

İÇİNDEKİLER

İçindekiler	I
Tablolar Listesi	III
Şekiller Listesi	V
Özet	VI
Abstract	VIII
GİRİŞ	1
BÖLÜM 1	
1. REGRESYON ANALİZİ	3
1.1 Doğrusal Regresyon Analizi	4
1.2 Çoklu Regresyon Analizi	8
1.3 Doğrusal Olasılık Modeli.....	11
BÖLÜM 2	
2. LOJİSTİK REGRESYON ANALİZİ	16
2.1 Lojistik Regresyon Analizinin Doğrusal Regresyonla İlişkisi	21
2.2 İki Grup Lojistik Modeller.....	23
2.2.1 Lojit Dönüşüm ve Lojit Model	23
2.2.2 Lojistik Model Varsayımları	25
2.2.3 Lojistik Ayrımsama	28
2.2.4 İki Grup Lojistik Modellerde Kestirim Yöntemleri	29
2.2.4.1 En Çok Olabilirlik Yöntemi	30
2.2.4.2 Yeniden Ağırlıklandırılmış İteratif En Küçük Kareler Yöntemi	33
2.2.4.3 Minimum Lojit Ki-Kare Yöntemi.....	33
2.2.5 Uyum İyiliği ve Sapma Ölçütleri.....	35

2.3 Çoklu Grup Lojistik Modeller	39
2.3.1 Çoklu Grup Lojistik Model Varsayımları	43
2.3.2 Çoklu Grup Lojistik Modellerde Kestirim Yöntemleri	44
2.3.2.1 En Çok Olabilirlik Yöntemi	45
2.3.2.2 Yeniden Ağırlıklandırılmış İteratif En Küçük Kareler Yöntemi.....	46
2.3.2.3 Minimum Lojit Ki-Kare Yöntemi.....	47
2.3.3 Uyum İyiliği Testleri	49
2.3.4 Değişken ve Temel Sınıf Seçimi	49
2.4 Lojistik Regresyon İle Diskriminat Analizinin Karşılaştırılması	50

BÖLÜM 3

3. SİGORTACILIK SEKTÖRÜNDE LOJİSTİK REGRESYON ANALİZİNİN UYGULANMASI.....52

3.1 Uygulamanın Amacı ve Kapsamı	52
3.2 Uygulamada Kullanılan Veriler.....	53
3.3 Şirketlerin 2005 Yılı Üretimleri Baz Alınarak Uygulanan Lojistik Regresyon Analizi	56
3.4 Şirketlerin 2004-2005 Yılı Pazar Payları Baz Alınarak Uygulanan Lojistik Regresyon Analizi.....	63
3.5 Şirketlerin 2005 Yılı Ortalama Satışları Baz Alınarak Uygulanan Lojistik Regresyon Analizi.....	69
3.6 Şirketlerin 2005 Yılı Pazar Payları Baz Alınarak Lojistik Regresyon Analizi İle Diskriminat Analizinin Karşılaştırılması.....	75

SONUÇ

KAYNAKÇA

TABLolar LİSTESİ

Tablo 1. Çoklu Regresyonun Verisel Gösterimi	9
Tablo 2. Bağımsız değişken ikili kodlandığında lojistik regresyon modelinin değeri....	24
Tablo 3. Türk Sigorta Sektöründeki Şirketler.....	53
Tablo 4. Seçilen Gözlemler	56
Tablo 5. Başlangıç Adımı Sınıflandırma Tablosu	56
Tablo 6. Başlangıç Adımında Denkleme Giren Değişkenler	57
Tablo 7. Başlangıç Adımında Denkleme Girmeyen Değişkenler.....	57
Tablo 8. Anlamlılık Testi.....	58
Tablo 9. Model Özeti.....	58
Tablo 10. Sonuç Sınıflandırma Tablosu	58
Tablo 11. Denkleme Giren ve Girmeyen Değişkenler	59
Tablo 12. Korelasyon Matrisi	60
Tablo 13. Seçilen Gözlemler	63
Tablo 14. Başlangıç Adımı Sınıflandırma Tablosu	63
Tablo 15. Başlangıç Adımında Denkleme Giren Değişkenler	64
Tablo 16. Başlangıç Adımında Denkleme Giren Değişkenler	64
Tablo 17. Anlamlılık Testi.....	65
Tablo 18. Model Özeti.....	65
Tablo 19. Sonuç Sınıflandırma Tablosu	65
Tablo 20. Denkleme Giren Değişkenler	66
Tablo 21. Seçilen Gözlemler	69
Tablo 22. Başlangıç Adımı Sınıflandırma Tablosu	69
Tablo 23. Başlangıç Adımında Denkleme Giren Değişkenler	70
Tablo 24. Başlangıç Adımında Denkleme Girmeyen Değişkenler.....	70

Tablo 25. Anlamlılık Testi.....	71
Tablo 26. Model Özeti.....	71
Tablo 27. Sonuç Sınıflandırma Tablosu	71
Tablo 28. Denkleme Giren ve Girmeyen Değişkenler	72
Tablo 29. Seçilen Gözlemler	75
Tablo 30. Anlamlılık Testi.....	75
Tablo 31. Bağımlı Değişkenlerin Sınıflandırılması	76
Tablo 32. Modelin Açıklanma Oranı.....	77
Tablo 33. Anlamlılık Testi.....	77
Tablo 34. Kurulan Modelin Standartlaştırılmış Fonksiyonu	77
Tablo 35. Sonuç Sınıflandırma Tablosu	79

ŞEKİLLER LİSTESİ

Şekil 1. Doğrusal ilişki önerir	10
Şekil 2. Eğrisel ilişki önerir	10
Şekil 3. İlişki yoktur.....	10
Şekil 4. Lojistik regresyon fonksiyonu	19
Şekil 5. Lineer Diskriminat Analizi ile Sınıflandırma	61
Şekil 6. Regresyon Analizi ile Sınıflandırma	61
Şekil 7. Lojistik Regresyon Analizi ile Sınıflandırma	62
Şekil 8. Lineer Diskriminat Analizi ile Sınıflandırma	67
Şekil 9. Regresyon Analizi ile Sınıflandırma	67
Şekil 10. Lojistik Regresyon Analizi ile Sınıflandırma	68
Şekil 11. Lineer Diskriminat Analizi ile Sınıflandırma	73
Şekil 12. Regresyon Analizi ile Sınıflandırma	73
Şekil 13. Lojistik Regresyon Analizi ile Sınıflandırma	74
Şekil 14. Pazar Payları Artmayan Şirketlerin Analizi	78
Şekil 15. Pazar Payları Artan Şirketlerin Analizi	78
Şekil 16. Diskriminant Analizi ile Sınıflandırma	80

ÖZET

Araştırmacılar ya da model tasarımcıları gerçek olaylardan ya da deneylerden elde ettikleri verileri çeşitli modeller yardımı ile fonksiyonel yapılar haline dönüştürmek için daima çaba harcarlar. Matematiksel modelleri kurmak oldukça zor olmasına karşın sağladığı faydalar bağlamında oldukça yararlı bilgiler sağlamaktadırlar. Modellerde kullanılan verileri sınıflama çok değişkenli istatistiksel analizin önemli bir bölümünü oluşturmakla birlikte, sağlık başta olmak üzere çeşitli bilim dallarında çok geniş bir kullanım alanına sahiptir. Özellikle verilerin değerlendirilmesi sonucunda başarılı-başarısız, hasta-hasta değil, iyi-orta-kötü gibi kategorik sonuçların olduğu durumlarda lojistik regresyon son derece kullanışlı bir istatistiksel yöntem olmaktadır. Lojistik regresyonun bu özelliğinden yararlanılarak çalışmada sigorta sektöründe faaliyet gösteren 46 şirketin verileri değerlendirilerek araştırılmış ve ayrıca verilere diskriminant analizi de uygulanarak sonuçlar karşılaştırılmıştır.

Çalışmanın giriş bölümünde lojistik regresyonla ilgili son yıllarda yapılan yayınlar incelenmiş, neler yapıldığı hangi alanlara uygulandığı ve sigorta sektöründe ne şekilde kullanıldığı araştırılmıştır.

