyulup uyulmadığı gibi konulardaki belirsizlikler işletmeler için elde edilen kar değerini etkilemektedir.³⁸⁸

Bir ekonomide imalat ve hizmet sektöründe varlık devir hızlarını arttırmaksızın sorunlara çözüm üretmek zorlaşmaktadır. Bu bağlamda işletmelerin gerçek anlamda başarılı olması finansman gereksinimlerini ve giderlerini azaltmak için kar marjları negatif olmaması koşuluyla devir hızlarını arttırmalarına bağlı olmaktadır.³⁸⁹ Çalışmada da Varlık Devir Hızı oranı hem Model 5 (Verimlilik-Büyüme) hem de Model 3'te (Karlılık-Büyüme) yer almaktadır. Ayrıca Model 3'te Hazır Değerler Devir Hızı oranı negatif işaretli olarak bulunmaktadır. İşletmelerin sağladığı rekabet avantajı satış hacimlerini yükseltmiş ve Satışlar/Toplam Varlık oranını Model 3 (Karlılık-Büyüme) ve Model 5'te (Verimlilik-Büyüme) önemli bir bağımsız değişken haline getirmiştir. İşletmelerin 2007 yılında artan satışlarının ve ekonomik istikrarın, ticari alacakların vadelerini etkilediği ve satış politikalarının esnek bir yapı kazanmasında, müşterilerinin taleplerine göre vadeli satışlarının vadelerinin uzamasında önemli bir etken olduğu düşünülmektedir. Varlık devir hızı bir işletmenin karlılığını belirleyen önemli faktörlerden biridir. Bir işletme diğer koşullar aynı kalmak üzere, varlık devir hızını arttırmak yoluyla öz sermaye karlılığını yükseltebilir. Faaliyet etkinliği göstergesi olarak kullanılan varlık devir hızı Model 3

³⁸⁸ Hüseyin Yılmaz, "Finansal Verimlilik Artışının Finans Maliyetlerine Etkisini Belirlemeye Yönelik Bir İnceleme", **V. Ulusal Üretim Araştırmaları Sempozyumu**, İstanbul Ticaret Üniversitesi, 25-27 Kasım 2005, s. 484.

³⁸⁹ **A.e.**, s.484.

ve Model 5 için anlamlı bir değişken olarak analize girmiştir. Varlık devir hızı ile performans göstergeleri arasındaki pozitif ilişki işletmelerin yatırımlarının 2007 yılında artması nedeniyle ortaya çıkmıştır.

Son olarak, işletme performansını değerlendirmek amacıyla yapılan temel analizlerde* kullanılan “fiyat-kazanç” değerinin “aktif karlılığı” ile birlikte modellenmeye çalışıldığı Model 4’te, pazar payı değişkeninin de pozitif bir etki yarattığı gözlenmiştir. Demir ve Kocabıyık (2008)³⁹⁰ yaptıkları çalışmada bireysel yatırımcıların hisse senedi yatırımlarını yaparken dikkat ettikleri faktörleri anket yolu ile analiz etmişlerdir. Yatırımcıların işletmelere yatırım yaparken işletmelerin üretim teknolojisi, pazar payı, esas faaliyet karı, dönem net karı, tanınmışlık, kar payı, işlem hacmi gibi faktörlerden etkilendikleri ortaya koymuşlardır. Çalışmada performans göstergelerinden biri olarak ele alınan fiyat-kazanç oranının varlığının, pazar payının açıklayıcı değişken olarak Model 4’te yer almasına neden olduğunu düşündürmektedir.

Çalışmanın sonuçlarında elde edilen Model 4 incelendiğinde pazar payının açıklayıcı değişken olarak devreye girmesi, piyasa göstergelerinden fiyat-kazanç oranında yani sadece Model 4’te bulunmasını açıklayabilir.

Özellikle Türkiye gibi gelişmekte olan ülkelerde yatırımcılar, ortaklar ve işletme yöneticileri açısından işletmelerin performanslarının sürekliliğinin sağlanmasının önemi artmakta ve yapılan araştırmalarla da desteklenmektedir. Hem işletmeler açısından yatırımcı ve ortakların kararlarında, hem de ülke ekonomilerinin verimliliğinin ve etkinliğinin belirlenmesinde işletme performanslarını etkileyen faktörlerin en iyi şekilde belirlenmesi her geçen yıl daha da önem kazanmaktadır.

Bu çalışmada 2007 yılına ait imalat sanayi sektöründe yer alan işletmelerin Lojistik Regresyon ve Diskriminant Analizi kullanılarak performanslarını etkileyen faktörler belirlenmeye çalışılmıştır. Sonuçların benzer çalışmalar ve teori ile tutarlı olduğu,

* “Temel Analiz: Halka açıklanan bilgilerden yararlanarak bir firmanın değerinin araştırılmasıdır.” Bkz. Mehmet Baha Karan, **a.g.e.**, s. 460.

³⁹⁰ Yusuf Demir, Turan Kocabıyık, “Türkiye’de Kurumsal Yatırımcılar ve Kurumsal Yatırımcıların Portföy Oluşturulurken Gösterdikleri Davranış Şekillerine İlişkin Bir Araştırma”, **Marmara Üniversitesi İ.İ.B.F. Dergisi**, Cilt. XXIV, Sayı. 1, 2008, ss. 181–197.

iřletmelerin performanslarının finansal yapı, faaliyet, sermaye yoğunluęu ve likidite oranları ile daha iyi açıklandığı gözlenmiştir. Bunun yanı sıra ülkelerin ekonomik politikalarıyla beraber iřletmelerin performansı deęerlendirildiğinde yabancı sermaye ve pazar payı deęiřkenlerinin model de yer aldığı görölmektedir. Çalışma sadece 2007 yılıyla sınırlı kalmayarak başka yıllar içinde yapılarak sonuçların karşılaştırılması veya daha uzun bir dönemi içerecek şekilde Panel Lojistik Regresyon Analizi kullanılarak iřletme performansını etkileyen faktörlerin zaman içerisindeki deęişimlerini de incelemek üzere genişletilebilir.

KAYNAKÇA

- Agresti, Alan : **An Introduction to Categorical Data Analysis**, John Wiley & Sons, 1996, p. 72.
- Aitchison, J.,
Bennett, J. A. : “Polychotomous Quantal Response by Maximum Likelihood Principle”, **Biometrika**, Vol: 57, No: 2, 1970, pp. 253–262.
- Akaike, H. : “Information Theory and an Extension of the Maximum Likelihood Principle”, **Perspective in Statistics Selected Paper of Hirotugu Akaike**, Springer, 1998, p. 209.
- Akal, Z. : **İmalatçı Kamu Kuruluşlarında İşletmelerarası Toplam Performans, Verimlilik, Karlılık ve Maliyet Karşılaştırması**, Ankara, MPM: 538, 1994, s. 53.
- Akdoğan, N.,
Tenker, N. : **Finansal Tablolar ve Mali Analiz Teknikleri**, 4. Baskı İstanbul, 1994, s. 452.
- Akgüç, Ö. : **Mali Tablolar Analizi**, Avcıol Basım-Yayın, İstanbul, Genişletilmiş 9. Bası, 1995, ss. 363–364.
- Akgüç, Ö. : **Finansal Yönetim**, Avcıol Basım, Yenilenmiş 7. Baskı, İstanbul 1998, s. 71.
- Alam, M. Khorshed,
Rao, M. Bhaskara,
Cheng, Fu-Chih : “Sample Size Determination in Logistic Regression”, **The Indian Journal of Statistics**, Vol: 72-B, 2010, pp. 58–75.

- Albayrak, A. S. : **Uygulamalı Çok Değişkenli İstatistik Teknikler**, Asil Yayın Dağıtım Ltd. Şti., 2006, s. 65.
- Aldrich, John H.,
Nelson, Forrest D. : **Linear Probability Logit and Probit Models**, USA, Sage Publication, 1984, p. 41.
- Andersen, Erling B. : **Introduction to the Statistical Analysis of Categorical Data**, Springer, 1997, p. 30, 177.
- Aslan, M.H.,
Yıldırım, F. : “Endüstri Mühendisliği Yaşam Süresi Analizlerinde Yeniden Örnekleme Yöntemlerinin Kullanılması”, **2. Mühendislik ve Teknoloji Sempozyumu, 30 Nisan- 1 Mayıs 2009**, Çankaya Üniversitesi, s. 169.
- Bayyurt, N. : “İşletmelerde Performans Değerlendirmenin Önemi ve Performans Göstergeleri Arasındaki İlişkiler”, **Sosyal Siyaset Konferansları Dergisi**, Sayı 53, 2007/2 Prof. Dr. Haşmet Başar’a Armağan Özel Sayı, s. 580.
- Bilgin, C.,
Şahbaz, A. : “Türkiye’de Büyüme ve İhracat Arasındaki Nedensellik İlişkisi”, **Gaziantep Üniversitesi Sosyal Bilimler Dergisi**, 2009, (8).1, ss. 177–198.
- Bilodeau, M.,
Brenner, D. : **Theory of Multivariate Statistics**, Springer, 1999, p. 170.
- Buse, A. : “The Likelihood Ratio, Wald and Lagrange Multiplier Tests: An Expository Note”, **The American Statistician**, Vol: 36, No: 3, 1982, pp. 153–157.

- Bolat, Bilge A. : **Firma Performansını Etkileyen Faktörlerin Yapısal Eşitlik Modeli ile İncelenmesi**, Basılmış Doktora Tezi, İstanbul, 2010, ss. 32–39.
- Booth, Laurence, Aivazian, Varouj, Demirguc-Kunt, A., Maksimovic, V. : “Capital Structures in Developing Countries”, **The Journal of Finance**, The Journal of Finance, Vol: LVI, No. 1, February 2001, pp. 87–130.
- Borooah, Vani K. : **Logit and Probit Ordered and Multinomial Models**, Sage Publications, 2002, p. 51.
- Chatterjee, Samprit, Hadi, Ali S., Price, Bertram : **Regression Analysis by Example**, Third Edition John Wiley & Sons, 2000, p.36.
- Cheng, Simon, J. Long, Scott : “Testing for IIA in the Multinomial Logit Model”, **Sociological Methods & Research**, Vol:35, Number 4, May 2007, pp. 583–600.
- Chhibber, Pradeep K., Majumdar, Sumit K. : “Foreign Ownership and Profitability: Property Rights, Strategic Control and Corporate Performance in Indian Industry”, **The William Davidson Institute**, Working Paper Number 64, April 1997, pp. 1–39.
- Christensen, Ronald : **Log-Linear Models and Logistic Regression**, 2nd Edition, Springer, 1997, p. 297.

- Çiftçioğlu, S.,
Karaaslan, C.,
Demir, H. : “Saving, Investment, Exports and Economic Growth in Turkey”, **Review of Social, Economic&Business**, Vol. 3–4, pp. 20–30.
- Copas J. B. : “Plotting p Against”, **Applied Statistics**, Vol: 32, 1983, pp. 25–31.
- Cox, D. R.,
Snell, E. J. : **Analysis of Binary Data**, Second Edition, Chapman & Hall, 1989, p. 21.
- Cramer, J. : “Predictive Performance of the Binary Logit Model in Unbalanced Samples”, **Journal of the Royal Statistical Society**, Series D, Vol: 88, 1990, pp. 85–94.
- Crown, William H. : **Statistical Models for the Social and Behavioral Sciences**, Praeger, p. 135.
- Demaris, Alfred : **Logit Modeling**, Sage University Paper Series on Quantitative Applications in the Social Sciences, 1992, p. 76.
- Demir, Yusuf,
Kocabıyık, Turan : “Türkiye’de Kurumsal Yatırımcılar ve Kurumsal Yatırımcıların Portföy Oluşturulurken Gösterdikleri Davranış Şekillerine İlişkin Bir Araştırma”, **Marmara Üniversitesi İ.İ.B.F. Dergisi**, Cilt. XXIV, Sayı. 1, 2008, ss. 181–197.
- Dinçer, Ö. : **Stratejik Yönetim ve İşletme Politikası**, İstanbul, 1992, s. 116.

- Dirican, A. : “Tanı Testi Performanslarının Değerlendirilmesi ve Kıyaslanması”, **Cerrahpaşa Tıp Dergisi**, Cilt 32, Sayı 1, 2001, ss. 25–30.
- Ebaid, I. El-Sayed : “The Impact of Capital-Structure Choice on Firm Performance: Empirical Evidence from Egypt”, **Journal of Risk Finance**, Vol. 10, Issue 5, 2009, pp. 477–487.
- Efron, Bradley : “Regression and ANOVA with Zero-One Data: Measures of Residual Variation”, **Journal of the American Statistical Association**, Vol: 73, 1978, pp. 113–121.
- Ellem, A. B. : **Multivariate Analysis** (Stat501), January 10, 2008 (Çevirimiçi) <http://mcs.une.edu.au/~stat501/UnitNotes/unitnotes08.pdf>, 25 Ağustos 2011, p. 28.
- Eriotis, Nikolaos : “How Firm Characteristics Affect Capital Structure: An Empirical Study”, **Managerial Finance**, Vol: 33 No. 5, 2007, pp. 321–331.
- Fisher, R. A. : “The Use of Multiple Measurements in Taxonomic Problems”, **Annals of Eugenics**, 7(2), 1936, pp. 466–475.
- Freund, John E. : **Mathematical Statistics**, Fifth Edition, Prentice Hall International Editions, p. 224

- Friend, Irwin,
Lang, Larry H. P. : “An Empirical Test of the Impact of Managerial Self-Interest on Corporate Capital Structure”, **The Journal of Finance**, Vol: WLIII, No. 2, June 1988, pp. 271–281.
- Fry, Tim R.,
Harris, Mark N. : “A Monte Carlo Study of Tests for the Independence of Irrelevant Alternatives Property”, **Transportation Research Part B: Methodological**, Vol: 30, 1996, pp. 19–30.
- Gilbert, Ethel S. : “The Effect of Unequal Variance-Covariance Matrices on Fisher’s Linear Discriminant Function”, **Biometrics**, 1969, 25, pp. 505–515.
- Gill, Jeff : **Generalized Linear Models**, USA Sage University Paper, 2001, p. 2.
- Grabowski, Henry G.,
Mueller, Dennis C. : “Industrial Research and Development, Intangible Capital Stocks, and Firm Profit Rates”, **The Bell Journal of Economics**, Vol: 9, No. 2, 1978, pp. 328–343.
- Griliches, Z. : “Productivity, R&D and Basic Research at the Firm Level in the 1970’s”, **The American Economic Review**, 1986, pp. 141–153.
- Gujarati, Damodar N. : **Temel Ekonometri**, Literatür Yayıncılık, 1999.
- Hagle, T. M.,
Mitchell II, G. E. : “Goodness-of-fit Measures for Probit and Logit”, **American Journal of Political Science**, Vol:36, No:3, 1992, pp. 762–784.