Çalışmanın birinci bölümünde, lojistik regresyon analizinin çıkışını sağlayan ve birçok temel özelliği içinde barındıran doğrusal regresyon analizi ile çoklu regresyon analizi incelenmiştir. Ayrıca, bu analizlerin özellikleri incelenirken niçin lojistik regresyona ilginin kaydığına da değinilmiştir.

İkinci bölümde iki grup ve çoklu grup lojistik regresyon analizinin teorik anlatımı ve doğrusal regresyonla ilişkisi incelenmiştir. Sınıflandırma analizlerinin en çok kullanılanlarından biri olan ve birçok konuda uygulama alanı bulunan lojistik regresyonun parametre tahminine ilişkin kestirim yöntemleri de detaylı bir şekilde incelenmiştir.

Çalışmanın üçüncü bölümünde ise, sigorta sektöründe faaliyet gösteren 46 şirketin 2004–2005 bilânçoları, toplam üretimleri, poliçe sayıları, pazar payları ile yangın, nakliyat, mühendislik, tarım, kasko, trafik, zorunlu sigortalar, ferdi kaza, sağlık ve hayat ürünleri branşlarındaki verileri dikkate alınarak lojistik regresyon analizi ile gözlemler gruplara ayrılarak, veriler sınıflandırılmış ve sigorta şirketleri başarılı - başarısız biçiminde gruplandırılarak ileriye yönelik tahminlerde kullanılacak modellerin oluşturulması sağlanmış, ayrıca benzer özelliklerin diskriminant analiziyle olan farkları ortaya konup birbirlerine karşı üstünlükleri anlatılmıştır. Çalışmada kullanılan üç ayrı sınıflama için üç lojistik model kurulmuş ve bu modelleri oluşturacak değişkenlerin amaca uygun seçilmesine çalışılmıştır.

Şirketlerin gelişmişliğini, tercih sebebi olmalarını etkileyecek olan değişkenlerin bulunup modele konması, aynı özelliklere sahip farklı değişkenlerin tespit edilip sadece birinin modele alınması ile modelin anlamlılığının sağlanmasına çalışılmıştır. Bu seçimin doğru yapılması, sınıflandırma da doğruluk oranını arttıracaktır. Ayrıca uygulamada yukarıda sözü edilen ve değişken grubunu oluşturan bu ürünlerin birbirleri üzerindeki etkileri de araştırılmıştır.

ABSTRACT

The researchers or model designers of fashion struggle the data which are taken from real events or experiments in order to convert functional positions with the aid of various models. Despite of the establishment of mathematical models, they obtain very useful data. The classifications of data have wide usage of area in various scientific informations with part of statistical analysis. Especially logistical regression constitutes usable statistical method at the end of evaluation of data as successful –unsuccessful, patient-not patient, bad or good. The data of 46 firms are evaluated by utilizing the method of statistical in insurance sector and compared the results in applying discrimination analysis.

In the prologue edits are examined related with logistic regression that have made in recent years, and investigated about what kind of fields it is applied, how it is used in insurance sector.

In first part of studying, the analysis of linear regression and multiple regression were investigated which concludes logistical regression. Also, mentioned why caring to logistical regression with observing features of analysis.

In second part, the relation between linear regression and logistical regression of multiple groups have been observed, because of logistical regression have wide area of usage in classification analysis and many practical area, conjecture methods used in parameter guessing have been observed in detail. Also, in this part the theoretical comparison of logistical regression and discrimination regression were realized.

In third part of studying, balance sheet of 46 firms worked in insurance sector in 2004-2005, total production, numbers of policies, market shares and models were realized by caring in datas of firing, transportation, engineering, agriculture, automobile insurance, traffic, compulsive insurances, healthy and life products with regression analysis in order to classify to groups and classification to companies at the view of successful or unsuccessful, also creating of differences of similar products were realized by discrimination analysis. Three models were established to be used for different three classifications in studying. The best important function is to choose the variables which constitute these models.

By finding of variables which will effect the development of firms and putting to model the meaningful of model must be constituted. The ratio of validity will increase in true classification, also the effects of these products have been investigated.

GİRİŞ

Çalışmaya başlamadan önce dünyada lojistik regresyonla yapılan bazı araştırmalar incelenmiştir. Michigan Üniversitesi'nden William G. Weissert, Jennifer M. Elston ve North Carolina Üniversitesi'nden Gary G. Koch (1990) yapmış oldukları çoklu lojistik regresyon analiziyle Amerika'da 1976–1986 yılları arasında sağlık harcamalarının kısıtlanması için çıkarılan gezer hastabakıcılık sisteminin eski sistemle kıyaslanmasını incelemişlerdir.¹ Boston Üniversitesi'nden John B. Williamson ve Tay K. McNamara (2001) insanların yaşma, cinsiyetine, ekonomik ve sosyal statülerine göre emeklilik yaşının hesaplanmasında ikili lojistik regresyon kullanmıştır. George Washington Üniversitesi'nden Langche Zeng ve Harvard Üniversitesi'nden Gary King (2001) çok seyrek gerçekleşen olaylar karşısında lojistik regresyonun ve veri seçiminin nasıl uygulanması gerektiğini anlatmıştır.² Ran Braniv, Anurag Agarwal ve Robert Leach (2002), 1978 yılında iflas ettiği kabul edilen 207 Amerikan şirketi üzerinde yapmış oldukları çalışmada şirketlerin son iflas kararlarını finansal ve finansal olmayan değerleri lojistik analizlere sokarak sınıflandırmaya ve tahmin etmeye çalışmışlardır. Yine sağlık sigortasıyla ilgili Amerika' da Robin Fisher ve Jennifer Campbell (2002) sağlık sigortası tahminlemesinde sigortası olmayanların cinsiyet, yaş, ekonomik durumlarına göre sınıflandırılmasında lojistik regresyon modelinden yararlanmışlardır. Avusturya Monash Üniversitesinden Jeromey Temple (2002) Avusturyalı ev halkının 1960 yılından itibaren özel sağlık sigortasına geçişinin hızını ve bu geçişin nedenleri olan ekonomik, sağlık ve sosyal olguları çoklu lojistik regresyon analiziyle açıklamıştır.³ Portland State Üniversitesi'nden Kerri Sullivan (2003) yapmış olduğu çalışmada Amerika'daki araba kazalarını sınıflandırmak için sürücülerin eğitim düzeyi, okur-yazarlık oranı, ekonomik durumu ve cinsiyeti gibi değişkenleri kullanarak lojistik regresyon analizi yapmıştır. Barcelona Ekonomi Enstitüsü'nden Montserrat

¹ Weissert, William G., Elston, Jennifer M. & Koch, Gary G.; **Risk of Institutionalization**, <http://aspe.hhs.gov/daltcp/reports/instrkes.pdf>, 2006.

² King, Gary. & Zeng, Langche.; **Logistic Regression in Rare Events Data**, <http://gking.harvard.edu/files/0s.pdf>, 2006.

³ Temple, Jeromey.; **Explaining the private health insurance coverage for older Australians**, <http://www.elecpress.monash.edu.au/pnp/cart/download/free.php?paper=178>, 2006.