- Hair, Joseph F.,
Black, William C.,
Babin, Barry J., : **Multivariate Data Analysis**, Sixth Edition, Pearson,
Anderson, Rolph E., : 2006, p. 411.
Tatham, Ronald L.
- Hauck, W. W., : “Wald’s Test as Applied to Hypotheses in Logit
Donner, A. : Analysis”, **American Statistical Association**, Vol: 72,
No: 360, 1977, pp. 851–853.
- Hausman, J., : “Specification Tests for the Multinomial Logit Model”,
McFadden, D. : **Econometrica**, Vol: 52, No: 5, 1984, pp. 1219–1240.
- Hausman, Jerry A., : “A Conditional Probit Model for Qualitative Wise
David A. : Choice: Discrete Decision recognizing Interdependence
and Heterogeneous Preferences”, **Econometrica**, Vol:
46, No: 2, 1978, pp. 403–426.
- Hoel, P. G., : “A Solution to the Problem of Optimum Classification”,
Peterson, R. P. : **The Annals of Mathematical Statistics**, 1949, Vol: 20,
No. 3 (Sep.), p. 433.
- Hosmer, David W., : “Best Subsets Logistic Regression”, **Biometrics**, Vol:
Jovanovic, B., : 45, 1989, pp. 1265–1270.
Lemeshow, S.
- Hosmer, David W., : **Applied Logistic Regression**, 2nd Edition, USA, John
Lemeshow, Stanley : Wiley & Sons, 2000, p.1.

- Hsieh, F. Y.,
Bloch, Daniel A.,
Larsen, Michael D. : “A Simple Method of Sample Size Calculation for Linear and Logistic Regression”, **Statistics in Medicine**, Vol: 17, 1998, pp. 1623–1634.
- Huberty, Carl J. : **Applied Discriminant Analysis**, John Wiley & Sons INC., 1994, p. 43.
- Johnson, Richard A.,
Wichern Dean W. : **Applied Multivariate Statistical Analysis**, Third Edition, Prentice Hall, New Jersey, 1992, p. 133.
- Karagöz, M.,
Demirgil, H. : “Firma Büyüme Performansını Etkileyen Faktörler”, **Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi**, 2009, C. 14, S.2, s. 48.
- Karan, M. B. : **Yatırım Analizi ve Portföy Yönetimi**, Gazi Kitapevi, Ankara, 2001, s. 356.
- Kay, R.,
Little, S. : “Transformation of the Explanatory Variables in the Logistic Regression Model for Binary Data”, **Biometrika**, Vol: 74, 1987, pp. 495–501.
- Kayo, Eduardo K.,
Kimura, Herbert : “Hierarchical Determinants of Capital Structure”, **Journal of Banking & Finance**, 35, 2011, pp. 358–371.
- Kester, W. Carl : “Capital and Ownership Structure: A Comparison of United States and Japanese Manufacturing Corporations”, **Financial Management in Japan**, Spring 1986, pp. 5–16.

- Kurnaz, N. : “Finansal (Mali) Tablolar Analizi”, (Çevrimiçi)
: <http://www.niyazikurnaz.net>, 12 Temmuz 2011.
- Lachenbruch, P. A. : “An Almost Unbiased Method of Obtaining Confidence
: Intervals for the Probability of Misclassification in
Discriminant Analysis”, **Biometrics**, Vol. 23, No. 4,
1967, pp. 639–645.
- Lachenbruch, Peter A.,
Sneeringer, C., : “Robustness of the Linear and Quadratic Discriminant
: Function to Certain Types of Non-Normality”,
Revo, L. T. **Communications in Statistics**, 1973, Vol: 1, pp. 39–
56.
- Liao, Tim Futing : **Interpreting Probability Models Logit Probit and
: Other Generalized Linear Models**, Sage University
Paper, 1994, pp. 1–2.
- Long, J. Scott : **Regression Models for Categorical and Limited
: Dependent Variables**, USA, Sage Publication, 1997, p.
36.
- Long, J. Scott,
Cheng, Simon : **The Handbook of Data Analysis**, Edited by Hardy and
: Bryman pp. 261–262.
- Long, Michael S.,
Malitz, Ilen B. : “Investment Patterns and Financial Leverage”, **National
: Bureau of Economic Research**, Ed. Benjamin M.
Friedman, Chapter 9, 1985, pp. 325–352.

- Longford, Nicholas T. : **Handbook of Statistical Modeling for the Social and Behavioral Sciences**, Plenum Press, Edited by Gerhard Arminger, Clifford C. Clogg and Michael E. Sobel, 1995, p. 551.
- Maddala, G. S. : **Limited-Dependent and Qualitative Variables in Econometrics**, Cambridge University Press, 1983, p.39.
- Magee, Lonnie : “ R^2 Measures Based on Wald and Likelihood Ratio Joint Significance Tests”, **The American Statistician**, Vol: 44, 1990, pp. 250–253.
- Marks, S.,
Dunn, O. J. : “Discriminant Functions When Covariance Matrices are Unequal”, **Journal of the American Statistical Association**, 1974, 69, pp. 555–559.
- McCullagh, P. : “The Conditional Distribution of Goodness-of-fit Statistics for Discrete Data”, **Journal of the American Statistical Association**, Vol: 81, No: 393, 1986, pp. 104–107.
- McCullagh, P.,
Nelder, J. A. : **Generalized Linear Models**, 2nd Edition, Chapman & Hall, 1989, p. 30.
- McFadden, D. : “Conditional Logit Analysis of Quantitative Choice Behavior” Edited by Zarembka, P., **Frontiers in Econometrics**, New York Academic Press, 1973, p. 121.

- McFadden, Daniel,
Train, Kenneth,
Tye William B. : “An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model”, **Transportation Research Board Record**, 637, 1981, pp. 39–46.
- McKay, R. J.,
Campbell, N. A. : **Variable Selection Techniques in Discriminant Analysis I. Description**, British Journal of Mathematics Statistical Psychology, 35, 1982a, pp. 1–29.
- McKelvey, R. D.,
Zavoina, W. : “A Statistical Model for the Analysis of Ordinal Level Dependent Variable”, **Journal of Mathematical Sociology**, Vol: 4, 1975, pp. 103–120.
- Menard, Scott : **Applied Logistic Regression Analysis**, 2nd Edition, Sage Publications, 2002, p.91.
- Mickey J.,
Greenland S. : “A Study of the Impact of Confounder-selection Criteria on Effect Estimation”, **American Journal of Epidemiology**, 129, pp. 125–137.
- Moschos, Demetrios : “Export Expansion, Growth and the Level of Economic Development: An Empirical Analysis”, **Journal of Development Economics**, Vol: 30 Issue 1, 1989, p. 93.
- Morbey, Graham K.,
Reithner, Robert M. : “How Research and Development Affects Sales Growth, Productivity and Profitability”, **Research-Technology Management**, Vol: 33, 3, 1990, pp. 11–15.

- Mucuk, İ. : **Modern İşletmecilik**, 7. Baskı, Türkmen Kitapevi, İstanbul, 1997, ss. 98–99.
- Myers, Stewart C. : “Capital Structure”, **Journal of Economic Perspectives**, Vol: 15, Number 2, Spring 2001, pp. 81–102.
- Nelder, J. A.,
Wedderburn, R. W. M. : “Generalized Linear Models”, **Journal of the Royal Statistical Society Series A**, Vol: 135, No: 3, 1972, pp. 370–384.
- Newbold, Paul,
Carlson, William L.,
Thorne, Betty : **Statistics for Business and Economics**, Sixth edition, Prentice Hall 2007, p.539.
- Obuchowski, N. A. : “ROC Analysis”, **American Journal of Roentgenology**, 184, 2005, pp. 364–372.
- O’Connell, Ann A. : **Logistic Regression Models for Ordinal Response Variables**, Sage Publication, 2006, pp. 30–31.
- Onaolapo, Adekunle A.,
Kajola, Sunday O. : “Capital Structure and Firm Performance: Evidence from Nigeria”, **European Journal of Economics, Finance and Administrative Sciences**, Issue 25, 2010, pp. 70–82.
- Orhunbilge, Neyran : **Tanımsal İstatistik Olasılık ve Olasılık Dağılımları**, Avcıol Basım, 2000, s. 192.

- Orhunbilge, Neyran : **Uygulamalı Regresyon ve Korelasyon Analizi**, İstanbul, 2002, s. 241.
- Orhunbilge, Neyran : **Çok Değişkenli İstatistik Yöntemler**, İstanbul Üniversitesi Basım ve Yayımevi Müdürlüğü, İstanbul, 2010, s. 191.
- Öğüdücü, Ş. G. : “Veri Madenciliği Farklı Sınıflandırma Yöntemleri”, (Çevirimiçi) <http://web.itu.edu.tr/~sgunduz/courses/verimaden> 15 Ağustos 2011.
- Özdamar, Kazım : **Paket Programlar ile İstatistiksel Veri Analizi**, Cilt 2, Yenilenmiş 5. Baskı, 2004, s. 355.
- Pampel, Fred C. : **Logistic Regression A Primer**, USA, Sage Publication, 2000, pp. 54–55.
- Panas, Epaminondas, Wamvoukas, George : “Further Evidence on The Export-Led Growth Hypothesis”, **Applied Economics Letters**, 9, 2002, p.731.
- Pandya, Anil M., Rao, Narender V. : “Deversification and Firm Performance an Empirical Evaluation”, **Journal of Financial and Strategic Decisions**, V.II. N. 2, 1998, p. 67.

- Pedduzi, P. N.,
 Concato, J.,
 Kemper, E.,
 Holford, T. R.,
 Feinstein, A.,
 : “A Simulation Study of the Number of Events per
 Variable in Logistic Regression Analysis”, **Journal of
 Clinical Epidemiology**, Vol: 49, 1996, pp. 1373–1379.
- Pratt, J. W.
 : “Concavity of the Log-Likelihood”, **Journal of the
 American Statistical Association**, Vol: 76, 1981, pp.
 103–106.
- Pregibon, D.
 : “Logistic Regression Diagnostic”, **Annals of Statistics**,
 Vol: 9, 1981, pp. 705–724.
- Raftery, Adrian E.
 : “Bayesian Model Selection Social Research”,
Sociological Methodoloji, Vol: 25, 1995, pp. 111–165.
- Rao, C. R.
 : “The Utilization of Multiple Measurements in Problem
 of Biological Classification”, **Journal of the Royal
 Statistical Society**, Series B (Methodological), 1948,
 Vol: 10, No. 2, pp. 159–203.
- Narendar V. Rao,
 Al-Yahyaee, K. H. M.,
 Syed, Lateef A. M.
 : “Capital Structure and Financial Performance: Evidence
 from Oman”, **Indian Journal of Economics and
 Business**, June 2007.
- Rencher, Alvin C.,
 Schaalje, G. Bruce
 : **Linear Models in Statistics**, A John Willey & Sons,
 2008, p. 514.

- Rothenberg, T. J. : “Hypothesis Testing in Linear Models When the Error Covariance Matrix is Non-scalar”, **Econometrica**, 1984, Vol: 52, pp. 827–842.
- Royston, P.,
Altman, D. G. : “Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling”, **Applied Statistics**, Vol: 43, 1994, pp. 429–467.
- Sauerbrei, W.,
Meier-Hirmer, C.,
Benner, A.,
Royston, P. : “Multivariable Regression Model Building by Using Fractional Polynomials: Description of SAS, STATA and R Programs”, **Computational Statistics & Data Analysis**, Vol: 50, 2006, p. 3466.
- Schmidt, P.,
Strauss, R. P. : “The Prediction of Occupation Using Multiple Logit Models”, **International Economic Review**, Vol: 16, No: 2, 1975, pp. 471–486.
- Sharma, Subhash : **Applied Multivariate Techniques**, John Wiley & Sons, Inc., 1996, p. 241.
- Silva, A. Pedro Durate,
Stam, Antonie : **Reading and Understanding Multivariate Statistics**, Edited by Laurence G. Grimm, Paul R. Yarnold, American Psychological Association Washington, DC., 1994, p. 277.
- Small, Kenneth A.,
Hsiao, C. : “Multinomial Logit Specification Tests”, **International Economic Review**, Vol: 26, No: 3, 1985, pp. 619–627.

- Smith, C. A. B. : “Some Examples of Discrimination”, **Annals of Eugenics**, 1947, 13, pp. 272–282.
- (Çevirimiçi)
- SPSS : http://support.spss.com/ProductsExt/SPSS/Documentation/Statistics/algorithms/14.0/app14_boxsm.pdf, 10 Eylül 2011.
- Suiçmez, H. : “Türkiye’de Ekonomik Büyüme ve Verimlilik Artış Performansı Işığında Nasıl bir Kalkınma Politikası Benimsenmeli”, **Türkiye İşveren Sendikaları Konfederasyonu**, Ocak, 2007.
(Çevirimiçi) [http://www.tisk.org.tr/isveren_sayfa.asp?yazi_id=1618 &id=84](http://www.tisk.org.tr/isveren_sayfa.asp?yazi_id=1618&id=84), 11 Haziran 2011
- Szymanski, David M., Bharadwaj, Sundar G., Varadarjan, P. Rajan : “An Analysis of the Market Share-Profitability Relationship”, **The Journal of Marketing**, Vol: 57, No. 3, 1993, pp. 1–18.
- Tacq, Jacques : **Multivariate Analysis Techniques in Social Science Research**, Sage Publication, 1997, p. 248.
- Tatlıdil, Hüseyin : **Uygulamalı Çok Değişkenli İstatistiksel Analiz**, Cem Ofset Ltd. Şti., Ankara, 1996, s.293.
- Tekin, M. : “Sanayi İşletmelerinde Verimlilik ve Önemi”, **Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi**, Sayı: 1, 1992, s. 170.