Álvarez (2003) Katalan şirketlerinin sadece kendi bünyesinde yatırım yapmaları ile diğer yabancı şirketlerle birleşerek yatırım yapmalarının hangisinin daha yararlı olacağı ve hangisinin tercih edilmesi gerektiğinin anlaşılması için birçok kriter baz alınarak ikili lojistik regresyon analizi yapılmıştır.⁴ Japonya da Hitotsubashi Üniversitesinden Tsun-Siou Lee, Yin-Hua Yeh ve Rong-Tze Liu (2003) Japon ekonomisine yön veren 57 büyük aile şirketini finansal açıdan stres altında ve huzurlu şekilde sınıflandırmışlardır.⁵ Michiel de Hoon (2004) insan genleri üzerinde yapmış olduğu çalışmasında lojistik regresyon analiziyle genleri özelliklerine göre sınıflandırmıştır. Michele E. Capella (2005) kör ya da görsel zayıflığı bulunan kişilerin yeniden iş bulup çalışabilmeleri için kurulan bir iyileştirme merkezinin yararını lojistik modellemeyle analiz etmiştir. Harvard Üniversitesi'nden Amitabh Chandra ve Dartmouth Koleji'nden Andrew A. Samwick (2005) Amerika'daki sakatlık riskini ve sakatlık risk sigortalarını incelemiş, bu çalışmada lojistik regresyon kullanarak malulen emekli olanları çalışabilme sınırlarına, çalışamama durumlarına ve sağlık yardımı almak zorunda olmalarına göre sınıflandırmıştır. Dortmund Üniversitesi'nden Marcos Marin Galiano ve Andreas Christmann (2005) tarafından yapılan çalışmada sigorta şirketlerinin araba sigortalarındaki tarifelerinin belirlenmesinde ve sınıflandırılmalarında lojistik regresyondan yararlanılmıştır.⁶

Dünyada lojistik regresyon analizinin kullanım alanının çok geniş olduğu literatür taramasında açıkça gözükmektedir. Bu çalışmada amaç, ikili ve çoklu grup lojistik regresyon analizine niçin ihtiyaç duyulduğunu, lojistik regresyon analizinin diğer çok değişkenli analizlerden farkını, kullanım alanlarını, teorik yapısını ortaya koymak, daha sonra Türkiye'deki sigortacılık sektöründeki tüm şirketleri faaliyetlerine, satışlarına, üretimlerine, pazar paylarına ve büyümelerine göre sınıflandırıp sigortacılık sektörünün matematiksel bir yüzünün görülmesine yardımcı olmaktır.

⁴ Álvarez, Montserrat.; Wholly-Owned Subsidiaries Versus Joint Ventures: **The Determinant Factors in the Catalan Multinational Manufacturing Case**. <http://www.pcb.ub.es/ieb/serie/doc2003-5.pdf>. 2006.

⁵ Lee ,Tsun-Siou., Yeh, Yin-Hua. & Liu, Rong-Tze.; **Center for Economic Institutions**, <http://cei.ier.hit-u.ac.jp/working/2003/2003WorkingPapers/wp2003-14.pdf>, 2006.

⁶ Galiano, Marcos Marin. & Christmann, Andreas.; **An R-Program to Model Insurance Data**, <http://www.sfb475.uni-dortmund.de/berichte/tr49-04.pdf>, 2006.

BİRİNCİ BÖLÜM

1. REGRESYON ANALİZİ

Çoğu kez, bir araştırmacı ya da deney yapan kimse iki ya da daha çok değişken arasında bir ilişki olup olmadığını bulmak ve bu ilişkinin bir denklemle nasıl ifade edilebileceğini göstermek ister.⁷ Örneğin bir mühendis basınç ve sıcaklık, bir ekonomist gelir düzeyi ve tüketim harcamaları, bir sigortacı satılan poliçe sayısı ile karlılığını, bir eğitimci öğrencilerin devamsızlık gösterdiği gün sayıları ve başarı dereceleri arasındaki ilişkiyi bilmek isteyebilir. İki (ya da daha çok) değişken arasındaki ilişkiyi gösteren denklem, değişkenler arasındaki ilişkiyi gösteren denklem, değişkenler arasındaki ilişkinin fonksiyonel şeklini göstermekle kalmaz, değişkenlerden birinin değeri bilindiğinde diğeri hakkında tahmin yapmasını sağlar. Örneğin, bir sigorta şirketinin çıkarmış olduğu bir ürünün satış rakamlarıyla ilgili bir model oluşturalım. Satış miktarı kısmen ürünün fiyatına bağlıdır. Satış miktarını Y ile, fiyatı X ile gösterelim. Amaç bir ürün fiyatını bildiğimizde oluşacak satışın miktarını tahmin edebilmektir. Ürünün fiyatı bazı X değerlerinde değişmese bile, ürünün satışı diğer rastgele etkiler nedeniyle hala değişebilecektir. Örneğin $X=1000$ YTL fiyatında çeşitli bölgelerde o ürünün satış rakamları alınmış ise, bu rakamlar değer olarak değişebilecektir. Bu nedenle X için Y/x ile gösterilen koşullu bir rastgele değişkenle ilgileniliyor demektir. Bu koşullu rastgele değişkenin ortalaması $u_{Y/x}$ ile gösterilir ve açıkça, ürünün ortalama satış rakamı fiyatına bağlıdır. $X=1000$ YTL fiyattan satış rakamı ile $X=500$ YTL fiyattan oluşacak satış rakamının aynı olacağı beklenir. Yani $u_{Y/x}$ in x in fonksiyonu olduğunu kabul etmek akla uygun gelmektedir. Bu fonksiyonun grafiğine X üzerinde Y nin regresyon eğrisi denir. X in değeri önceden biliniyor kabul edildiğinde ve Y tarafından kabul edilen değer kısmen X in özel değerine bağlı olduğundan Y ye bağımlı değişken, X e bağımsız değişken adı verilir.

⁷ Fikri Akdeniz, **Olasılık ve İstatistik**, Adana: Ç.Ü. Basımevi, 1998, s.441.

Problem; X in bazı seçilmiş $X_1, X_2, X_3, \dots, X_N$ değerinde bulunan verilere dayanarak $u_{Y/X}$ in formunu tahmin etmektir. Fonksiyonel ilişki varsa, bu ilişki X in seçilmiş değerine bağlı olarak elde edilecektir.

1.1 Doğrusal Regresyon Analizi

İki veya daha çok değişken arasındaki ilişkinin saptanması genellikle iki tür bilgi için gerekli olur. Bunlar, bir değişkene ilişkin gözlem sonuçları yardımıyla diğer değişkenin alabileceği değerlerin ne kadar doğrulukla tahmin edilebileceği ve değişken değerlerinde gözlenen farklılıkların ne dereceye kadar belirleyici bazı etmenlere bağlanabileceğidir⁸. Yani temelde yürütülen mantık, iki rassal değişkenin birbirine bağımlı olması durumunda, bu değişkenlerden birinin içerdiği bilginin diğer değişkenin değerlerini kestirmek için kullanılabilir⁹. Bu sebeplerle, değişkenler arasındaki ilişkinin fonksiyonel şekli, yönü ve derecesi bilinmelidir. İlişkinin fonksiyonel şekli ilişkinin varlığı durumunda regresyon analizinin konusunu oluştururken, iki değişken arasındaki ilişkinin ölçülmesi korelasyon analizinin konusunu oluşturur. Yani korelasyon en az iki değişken arasındaki ilişki iken regresyon, en az iki değişken arasındaki ilişkinin denklem ile ifade edilmesidir.¹⁰

Regresyon ve korelasyon ile tahminleme arasında da çok yakın bir ilişki vardır. Bu ilişki şöyle ifade edilebilir: Regresyon bilinenlerden yararlanılıp bilinmeyen durumların tahmin edilmesinde kullanılan bir tekniktir. Öte yandan, korelasyon katsayısının değeri yapılan tahminin güvenilirlik derecesinin bir göstergesidir.¹¹

Regresyon verileri nitel ve nicel olmak üzere iki türdür. Bağımlı değişken verileri sürekli değişken olmak zorunda iken bağımsız değişken verileri için böyle bir

⁸ Özer Serper, **Uygulamalı İstatistik 2**, 3. Baskı, İstanbul: Filiz,1996, s.187.

⁹ Necmi Gürsakal, **Bilgisayar Uygulamalı İstatistik 2**, 1.Baskı,İstanbul:Alfa, 2001,s.307.

¹⁰ Burhan Çil, **İstatistik**, 4.Baskı, Ankara: Detay Yayınları, 2004, s.282.

¹¹ Serper, **a.g.e**, s.190.

zorunluluk yoktur. Regresyon modelinin genel kullanım amaçları; salt tanımlama, katsayı kestirimi, denetim, uyum kestirimi ve önkestirimidir.

İki değişkenli dağılımlarda regresyon kuramının ana sorunu, bileşik olasılık dağılımının uygun yerinden geçirilecek bir doğru ya da bir eğri bulmaktır.¹² Regresyon lineer olarak düşünüldüğünde X üzerine Y nin regresyon doğrusu tahmin edilecektir. Bu durumda $u_{Y/X}$;

$$E(Y/X = x) = u_{Y/X} = \alpha + \beta x \quad (1)$$

ile verilecektir. Buna regresyon doğrusu denir. Basit lineer regresyon modelinde, β regresyon doğrusunun eğimi, α doğrunun Y eksenini kestiği noktadır.