- Tezeren, A. : **İmalat Sanayinde Verimliliği Etkileyen Faktörler**,
: 319 Nolu Milli Prodüktivite Merkezi Yayınları, Ankara,
1985.
- Titman, Sheridan,
Wessels, Roberto : “The Determinants of Capital Structure Choice”, **The
Journal of Finance**, Vol: XLIII, No. 1, March 1988,
pp. 1–19.
- Tükenmez, M., vd. : **Finansal Yönetim**, Vizyon Eğitim ve Danışmanlık Ltd.
Şti., İzmir 1999, ss. 413–414.
- Wald, John K. : “How Firm Characteristics Affect Capital Structure: An
International Comparison”, **The Journal of Financial
Research**, Vol: XXII, No. 2, Summer 1999, pp. 161–
187.
- Wehrens R.,
Putter, H.,
Buydens, L.M.C. : “The Bootstrap: A Tutorial”, **Chemometrics and
Intelligent Laboratory Systems**, Vol: 54, 2000, pp.
35–52.
- Whittemore, Alice S. : “Sample Size for Logistic Regression with Small
Response Probability”, **Journal of the American
Statistical Association**, Vol: 76, 1981, pp. 27–32.
- Wilks, S.S. : **Certain Generalizations in the Analysis of Variance**,
: *Biometrika*, Vol. 24, No. 3/4, 1932, pp. 471–494.

Wiwattanakantang, Y. : “An Empirical Study on the Determinants of the Capital Structure of Thai Firms”, **Pacific-Basin Finance Journal**, Vol: 7, 1999, pp. 371–403.

Wright, Raymond E. : **Reading and Understanding Multivariate Statistics**, American Psychological Association, Edited By Laurence G. Grimm and Paul R. Yarnold, 1994, p. 220.

Yılmaz, Hüseyin : “Finansal Verimlilik Artışının Finans Maliyetlerine Etkisini Belirlemeye Yönelik Bir İnceleme”, **V. Ulusal Üretim Araştırmaları Sempozyumu**, İstanbul Ticaret Üniversitesi, 25–27 Kasım 2005, s. 484.

İstanbul Sanayi Odası Dergisi, “Türkiye’nin 500 Büyük Sanayi Kuruluşu 2007”, Özel Sayı, Sayı 509, Ağustos 2008, s. 165.

Sermaye Piyasası ve Borsa Temel Bilgiler Kılavuzu, “İstanbul Menkul Kıymetler Borsası Yayınları”, 17. Basım, Temmuz 2002, s.141.

Açık Gazete (Çevirimiçi) : “Yabancı Sermaye Nedir? Hazırlayan Faruk Eskiöğlü”, <http://www.acikgazete.com/ozel-dosyalar/2009/03/24/yabanci-sermaye-nedir.htm?aid=27995>, 01 Temmuz 2011

- Bölüm-1 : “Niçin İhracat”, **100 Soruda Dış Ticaret, İhracat Geliştirme Etüd Merkezi**, Ekim, 2006, (Çevirimiçi) http://www.dkib.org.tr/dosya/pratik_100soru.pdf, 10 Temmuz 2011.
- Bölüm-10 : “Teknolojik Yenilik ve Ekonomik Performans”, (Çevirimiçi) <http://www.inovasyon.org/pdf/blm10.pdf>, 01.07.2011.
- (Çevirimiçi) : <http://notoku.com/pims-analizi/> , 12.Temmuz 2011
- (Çevirimiçi) : <http://www.havvatunc.com/ticaret-acigi-buyuyor-ihracatin-ithalati-karsima-orani-kuculuyor-1107.html>, 10 Eylül 2011.
- (Çevirimiçi) : <http://www.bilgaz.net/dosyalar/OranAnalizi.pdf>, 10Ağustos 2011.
- (Çevirimiçi) : <http://www.bacegrup.com/3-4-oran-analizi.ntml>, 10Temmuz 2011.
- (Çevirimiçi) : “Yatırım Ortamı Değerlendirmesi Krizden Özel Sektör Öncülüğünde Büyümeye”. <http://www.hazine.gov.tr>, 07Temmuz 2011.

EKLER

EK.1 Karlılık Oranları

Kendall Uygunluk Katsayısı

N	8
Kendall's W ^a	.802
Chi-square	930.582
df	145
Asymp. Sig.	.000

a. Kendall's Coefficient of
Concordance

Correlations

Spearman's rho		OzSermaye	Ekonomik Rant	AktifKar	FaalKar Marj	BrutKar Marj	NetKarMarj	VOKarMarj	Faiz, Karsilama
OzSermaye	Correlation Coefficient	1.000	.823**	.907**	.743**	.462**	.852**	.840**	.747**
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000	.000	.000
	N	146	146	146	146	146	146	146	146
EkonomikRant	Correlation Coefficient	.823**	1.000	.896**	.767**	.530**	.832**	.832**	.672**
	Sig. (2-tailed)	.000	.	.000	.000	.000	.000	.000	.000
	N	146	146	146	146	146	146	146	146
AktifKar	Correlation Coefficient	.907**	.896**	1.000	.810**	.542**	.953**	.945**	.805**
	Sig. (2-tailed)	.000	.000	.	.000	.000	.000	.000	.000
	N	146	146	146	146	146	146	146	146
FaalKarMarj	Correlation Coefficient	.743**	.767**	.810**	1.000	.686**	.852**	.859**	.871**
	Sig. (2-tailed)	.000	.000	.000	.	.000	.000	.000	.000
	N	146	146	146	146	146	146	146	146
BrutKarMarj	Correlation Coefficient	.462**	.530**	.542**	.686**	1.000	.628**	.652**	.550**
	Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.000	.000
	N	146	146	146	146	146	146	146	146
NetKarMarj	Correlation Coefficient	.852**	.832**	.953**	.852**	.628**	1.000	.991**	.809**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.	.000	.000
	N	146	146	146	146	146	146	146	146
VOKarMarj	Correlation Coefficient	.840**	.832**	.945**	.859**	.652**	.991**	1.000	.816**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.	.000
	N	146	146	146	146	146	146	146	146
FaizKarsilama	Correlation Coefficient	.747**	.672**	.805**	.871**	.550**	.809**	.816**	1.000
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.
	N	146	146	146	146	146	146	146	146

** . Correlation is significant at the 0.01 level (2-tailed).

EK.2 Büyüme Oranları

Kendall Uygunluk Katsayısı

N	4
Kendall's W ^a	.501
Chi-square	290.488
df	145
Asymp. Sig.	.000

a. Kendall's Coefficient of Concordance

Spearman Sıra Korelasyon Değerleri

Spearman's rho		Satartisreel	Karartisreel	Ozsermayeartis	Varlikartis
Satartisreel	Correlation Coefficient	1.000	.208*	.280**	.374**
	Sig. (2-tailed)	.	.012	.001	.000
	N	146	146	146	146
Karartisreel	Correlation Coefficient	.208*	1.000	.275**	.327**
	Sig. (2-tailed)	.012	.	.001	.000
	N	146	146	146	146
Ozsermayeartis	Correlation Coefficient	.280**	.275**	1.000	.544**
	Sig. (2-tailed)	.001	.001	.	.000
	N	146	146	146	146
Varlikartis	Correlation Coefficient	.374**	.327**	.544**	1.000
	Sig. (2-tailed)	.000	.000	.000	.
	N	146	146	146	146

*. Correlation is significant at the 0.05 level (2-tailed).

** . Correlation is significant at the 0.01 level (2-tailed).

EK.3 Piyasa Performans Değerleri

Kendall Uygunluk Katsayısı

Test Statistics	
N	3
Kendall's W ^a	.485
Chi-square	210.915
df	145
Asymp. Sig.	.000

a. Kendall's Coefficient of Concordance

Spearman Sıra Korelasyon Değerleri

		Correlations		
Spearman's rho		Hissebaskar	Fiyatkazanc	PiyasaDefDeg
Hissebaskar	Correlation Coefficient	1.000	.319**	.111
	Sig. (2-tailed)	.	.000	.182
	N	146	146	146
Fiyatkazanc	Correlation Coefficient	.319**	1.000	.253**
	Sig. (2-tailed)	.000	.	.002
	N	146	146	146
PiyasaDefDeg	Correlation Coefficient	.111	.253**	1.000
	Sig. (2-tailed)	.182	.002	.
	N	146	146	146

** . Correlation is significant at the 0.01 level (2-tailed).

EK.4 Model 1 için Diskriminant Analizi Sonuçları

Eşit Varyanslılık Varsayımını test etmek için Enter model kullanılarak hesaplanan Box's M test sonucu,

y	Rank	Log Determinant
0	24	36,640
1	24	34,862
Pooled within-groups	24	59,410

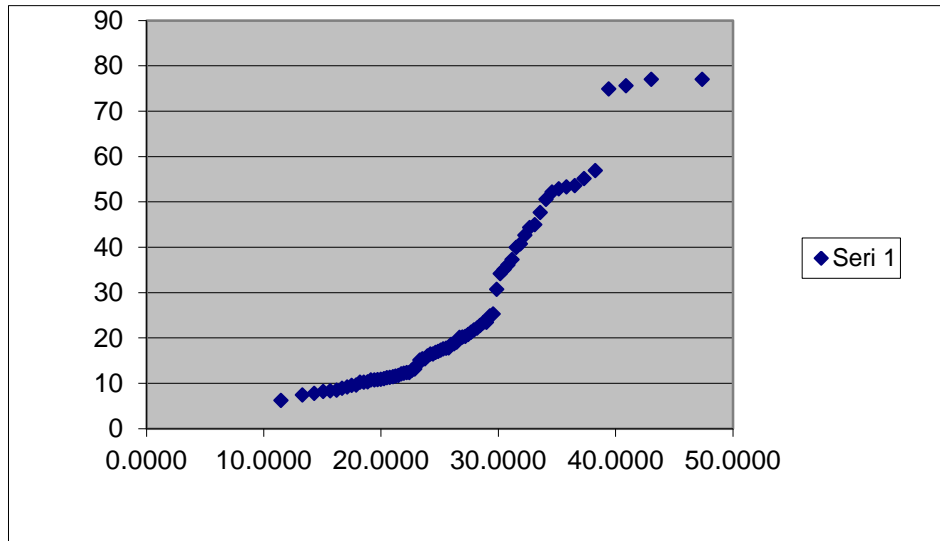
The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

Box's M	1835,091
F	Approx. 3,937
df1	300
df2	13603,759
Sig.	,000

Tests null hypothesis of equal population covariance matrices.

Çoklu Normallik Varsayımını için yapılan testler,

Sıralanmış Mahalanobis uzaklıkları ile χ^2 değerlerinin serpilme diyagramı,



Ek.4'ün Devamı

Analyze > Descriptive > Explore

Tests of Normality

y	Kolmogorov-Smirnov ^a			Shapiro-Wilk			
	Statistic	df	Sig.	Statistic	df	Sig.	
Likidite1	0	,285	32	,000	,582	32	,000
	1	,215	47	,000	,737	47	,000
Likidite2	0	,287	32	,000	,580	32	,000
	1	,207	47	,000	,700	47	,000
Likidite3	0	,308	32	,000	,520	32	,000
	1	,278	47	,000	,519	47	,000
Likidite4	0	,177	32	,012	,875	32	,002
	1	,113	47	,170	,936	47	,012
Likidite5	0	,135	32	,148	,921	32	,023
	1	,160	47	,004	,912	47	,002
FinYap1	0	,124	32	,200 [*]	,976	32	,688
	1	,108	47	,200 [*]	,968	47	,221
FinYap2	0	,084	32	,200 [*]	,975	32	,653
	1	,128	47	,051	,933	47	,009
FinYap3	0	,447	32	,000	,234	32	,000
	1	,141	47	,020	,885	47	,000
FinYap4	0	,228	32	,000	,719	32	,000
	1	,104	47	,200 [*]	,976	47	,445
FinYap5	0	,110	32	,200 [*]	,956	32	,218
	1	,091	47	,200 [*]	,966	47	,182
Faaliyet1	0	,355	32	,000	,416	32	,000
	1	,471	47	,000	,161	47	,000
Faaliyet2	0	,388	32	,000	,297	32	,000
	1	,210	47	,000	,734	47	,000
Faaliyet3	0	,453	32	,000	,225	32	,000
	1	,235	47	,000	,632	47	,000
Faaliyet4	0	,171	32	,018	,873	32	,001
	1	,132	47	,041	,919	47	,003
Faaliyet5	0	,244	32	,000	,765	32	,000
	1	,145	47	,015	,879	47	,000
Faaliyet6	0	,180	32	,010	,859	32	,001
	1	,154	47	,007	,866	47	,000

Faaliyet7	0	,229	32	,000	,785	32	,000
	1	,120	47	,086	,910	47	,002
SermayeYog1	0	,219	32	,000	,766	32	,000
	1	,197	47	,000	,760	47	,000
SermayeYog2	0	,226	32	,000	,818	32	,000
	1	,186	47	,000	,856	47	,000
İhracat	0	,404	32	,000	,399	32	,000
	1	,404	47	,000	,349	47	,000
YabSermaye	0	,512	32	,000	,404	32	,000
	1	,323	47	,000	,713	47	,000
Yenilik	0	,269	32	,000	,701	32	,000
	1	,530	47	,000	,135	47	,000
PazarPay	0	,203	32	,002	,750	32	,000
	1	,285	47	,000	,554	47	,000