X bağımsız değişkenin değerini gösterir ve bir gerçel sayıdır Ayrıca; bir rastgele değişken kendi ortalama değeri etrafında değişir. O halde Y/x_i ile $u_{Y/X}$ arasındaki fark E_i ile gösterilir ve E_i farkının beklenen değeri sıfır kabul edilir. O halde;

$$Y/x_i = \alpha + \beta x_i + E_i \quad (2)$$

şeklinde ifade edilir. Burada $E(E_i) = 0$ kabul edilir ve yaygın kullanım olarak koşullu gösterimin atılmasıyla basit doğrusal regresyon modeli;

$$Y_i = \alpha + \beta x_i + E_i \quad (3)$$

şeklinde ifade edilir. Verilerimiz (x_i, y_i) $i = 1, 2, 3, \dots, N$ çiftlerinden oluşmaktadır, x_i , X değişkeninin gözlenen değeri, y_i , Y rastgele değişkeni için karşılık gelen gözlemlerdir. Bir rastgele değişkenin gözlem değeri onun ortalamasından bir rastgele miktar olarak farklı olacağından denklemdeki E_i rastgele değişkeninin gerçekleşen değerine ε_i diyebiliriz. α ve β nin gerçek değerleri ile ε 'nın da gerçek değeri bilinemez.¹³

¹² H.Ceyhan İnal, Süleyman Günay, **Olasılık ve Matematiksel İstatistik**, Ankara, H.Ü.Fen Fakültesi Basımevi,1999,s.462.

¹³ Akdeniz, **a.g.e.**, s.443.

Regresyon modelinde teorik parametreler yerine onların tahmini değerleri olan a ve b ile yer değiştirir ise;

$$\hat{y} = u_{Y/X} = a + bx \quad (4)$$

Şeklinde tahmin edilmiş regresyon doğrusunu elde edilebilir. Tahminlerdeki tutarlılık bir başka deyişle sapma da $e_i = y_i - \hat{y}$ den elde edilebilir.

Regresyon analizinin uygulanabilmesi için ya da regresyon analiziyle elde edilen sonuçların güvenilir olabilmesi için bazı varsayımların sağlanmış olması gereklidir. Kullanım sahası son derece geniş olan regresyon analizinin varsayımlarından bahsedilmeden önce hata teriminden bahsedilecektir. Sonucu ortaya çıkaran bütün sebepler belirlenememiş dolayısıyla modele dahil edilememiş olduğundan, modele ε hata terimi girmek zorundadır.

Doğrusal regresyon denkleminin varsayımları;

i-) x_i değişkenine ilişkin değerler sabittir. Her x değeri için y değerlerinden oluşan bir altküme vardır. Söz konusu bu alt kümelerin dağılımları normaldir. Y/X rassal değişkenin, ortalaması $u_{Y/X}$ ve varyansı $\sigma^2_{Y/X}$ tir. Eğer $x=x_i$ ise, Y/x_i , Y_i rassal değişkenini gösterir. Bu bahsedilen Y_i nin ortalaması $u_{Y/X(i)}$ dir, varyansı da $\sigma^2_{Y/X(i)}$ dir. Y_i lerin dağılımı ile ilgilenilir. Y_i lerin bağımsız olduğu varsayılır. Y_i için aralık tahmini yapılacağından ve hipotez testleri ile ilgilenileceğinden $Y_1, Y_2, Y_3, \dots, Y_N$ in normal dağılması gerekir¹⁴.

ii-) Bu alt kümelerin varyansları eşittir.

iii-) Bu alt kümelerin ortalamaları aynı doğru üzerindedir¹⁵.

¹⁴ E.Ronald Walpole, **Introduction To Statics**,3.Baskı,New York: Mac Millan,1982,s.353.

¹⁵ Gürsakal, **a.g.e**,s.310.

Bilindiği gibi modelin matematiksel kalıbından, model dışı bırakılan değişkenlerden bağımlı değişkendeki ölçme hatalarından vb. kaynaklanan çeşitli hataların hesaba katılması amacıyla hata terimine modelde yer verilmektedir. Hata teriminin rassal olabilmesi için, model dışı bırakılmış değişkenlerin çok sayıda, her birinin tek başına önemsiz olması ve bunların farklı yönlerde değişerek bağımlı değişken üzerindeki genel etkilerinin kestirilememesi gerekir. Ayrıca ölçüm sırasında yapılan hataların da rastgele olması gerekmektedir.

Doğrusal regresyonda hata terimi için ise şu varsayımlar yapılır ;

i-) Hata terimi rassal bir değişkendir.

ii-) Hata teriminin ortalaması sıfırdır.

iii-) Hata teriminin alabileceği bütün değerler göz önüne alındığında, herhangi bir x değeri için hata terimlerinin ortalaması sıfıra eşittir.

iv-) Hata teriminin varyansı x değerlerine göre değişmez.

v-) Bütün x değerleri için e hata terimleri kendi ortalamaları etrafında aynı değişkenliğe sahiptir.¹⁶

vi-) Hata terimi normal bölünmeye sahiptir.

vii-) Her x_i için hata teriminin değerleri kendi ortalamaları etrafında çan eğrisi biçiminde simetrik bir bölünme gösterir.

viii-) Farklı y değerlerine ilişkin rassal hatalar birbirinden bağımsızdır.¹⁷

¹⁶ James R. Evans, **Statistics, Data Analysis, And Decision Modeling**, 1.Baskı, Upper Saddle River NJ: Prentice, 2003, s.193.

¹⁷ Paul Newbold, William L. Carlson, Betty Thorne, **Statistics For Business And Economics**, 5.Baskı, Upper Saddle River NJ: Prentice Hall, 2003, s.582.

ix-) Bağımsız değişken değerleri sabit sayılar olup hata terimi bağımsız değişkenden bağımsızdır.

X ve Y değerleriyle ilgili çok sayıda örneklem alındığında, bütün örneklemdaki x_i değerleri aynıdır. Yani x_i 'ler sabit sayılardır. Fakat e_i ve dolayısıyla y_i değerleri farklı rassal nedenlerle değişir.

x-)Bağımsız değişken x 'in ölçüm hataları ihmal edilebilecek kadar küçüktür.

xi-)Bağımsız değişkenler arasında yüksek derecede ilişki yoktur.¹⁸

xii-)Modelin matematiksel kalıbı tektir ve ilişkinin belirlenişi doğrudur.¹⁹

1.2 Çoklu Regresyon Analizi

Bir bağımsız, bir bağımlı değişkenli doğrusal regresyon çok sayıda bağımsız değişken için genişletilebilir. Örneğin y değişkeninin x_1 ve x_2 bağımsız değişkenlerin fonksiyonu olduğunu kabul edelim. O halde aşağıdaki denklem yazılabilir;

$$Y = a + bx_1 + cx_2 \quad (5)$$

(5) denklemi üç boyutlu uzayda bir düzlem gösterir. O halde şimdi (y_i, x_{1i}, x_{2i}) i inci noktanın koordinatları olmak üzere Tablo 1 de verilmiş N nokta için en iyi düzlem bulunmaya çalışılır.

¹⁸ Reha Alpar, **Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1**, Ankara: Kültür Ofset, 1997, s.201.

¹⁹ Serper, **a.g.e** ,ss.190 - 198.

Tablo 1. Çoklu Regresyonun Verisel Gösterimi

y	x ₁	x ₂
y ₁	x ₁₂	x ₂₂
y ₂	x ₁₂	x ₂₂
•	•	•
•	•	•
y _N	x _{1N}	x _{2N}

Diğer bir deyimle, N nokta düzleme olabildiği kadar yakın olacak şekilde a, b, c, değerleri belirtilir. Hata terimini bulmak için gerçek y değeriyle tahmini \hat{y} değerinin birbirinden ayrılması gerekir.

$$\hat{y} = u_{Y/X} = a + bx_1 + cx_2 \quad (6)$$

$$u_{Y/x_1, x_2} = \alpha + \beta x_1 + \gamma x_2 \quad (7)$$

Bu değerler basit doğrusal regresyonda olduğu gibi $\sum_{i=1}^N (y_i - \hat{y}_i)^2$ toplamı minimum olacak şekilde belirtilir. x₁, x₂ ve y nin N değeri verildiğinde

$$\hat{y} = u_{Y/X} = a + bx_1 + cx_2$$

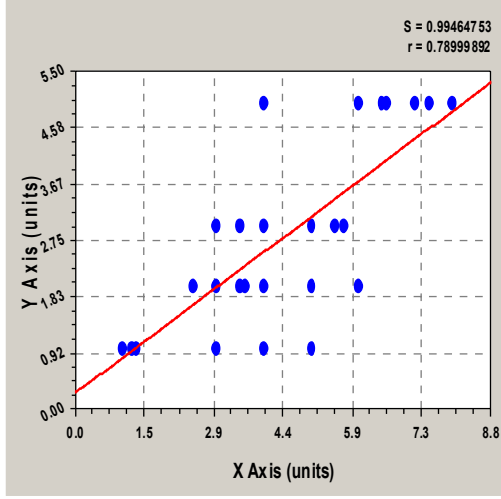
Denklemindeki a, b ve c parametreleri aşağıdaki normal denklem sisteminin çözümüyle bulunur.

$$\sum_{i=1}^N y_i = aN + b \sum_{i=1}^N x_{1i} + c \sum_{i=1}^N x_{2i} \quad (8)$$

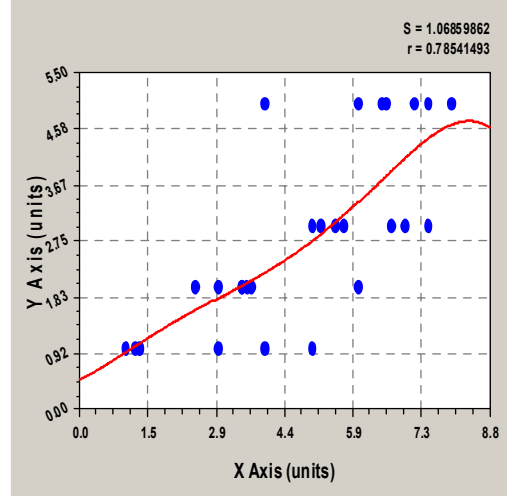
$$\sum_{i=1}^N y_i x_{1i} = a \sum_{i=1}^N x_{1i} + b \sum_{i=1}^N x_{1i}^2 + c \sum_{i=1}^N x_{1i} \cdot x_{2i} \quad (9)$$

$$\sum_{i=1}^N y_i x_{2i} = a \sum_{i=1}^N x_{2i} + b \sum_{i=1}^N x_{1i} \cdot x_{2i} + c \sum_{i=1}^N x_{2i}^2 \quad (10)$$

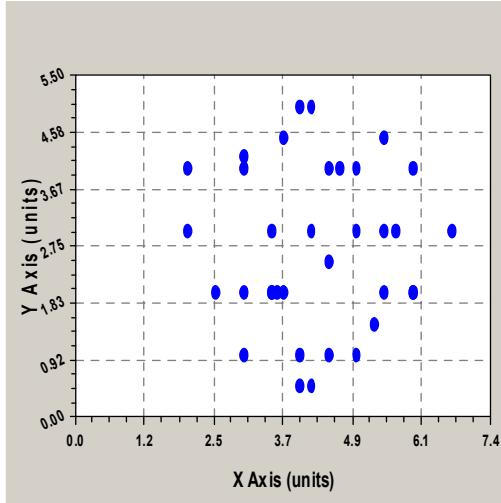
Koordinat sisteminde (x_i, y_i) çiftlerinin kümesinin grafiğine serpilme diyagramı denir.²⁰



Şekil 1. Doğrusal ilişki önerir.



Şekil 2. Eğrisel ilişki önerir.



Şekil 3. İlişki yoktur.

Bu şekiller değişkenler arasındaki ilişkinin daha rahat görünmesini sağlar. Kurulması gereken regresyon modelinin doğrusal mı, eğrisel mi olması gerektiğine diyagram yardımıyla karar verilir.

²⁰ Akdeniz, a.g.e., s.443.

1.3 Doğrusal Olasılık Modeli

Genel doğrusal regresyon modelinin çeşitli gösterim biçimleri vardır. Bunlardan bir tanesi de, sol tarafın koşullu beklenen değer olduğu tanım biçimidir.

$$E (y_i/x_{i1}, \dots, x_{ip}) = \sum_{k=0}^p B_k x_{ik} ; \quad i = 1, 2, 3, \dots, n \text{ için} \quad (11)$$

(p: bağımsız değişken sayısı. i: gözlem sıra no, k: bağımsız değişken sıra no)

Bu modelde açıklayıcı değişkenler üzerinde kısıt yok iken, y bağımlı değişkeninin sürekli olması koşulu vardır. Herhangi bir i' inci gözlem için,

$$y_i = \sum_{k=0}^p B_k x_{ik} + \varepsilon_i \quad (12)$$

biçiminde ifade edilen modelde açıklayıcı değişkenler üzerinde bir kısıt olmadığından y sonuç değeri (bağımlı değişken), - ile + arasında tüm değerleri alabilmektedir.²¹ Bağımlı değişkenin 0, 1 gibi değerler aldığı durumda bu kural bozulmakta ve P (y = 1), i' inci gözlemin 1 değerini alma olasılığı olmak üzere, beklenen değer,

$$E (y_i) = 1 \cdot P (y_i = 1) + 0 \cdot P (y_i = 0) = P(y_i = 1) \quad (13)$$

olmaktadır, bu sonuç regresyon denklemi olarak yazılacak olursa;

$$E(y_i) = P (y_i = 1) = \sum_{k=0}^p B_k x_{ik} \quad (14)$$

²¹ Hüseyin Tatlıdil, **Uygulamalı Çok Değişkenli İstatistiksel Analiz**, Ankara: Cem Web Ofset, 1996, s.290.

elde edilir. Sol tarafı 0–1 arasında olasılık değerleri alan bu denkleme doğrusal olasılık modeli (DOM) adı verilmektedir.²²

DOM, tek açıklayıcı değişkenli bir basit doğrusal regresyon modelinden hareket edilerek aşağıdaki şekilde elde edilir.

$$y = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i \quad (15)$$

Bu modelde y değerlerinin beklenen değeri alınırsa yani y: 0-1 (dikotom), tanım gereği,

$$E(y) = 1 \cdot P(y = 1) + 0 \cdot P(y = 0) \quad (16)$$

$$E(y) = P(y=1) = p_i \quad (17)$$

Elde edilir. Diğer bir deyişle Y' lerin beklenen değeri (y=1) olma olasılığı olacaktır. Beklenen değer, bir de şu şekilde alınabilir:

$$E(y) = E(\beta_0 + \beta_1 \cdot x_i + \varepsilon_i) \quad (18)$$

$$E(y) = \beta_0 + \beta_1 \cdot x \quad (19)$$

İki beklenen değer sonucu birleştirilirse,

$$P = \beta_0 + \beta_1 \cdot x \quad (20)$$

modeli elde edilir. Bu model Doğrusal Olasılık Modeli'dir. Modelden görülebileceği gibi, y'nin 1 'e eşit olması, x'in bir lineer fonksiyonudur. Bu modelde β nin yorumu; x bağımsız değişkenindeki 1 birimlik değişme, y' nin 1 'e eşit olma olasılığında β kadar bir değişme sağlar şeklinde olacaktır.

²² Tatlıdil, a.g.e., s.290.

Bağımlı değişkenin dikotom olması durumunda, yukarıda belirtilen regresyon varsayımlarının ne ölçüde geçerli olacağı önemlidir. Y bağımlı değişkeninin dikotom olması durumunda, ε_i sadece iki değer alabilir.