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Ek.4'ün Devamı

Discriminant

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Likidite1	,978	1,759	1	77	,189
Likidite2	,977	1,836	1	77	,179
Likidite3	,943	4,697	1	77	,033
Likidite4	,967	2,631	1	77	,109
Likidite5	,920	6,668	1	77	,012
FinYap1	,820	16,956	1	77	,000
FinYap2	,815	17,437	1	77	,000
FinYap3	,974	2,032	1	77	,158
FinYap4	,974	2,071	1	77	,154
FinYap5	,979	1,675	1	77	,199
Faaliyet1	,991	,669	1	77	,416
Faaliyet2	,990	,809	1	77	,371
Faaliyet3	,993	,577	1	77	,450
Faaliyet4	1,000	,000	1	77	,999
Faaliyet5	,970	2,401	1	77	,125
Faaliyet6	,917	6,963	1	77	,010
Faaliyet7	,994	,499	1	77	,482
SermayeYog1	,861	12,480	1	77	,001
SermayeYog2	,828	16,017	1	77	,000
İhracat	,995	,349	1	77	,557
YabSermaye	,940	4,936	1	77	,029
Yenilik	,991	,669	1	77	,416
PazarPay	,966	2,726	1	77	,103
Buyukluk	,901	8,507	1	77	,005

Ek.4'ün Devamı
Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	,874 ^a	100,0	100,0	,683

a. First 1 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	,534	46,462	6	,000

Canonical Discriminant

Function Coefficients

	Function
	1
Likidite5	4,206
FinYap2	3,786
SermayeYog1	,000
SermayeYog2	,000
YabSermaye	-,015
Buyukluk	-1,444
(Constant)	-,871

Unstandardized coefficients

Classification Statistics

Prior Probabilities for Groups

y	Prior	Cases Used in Analysis	
		Unweighted	Weighted
0	,405	32	32,000
1	,595	47	47,000
Total	1,000	79	79,000

Classification Results^a

y		Predicted Group Membership		Total
		0	1	
Original	Count	0	1	
		23	9	32
		6	41	47
	%	0	1	
		71,9	28,1	100,0
		12,8	87,2	100,0

a. 81,0% of original grouped cases correctly classified.

EK.5 Model 1 için Lojistik Regresyon Analizi Sonuçları

STATA Çıktısı

```
. sw logit y Likidite1 Likidite2 Likidite3 Likidite4 Likidite5 FinYap1 FinYap2  
FinYap3 FinYap4 FinYap5 Faaliyet1 Faaliyet2 Faaliyet3 Faaliyet4 Faaliyet5  
Faaliyet6 Faaliyet7 SermayeYog1 SermayeYog2 Ihracat YabSermaye Yenilik  
PazarPay Buyukluk, forward pe(.1) pr(.2)
```

begin with empty model

```
p = 0.0005 < 0.1000 adding FinYap1  
p = 0.0032 < 0.1000 adding SermayeYog2  
p = 0.0077 < 0.1000 adding Buyukluk  
p = 0.0583 < 0.1000 adding YabSermaye  
p = 0.0989 < 0.1000 adding Likidite5  
p = 0.0427 < 0.1000 adding FinYap5  
p = 0.0349 < 0.1000 adding FinYap2  
p = 0.5408 >= 0.2000 removing FinYap1
```

```
Logistic regression                Number of obs =    79  
LR chi2(6)    =    56.84            Prob > chi2    =    0.0000  
Log likelihood = -24.904914        Pseudo R2     =    0.5330
```

```
      y |   Coef.   Std. Err.      z    P>|z|   [95% Conf. Interval]  
-----+-----  
      FinYap2 | -9.932728   3.326845   -2.99   0.003   -16.45322   -3.412232  
      SermayeYog2 | 5.45e-06   1.64e-06    3.32   0.001    2.23e-06   8.67e-06  
      Buyukluk | 3.737337   1.389887    2.69   0.007    1.013209   6.461466  
      YabSermaye | .0415742   .0161685    2.57   0.010    .0098845   .0732638  
      Likidite5 | -10.85362   4.46874    -2.43   0.015   -19.61219   -2.095047  
      FinYap5 | 8.50793    3.335086    2.55   0.011    1.971282   15.04458  
      _cons | -1.688702   1.251986   -1.35   0.177   -4.14255    .7651456
```

Ek.5'in Devamı

SPSS Çıktısı

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	56.842	6	.000
	Block	56.842	6	.000
	Model	56.842	6	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	49.810 ^a	.513	.693

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	2.523	8	.961

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	FinYap2	-9.933	3.327	8.914	1	.003	.000
	SermayeYog2	.000	.000	11.026	1	.001	1.000
	Buyukluk	3.737	1.390	7.230	1	.007	41.986
	YabSermaye	.042	.016	6.612	1	.010	1.042
	Likidite5	-10.854	4.469	5.899	1	.015	.000
	FinYap5	8.508	3.335	6.508	1	.011	4953.902
	Constant	-1.689	1.252	1.819	1	.177	.185

a. Variable(s) entered on step 1: FinYap2, SermayeYog2, Buyukluk, YabSermaye, Likidite5, FinYap5.

Classification Table^a

Observed			Predicted		
			y		Percentage Correct
			0	1	
Step 1	y	0	25	7	78.1
		1	5	42	89.4
Overall Percentage					84.8

a. The cut value is .500

EK.6 Kesirli Polinomlar Çıktısı

(FinYap2 için)

fracpoly logit y FinYap2 SermayeYog2 Buyukluk YabSermaye Likidite5 FinYap5, compare

.....

-> gen double IFinY__1 = FinYap2^3-.0277117306 if e(sample)

-> gen double IFinY__2 = FinYap2^3*ln(FinYap2)+.0331238266 if e(sample)

Iteration 0: log likelihood = -53.325894

Iteration 1: log likelihood = -31.470079

Iteration 2: log likelihood = -26.261177

Iteration 3: log likelihood = -23.069584

Iteration 4: log likelihood = -21.620449

Iteration 5: log likelihood = -21.326689

Iteration 6: log likelihood = -21.30554

Iteration 7: log likelihood = -21.305182

Iteration 8: log likelihood = -21.305182

Logistic regression Number of obs = 79

LR chi2(7) = 64.04 Prob > chi2 = 0.0000

Log likelihood = -21.305182 Pseudo R2 = 0.6005

 y | Coef. Std. Err. z P>|z| [95% Conf. Interval]

```
-----+-----
IFinY__1 | -51.9244   29.92416   -1.74   0.083   -110.5747   6.725868
IFinY__2 | -46.33207   50.87512   -0.91   0.362   -146.0455   53.38134
ISerm__1 | 8.77e-06   2.87e-06   3.06   0.002   3.14e-06   .0000144
Buyukluk | 3.613839   1.518983   2.38   0.017   .6366866   6.590991
IYabS__1 | .0677053   .0248941   2.72   0.007   .0189136   .1164969
ILiki__1 | -8.967307   4.247005   -2.11   0.035   -17.29128   -.6433298
IFinYa_1 | 8.770861   3.452947   2.54   0.011   2.003209   15.53851
      _cons | 2.305496   .9231263   2.50   0.013   .4962014   4.11479
-----
```

Fractional polynomial model comparisons:

FinYap2 df Deviance Gain P(term) Powers

```
-----
```

Model	df	Deviance	Gain	P(term)	Powers
Not in model	0	63.389	--	--	
Linear	1	49.810	0.000	0.000	1
m = 1	2	43.716	6.094	0.014	3
m = 2	4	42.610	7.199	0.575	3 3

Ek.6'nın Devamı

(Buyukluk için)

fracpoly logit y Buyukluk YabSermaye Likidite5 FinYap5 FinYap2 SermayeYog2, compare

.....