Sonuç değişkenin ikili olduğu (12) no' lu denklemde ifade edilen modelin hata terimi, $y = 0$ ve $y = 1$ durumları için sırasıyla;

$$y_i = \sum_{k=0}^p \beta_k x_{ik} + \varepsilon_i = 0 \quad \text{için} \quad \varepsilon_i = - \sum_{k=0}^p \beta_k x_{ik} + \varepsilon_i \quad (21)$$

$$y_i = \sum_{k=0}^p \beta_k x_{ik} + \varepsilon_i = 1 \quad \text{için} \quad \varepsilon_i = 1 - \sum_{k=0}^p \beta_k x_{ik} + \varepsilon_i \quad (22)$$

değerlerini almaktadır. ε_i 'nin sadece iki değer alması, onun normal dağılıma sahip olmasını engeller. Bu sebeple normallik varsayımı gerçekleşemez²³.

$$E(\varepsilon_i) = P(y_i=0) \left(- \sum_{k=0}^p \beta_k x_{ik} \right) + P(y_i=1) \left(1 - \sum_{k=0}^p \beta_k x_{ik} \right) = 0 \quad (23)$$

$$\text{Var}(\varepsilon_i) = E(\varepsilon_i)^2 P(y_i=0) \left(- \sum_{k=0}^p \beta_k x_{ik} \right)^2 + P(y_i=1) \left(1 - \sum_{k=0}^p \beta_k x_{ik} \right)^2 \quad (24)$$

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

Görüldüğü gibi, gözlemden gözleme varyans değişecektir. $\text{Var}(\varepsilon_i)$ x'lerin bir fonksiyonudur.²⁴

²³ Paul D. Allison, **Logistic Regression-Using The SAS System: Theory and Application**, 2. Edition, Cary: SAS Institute, 2000, s.s 8-9.

²⁴ Damodar N.Gujarati, **Temel Ekonometri** (Çev. :Ümit Şenesen, Gülay Günlük Şenesen), 1. Basım, İstanbul: Çevik, 1999, s.543.

Hataların normalliği ve hata varyansının sabitliği varsayımlarının gerçekleştirilememesi, yansız tahmin elde edilmesini etkilemez. Sadece birinci ve ikinci varsayımlarla β ve $\hat{\beta}$ 'nin en küçük kareler kestirimi yansız, ancak en iyi olmayacaktır.²⁵

Hata terimlerinin normal dağılmaması da çok önemli değildir. Çünkü Merkezi Limit Teoremi gereği örneklem hacmi yeterince büyük olursa, normal dağılmasa da, katsayı tahminleri yaklaşık olarak normal dağılacaktır. Dolayısıyla şu durumda güven aralıkları hesabında halen normal dağılım tablosu kullanılabilir. Ancak sabit varyanslılığın gerçekleştirilememesi nedeniyle, katsayı tahminleri artık etkin olmayacaktır. Bu durumda standart hata tahminleri yanlış olacaktır. Standart hata ise, bilindiği gibi, test istatistiği hesabında kullanılmaktadır. Dolayısıyla test istatistiği de yanlış olacaktır. Değişen varyanslılığı ortadan kaldırmak için bir ağırlıklandırma yapılmalıdır. (24) denklemleri ile gösterilen modelden yararlanılarak bu ağırlıklandırma, çok değişkenli genel regresyon denkleminde şöyle gösterilebilir:

$$w_i = 1 / \left(\sum_{k=0}^p \beta_k x_{ik} \right) \left(1 - \sum_{k=0}^p \beta_k x_{ik} \right) \quad (25)$$

β_k 'ler, y_i 'nin x 'ler üzerindeki regresyonundan bulunacaktır ve ağırlık değerlerinde kullanılacaktır. Bu sayede yeni hata terimi $w_i \varepsilon_i$ olacaktır. Yeni hata terimi yansızlık özelliğinin yanı sıra sabit ve minimum varyans özelliğini de gösterecek, ilgili varsayımlar da böylece sağlanmış olacaktır.²⁶

Her bir y gözlem değeri için iki olası sonuç ve bir deneme vardır. 0 halde her bir yanıt değişkeni farklı Bernoulli dağılımına sahiptir. (Bu, kukla değişkenin varyansının bulunduğu formül yardımıyla da anlaşılabilir.) N bağımsız tekrarlı yanıt değişkeni ise Binom dağılımı gösterecektir.²⁷

²⁵ Allison, **a.g.e.**, s.10.

²⁶ Tatlıdil, **a.g.e.**, s.291.

²⁷ Gülay Başarır, **Çok Değişkenli Verilerde Ayrısama Sorunu ve Lojistik Regresyon Analizi**, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Yayınlanmamış Doktora Tezi, 1990, s.8.

Lojistik regresyon, doğrusal regresyon modeli için olduğu gibi çapraz tablolar için de bir alternatiftir. Bu nedenle veri kümesinde m grup ve her grupta n_j ($j=1, 2, 3, \dots, m$) gözlem olduğunda, j'inci grupta bulunan n_j tane y_i gözlemi y_{ij} ile gösterildiğinde,

$$y_i = \sum_{i=1}^{n_j} y_{ij} \quad (26)$$

eşitliği yazılabilmektedir. Burada y_{ij} Bernoulli (P_j) ve $y \approx \text{Binom}(P_j, n_j)$ dağılımlıdır. j'inci gruptaki n_j gözlemden y. tanesinin sonucu olumlu ($y_{ij} = 1$) ise olumlu olma (ya da başarı) oranı $f_j = y_j / n_j$ olacaktır. Bu durumda tekrarlı veriler için (12) ile gösterilen doğrusal olasılık modelini, $j = 1, 2, 3, \dots, m$ için,

$$f_j = \sum_{k=0}^p \beta_k x_{ik} + \varepsilon_j \quad (27)$$

olarak yazmak mümkündür. Bu modelden elde edilecek katsayı kestirimleri de değişen varyanslılık nedeniyle yansız olacak, ancak en iyi olmayacaktır. Bu nedenle yine,

$$w_j = \frac{n_j}{f_j (1 - f_j)} \quad (28)$$

ağırlık değerlerinden yararlanılmalıdır.

İKİNCİ BÖLÜM

2. LOJİSTİK REGRESYON ANALİZİ

Çok değişkenli istatistiksel verilerin sınıflandırılması, bu verilere uygulanabilecek çeşitli istatistiksel yöntemler için gerekli bir ön analiz olmanın yanı sıra pratikte (özellikle sosyo-ekonomik konulu araştırmalarda) başlı başına bir analiz olarak da sıkça kullanılmaktadır.

Gözlemlerin gruplara ayrılmasında kullanılan yöntemlerden üç tanesi; kümeleme (cluster), diskriminant (discriminant) ve lojistik regresyon (logistic regression) analizleridir. Kümeleme analizinde gözlemlerin atanacağı küme sayısı tam bilinmezken, diskriminant ve lojistik regresyon analizinde grup (küme) sayısı bilinmekte, mevcut veriler kullanılarak bir ayırimsama modeli elde edilmekte ve kurulan bu model yardımı ile veri kümesine eklenen yeni gözlemlerin gruplara atanması mümkün olabilmektedir.

Bu üç yöntemden diskriminant analizi sık kullanılan ve çok bilinen yöntem olup varsayımları daha ağırdır. Lojistik regresyon ise son yıllarda ünlenmiş ve yoğun bir biçimde kullanılmaya başlanmıştır. Bu yöntem çeşitli varsayım (normallik, ortak kovaryansa sahip olma gibi) bozulmaları durumunda diskriminant analizi ve çapraz tablolara bir alternatif olurken, bağımlı değişkenin 0,1 gibi ikili (binary) ya da ikiden çok düzey içeren (polychotomous) kesikli değişken olması durumunda normallik varsayım kısıtı olmaması nedeniyle kullanım rahatlığının yanı sıra çözümlenmeden elde edilen modelin matematiksel olarak çok esnek olması, kolay yorumlanabilir olması yönüne olan ilgiyi artırmaktadır.²⁸ Bağımsız değişkenler, kategorik ve sürekli değişkenlerin bir karışımı olduğunda, diskriminant analizinin bağımsız değişkenlerin çok değişkenli normal dağılıma uygun olması varsayımı bozulur.²⁹

²⁸ Tatlıdil, a.g.e., s.289.

²⁹ Subhash Sharma, **Applied Multivariate Techniques**, NewYork: John Wiley&Sons, mc., 1996, s.3 17.

Çalışmalardan sağlıklı sonuçlar elde etmek için bilginin mümkün olduğu kadar ölçme hatalarından arındırılmış olması gerekmektedir. Doğru bilgiyi elde etmek için doğru soruyu sormak gerekir. Ancak yeterliliği olan verilerle yüksek güvenilirlikte sonuçlar elde edilebilir. Ölçme hatalarını zayıf sonuçlar elde etmenin tek sorumlusu olarak ta görmemek gerekir.³⁰

Aslında veriler biraz dikkatli ele alındığında ya da yapıları hakkında dikkatli düşünüldüğünde, sürekli olduğu varsayılan birçok değişkenin gerçekte kesikli olduğu ya da kesikli olarak varsayılan birçok değişkenin sürekli olduğu görülebilir. Örneğin birinin boyunu ölçerken en yakın cm ya da mm cinsinden ölçü alınır oysa bu birimler arasına sonsuz sayıda birim ilave edilebilir. Dolayısıyla sürekli gibi görülen boy değişkeni çok sayıda değer alabilen bir kesikli değişken gibi düşünülebilir. Bazen bu duruma karşıt bir bakış açısıyla yaklaşılabilir. Eğer kategoriler sıralanabilirse, araştırmacı bu değerleri sürekli bir ölçekte yer alan noktalar olarak değerlendirebilir.

Sınırlı sayıda değerler alan değişkenlere kukla ya da dummy değişken de denir. Kukla değişkenler, kalitatif değişkenleri ifade etmek için kullanılmasının yanı sıra, zaman etkisini, yerel etkiyi ve kantitatif değişkenlerin geniş gruplandırılmalarını yansıtabilmek amacıyla kullanılır.³¹

Bağımlı değişkenin kalitatif ya da kategorik bir yapı göstermesi halinde başlıca 3 durum söz konusu olmaktadır. Bunlar;

i-) İki değerli değişken (dichotomous veya binary) : Modelin bağımlı değişkenli (Y), olası iki sonuçtan hangisinin ortaya çıktığına bağlı olarak 1 veya 0 değerini alır. Örneğin i şahsının kanser olması 0 ve olmaması 1 gibi.

³⁰ F. J.Hair, R.E. Anderson, R.L. Tatham, W.C.Black, **Multivariate Data Analysis**, Prentice Hall International, s.9.

³¹ Y.İşyar, **EkonometrİK Modeller**, Uludağ Üniversitesi Güçlendirme Vakfı Yayını 92, s 189.

ii-) Çok değerli deęişken (polytomous) : Baęımlı deęişken Y, çok sayıdaki olumlu sonuçtan hangisinin ortaya çıktığına baęlı olarak deęerler alır. Örneęin i şahsı işine özel otomobil ile gider, tren ile gider, otobüs ile gider.

iii-) Kısıtlı (limited) baęımlı deęişken: Bu durum yukarıdakileri özel durum olarak kapsar. Örneęin nicel bir baęımlı deęişken bir alt veya üst veya hem alt hem de üst sınırla sınırlandırılabilir. $Y > l$, $Y < O$ gibi.

Baęımlı deęişkenler bireylerin davranışlarını ifade eden kesikli deęerler aldığıında elimizde yeterli veri olduğunda, örnek dışındaki benzer bireylerin davranışlarını tahmin etmeye (öngörü) yönelik bir denklem tahmin edilebilir.

Regresyon denkleminde olduğu gibi burada da amaç bireyi tasvir eden davranışların bir kümesi ile o bireyin verilen bir seçimi yapması olasılığı arasındaki ilişkinin bulunmasıdır. Bu ilişkiyi yansıtan seçim modelinin hem ileriye yönelik tahmin için kullanışlı hem de kolayca tahmin edilebilir bir formda olmasına özen gösterilmelidir.

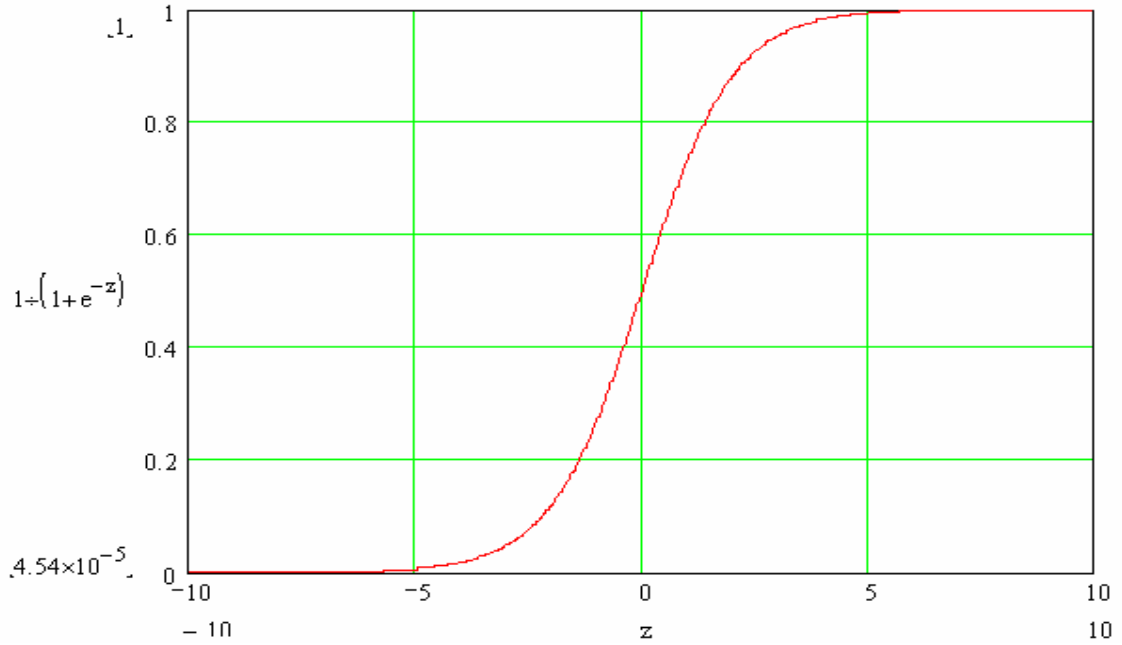
Baęımlı deęişkenin kesikli olması durumunda kullanılacak modeller çok çeşitlidir. Doğrusal Olasılık Modeli, Probit ve Lojit modeller arasında en çok tercih edilen yöntem Lojistik Regresyondur.

Lojistik regresyon, çeşitli x baęımsız deęişkenleriyle, y dikotom baęımlı deęişken arasındaki ilişkiyi tanımlamak amaçlı matematiksel modelleme yöntemidir.³² Lojistik regresyonun, bir önceki bölümde bahsedilen Doğrusal Olasılık Modeli'nden sıyrılıp öne çıkmasının sebebi ise, lojistik modelin dayandığı lojistik fonksiyondur. Lojistik fonksiyon;

$$f(z) = \frac{1}{(1 + e^{-z})} \quad (29)$$

³² David G. Kleinbaum, **Logistic Regression: A Self-Learning Text**, 1.Edition, NewYork: SpringerVerlag, 1994, s.5.

olup grafiđi Őekil 4. 'teki gibidir.



Őekil 4. Lojistik regresyon fonksiyonu

Görüldüđü gibi yatay ekseninde yer alan z 'nin tanım kümesi $(-\infty + \infty)$ dur. Buna karşılık deđer kümesi $[0,1]$ 'dir.

$Z \rightarrow -\infty$ iken lojistik fonksiyon $f(z) \rightarrow 0$ 'dır.

$Z \rightarrow +\infty$ iken lojistik fonksiyon $f(z) \rightarrow 1$ 'dir.

$f(z)$, x 'in tanım aralıđından bađımsız bir biçimde $(0,1)$ kümesinin dıŐına çıkmamaktadır. Bu, lojistik regresyon analizinin popüler olmasının çok önemli bir nedenidir. Bir olasılık belirlemek amacıyla tasarılan modellerde, y bađımlı deđiŐkeni riski göstermekte olup en önemli sorun, y bađımlı deđiŐken deđer kümesinin $(0,1)$ 'in dıŐına çıkmaması ve sonuçta y bir olasılık, risk olduđundan bunun kabul edilmez olmasıdır.³³

³³ Kleinbaum, a.g.e, s. 6.