-> gen double IBuyu__1 = X^2 if e(sample)

-> gen double IBuyu__2 = X^2*ln(X) if e(sample)

(where: X = (Buyukluk+1))

note: IBuyu__2 dropped due to collinearity

Iteration 0: log likelihood = -53.325894

Iteration 1: log likelihood = -32.455609

Iteration 2: log likelihood = -27.261292

Iteration 3: log likelihood = -25.209538

Iteration 4: log likelihood = -24.912528

Iteration 5: log likelihood = -24.904921

Iteration 6: log likelihood = -24.904914

Logistic regression	Number of obs = 79
LR chi2(6) = 56.84	Prob > chi2 = 0.0000
Log likelihood = -24.904914	Pseudo R2 = 0.5330

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
IBuyu__1	-4.983117	1.853182	-2.69	0.007	-8.615287 -1.350946
IYabS__1	.0415742	.0161685	2.57	0.010	.0098845 .0732638
ILiki__1	-10.85362	4.46874	-2.43	0.015	-19.61219 -2.095047
IFinY__1	8.50793	3.335086	2.55	0.011	1.971282 15.04458
IFinYa__1	-9.932728	3.326845	-2.99	0.003	-16.45322 -3.412232
ISerm__1	5.45e-06	1.64e-06	3.32	0.001	2.23e-06 8.67e-06
_cons	5.586161	1.853692	3.01	0.003	1.95299 9.219331

Fractional polynomial model comparisons:

Buyukluk	df	Deviance	Gain	P(term)	Powers
Not in model	0	60.355	--	--	
Linear	1	49.810	0.000	0.001	1
m = 1	2	49.810	0.000	1.000	-2
m = 2	4	49.810	0.000	1.000	-2 -2

Ek.6'nın Devamı

(YabSermaye için)

fracpoly logit y YabSermaye Likidite5 FinYap5 FinYap2 SermayeYog2 Buyukluk, compare

.....

-> gen double IYabS__1 = X^-2-.4906096535 if e(sample)

-> gen double IYabS__2 = X^3-2.910019277 if e(sample)

(where: X = (YabSermaye+.0599999986588955)/10)

Iteration 0: log likelihood = -53.325894

Iteration 1: log likelihood = -31.54844

Iteration 2: log likelihood = -26.249092

Iteration 3: log likelihood = -24.007801

Iteration 4: log likelihood = -23.61483

Iteration 5: log likelihood = -23.599609

Iteration 6: log likelihood = -23.599576

Logistic regression	Number of obs =	79
LR chi2(7) = 59.45	Prob > chi2 =	0.0000
Log likelihood = -23.599576	Pseudo R2 =	0.5574

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
IYabS__1	-.0000794	.0000379	-2.10	0.036	-.0001536	-5.17e-06
IYabS__2	.0021787	.0024727	0.88	0.378	-.0026677	.0070252
ILiki__1	-11.41502	4.780794	-2.39	0.017	-20.7852	-2.044838
IFinY__1	9.133216	3.585204	2.55	0.011	2.106344	16.16009
IFinYa__1	-10.5042	3.568437	-2.94	0.003	-17.49821	-3.510192
ISerm__1	5.72e-06	1.73e-06	3.31	0.001	2.33e-06	9.12e-06
Buyukluk	3.798341	1.469575	2.58	0.010	.9180261	6.678655
_cons	2.054038	.9676219	2.12	0.034	.1575337	3.950542

Fractional polynomial model comparisons:

YabSermaye	df	Deviance	Gain	P(term)	Powers
Not in model	0	58.557	--	--	
Linear	1	49.810	0.000	0.003	1
m = 1	2	48.102	1.708	0.191	-2
m = 2	4	47.199	2.611	0.637	-2 3

Ek.6'nın Devamı

(Likidite5 için)

fracpoly logit y Likidite5 FinYap5 FinYap2 SermayeYog2 Buyukluk YabSermaye, compare

.....

-> gen double ILiki__1 = Likidite5^.5-.4604126035 if e(sample)

-> gen double ILiki__2 = Likidite5^3-.0095254 if e(sample)

Iteration 0: log likelihood = -53.325894

Iteration 1: log likelihood = -32.42165

Iteration 2: log likelihood = -27.146739

Iteration 3: log likelihood = -24.833048

Iteration 4: log likelihood = -24.285829

Iteration 5: log likelihood = -24.252756

Iteration 6: log likelihood = -24.252618

Logistic regression	Number of obs	=	79
	LR chi2(7)	=	58.15
	Prob > chi2	=	0.0000

Log likelihood = -24.252618	Pseudo R2	=	0.5452
-----------------------------	-----------	---	--------

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
ILiki__1	-3.688889	5.14791	-0.72	0.474	-13.77861 6.40083
ILiki__2	-35.47479	25.96504	-1.37	0.172	-86.36533 15.41575
IFinY__1	8.705547	3.417771	2.55	0.011	2.006839 15.40426
IFinYa__1	-9.787869	3.445964	-2.84	0.005	-16.54184 -3.033903
ISerm__1	6.20e-06	1.91e-06	3.24	0.001	2.45e-06 9.94e-06
Buyukluk	4.0331	1.425731	2.83	0.005	1.23872 6.827481
IYabS__1	.0528796	.0212954	2.48	0.013	.0111415 .0946177
_cons	.9590403	.6000558	1.60	0.110	-.2170474 2.135128

Fractional polynomial model comparisons:

Likidite5	df	Deviance	Gain	P(term)	Powers
Not in model	0	58.797	--	--	
Linear	1	49.810	0.000	0.003	1
m = 1	2	48.998	0.812	0.368	3
m = 2	4	48.505	1.305	0.782	.5 3

Ek.6'nın Devamı

ROC curve çizmek için Stata Kodu ve Sonuçları

lroc

Logistic model for y

number of observations = 79

area under ROC curve = 0.9402

ÖZGEÇMİŞ

Serap Şahin ilk, orta ve lise öğrenimini Ankara'da sırası ile 13 Ekim İlköğretim, Mehmet Akif Ersoy ve Aydınlikevler Lisesin'de tamamlamıştır.

1993'de Hacettepe Üniversitesi, Fen Fakültesi, İstatistik bölümünü bitirmiştir. 1995'de Milli Eğitim Bakanlığı'nın Bursu ile Amerika'ya gitmiştir. 1997 yılında Virginia Tech Üniversitesi İstatistik Bölümünden yüksek lisans derecesini almıştır.

1998 – 2004 yılları arasında Kırıkkale Üniversitesin'de araştırma görevlisi olarak çalışmıştır. 2005 yılında İstanbul Üniversitesi İşletme Fakültesi Sayısal Yöntemler Anabilim Dalı'nda doktora başlamıştır. Halen İstanbul Üniversitesi İşletme Fakültesi Sayısal Yöntemler Anabilim Dalı'nda Araştırma Görevlisi olarak çalışmaktadır.