Sonuç olarak denilebilir ki, lojistik regresyon analizini bu denli popüler yapan iki önemli nokta, lojistik fonksiyonun tahminlerin 0 ile 1 aralığında olmak zorunda olduğu durumlarda bunu sağlaması ve lojistik fonksiyonun S şeklindeki grafiği, risk faktörlerinin kombinasyonunun temsili olarak düşünülebilecek z değişkeninin $-\infty$ değerlerine gitmesi durumunda sifıra yaklaşması, Z' nin $+\infty$ değerlerine gitmesi durumunda 1 'e yaklaşmasıdır. Bunun da bağımlı değişkenin risk olasılığı olarak tanımlandığı durumlarda dikkat çekici olmasıdır.³⁴

Lojistik regresyon konusundaki en önemli kavramlardan birisi “odds oranıdır”. Olasılık, birçok insan tarafından belirli bir sonuçla ilgili o sonucun gerçekleşebilme oranı olarak bilinir.³⁵ Fakat bir olayın gerçekleşme şansını tanımlamanın farklı yolları da vardır ve odds bunlardan biridir.³⁶ Bir olayın odds' u, o olayın gerçekleşme sayısının, gerçekleşmeme sayısına oranıdır. Olasılıkla odds arasında şöyle bir bağlantı mevcuttur.

$$O = \frac{P}{1-P} \quad (30)$$

O = Olayın gerçekleşme olasılığı / Olayın gerçekleşmeme olasılığı

$$P = \frac{O}{1+O} \quad (31)$$

P [0,1] iken O [0,∞)'dur. Bununla ilgili olarak aşağıdaki değerler verilebilir:

Olasılık: 0,1 0,2 0,3 0,4 0,5 0,6 0,7 0,8 0,9

Odds: 0,11 0,25 0,43 0,67 1,00 1,50 2,33 4,00 9,00

³⁴ Kleinbaum, **a.g.e.**, s. 7.

³⁵ G. Şebnem Uralcan, **Temel Sigorta Bilgileri ve Sigorta Sektörünün Yapısal Analizi**, İstanbul, Beta Yayınevi, 2004, s.7.

³⁶ Allison, **a.g.e.**, s.11

Odds'un 1' den küçük olması, olasılığın 0,5'ten küçük olması demektir. Odds oranları çoklu karşılaştırmalarda çok kullanılır, hassastırlar ve en yaygın kullanım alanları iki dikotom değişken arasındaki ilişkinin ölçüldüğü alanlardır ve lojistik regresyon analizinde önemli bir ölçüttür.³⁷

2.1 Lojistik Regresyon Analizinin Doğrusal Regresyonla İlişkisi

Doğrusal Regresyonda bağımlı değişken sürekli iken Lojistik Regresyon bağımlı değişkenin kesikli olduğu durumlar için geliştirilmiştir ve hem parametrik modelin seçiminde hem de varsayımlarda, Doğrusal Regresyondan farklılıklar gösterir. Birinci farklılık bağımlı değişken ile bağımsız değişken arasındaki ilişkinin doğasıdır. Herhangi bir regresyon probleminde anahtar değer bağımlı değişkenin ortalama değeridir. Bu ortalama değer koşullu ortalamadır ve $E(Y/x)$ ile gösterilir.

Burada x bağımsız değişkenin değerini, Y ise bağımlı değişkenin değerini göstermektedir. $E(Y/x)$ 'i x verildiğinde (değeri bilindiğinde) Y ' nin koşullu beklenen değeri olarak ifade edilir.³⁸

Doğrusal Regresyonda bu ortalamanın x ' e göre doğrusal bir denklem ile gösterilebileceği varsayılmaktadır.³⁹

$$E(Y / x) = \beta_0 + \beta_1 \cdot x \quad (32)$$

Bu ifadeye göre x , $(-\infty, +\infty)$ aralığında değişen değerler alıyorken $E(Y/x)$ herhangi bir değeri alabilir. Kesikli veri söz konusu olduğunda beklenen değer $0 \leq E(Y/x) \leq 1$ aralığında olması beklenir. Bu koşulu sağlamak üzere yukarıda tanımlı beklenen değere bazı dönüşümler uygulanır.

³⁷ Allison, a.g.e, s.12.

³⁸ H.Ceyhan İnal, Süleyman Günay, a.g.e., s. 49.

³⁹ S. Menard, **Applied Logistic Regression Analysis**, Sage Publications, s.12.

Bu durumda kullanılabilir bir çok kümülatif fonksiyon olduğu halde bazı avantajlarından dolayı Lojistik Regresyon daha çok kullanılır ve yapılan dönüşüme de lojit dönüşümü adı verilir. Matematiksel açıdan kolaylıkla kullanılabilen, esnek bir fonksiyon olması ve kolay yorumlanabilmesi bu avantajındandır. Lojistik regresyonun formu aşağıdaki gibidir.

$$E(Y/x) = \frac{e^{\beta_0 + \beta_1 \cdot x}}{1 + e^{\beta_0 + \beta_1 \cdot x}} \quad (33)$$

Doğrusal ile Lojistik Regresyon arasındaki diğer önemli ayrım bağımlı değişkenin koşullu dağılımı ile ilgilidir. Doğrusal Regresyonda bağımlı değişkenin gözlemi $y = E(Y/x) + \varepsilon$ ile ifade edilir. Bu durumda ε değeri hata olarak adlandırılır ve gözlemin koşullu ortalamadan sapmasını ifade eder. Bu durumda en yaygın varsayım ε 'nun sıfır ortalama ve sabit varyansa ve normal dağılıma sahip olduğudur. Kesikli bağımlı değişken, verilen x değişkenine göre $y = E(Y/x) + \varepsilon$ olur. bu durumda ε için şu iki durum düşünülür.

$$y = 1 \quad \text{ise} \quad \varepsilon = 1 - E(Y/x) \quad E(Y/x) \text{ olasılığıyla} \quad (34)$$

$$y = 0 \quad \text{ise} \quad \varepsilon = - E(Y/x) \quad 1 - E(Y/x) \text{ olasılığıyla} \quad (35)$$

Buna göre bağımlı değişkenin koşullu dağılımı $E(Y/x)$ koşullu ortalaması ile verilen Binom dağılımını gösterir. Özet olarak bir regresyon analizinde Y bağımlı değişkeni kesikli ise regresyon denkleminin koşullu ortalaması 0 ve 1 değeriyle sınırlı olmalıdır. (33) denklemi ile tanımlanan model bu koşulu sağlamaktadır. Dolayısıyla hata terimlerinin dağılımı normal dağılım yerine Binom dağılımını gösterir.

2.2 İki Grup Lojistik Modeller

Daha önce verilen (14) no lu eşitliğin doğrusal olasılık modeli olduğundan, bu eşitliğin sol tarafının 0–1 arasında sınırlı olasılık değerleri aldığından ve bu değerlerin sonsuz değerler alabilen açıklayıcı değişkenlerle ilişkilendirildiğinden söz edilmişti. Açıklayıcı değişkenlerin sınırsız değerler alması nedeniyle söz konusu eşitlik her zaman sağlanamamaktadır. Böylesi bir durumla karşılaşılması için en iyi çözüm sonuç değeri olarak ifade edilen olasılık değerinin çeşitli dönüşümlerle $(-\infty +\infty)$ arasında tanımlı hale getirilmesidir. Bu amaçla geliştirilen dönüşümlerden en yaygın kullanılan iki tanesi lojit ve probit dönüşümlerdir.⁴⁰

2.2.1 Lojit Dönüşüm Ve Lojit Model

Bölüm 1.3 de bahsedildiği gibi;

$$E(Y / x) = \beta_0 + \beta_1 \cdot x$$

$$P = \beta_0 + \beta_1 \cdot x \quad (36)$$

şeklinde yazılan doğrusal olasılık modelinin sol tarafının tanım aralığının $[0,1]$ olması gerekmektedir. Çünkü denklemin sol tarafı $(y=1)$ olma olasılığıdır. DOM’ daki en büyük problem, olasılıkların $[0,1]$ arasında çıkma gerekliliği dolayısıyla sınırlandırılmış olması ancak sağ taraftaki doğrusal fonksiyonun sınırlandırılmamış olmasıdır. Uygulamada bu kısıtın sağlanamama sebebi, DOM’da eşitliğin sol tarafı $[0,1]$ arasında sınırlı olasılık değeri alırken, bu değerlerin sonsuz değerler alabilen açıklayıcı değişkenlerle ilişkilendirilmesidir. Bu problemi çözmek için olasılıklar şekillendirilmelidir. Böylece bağımsız değişkenlerin tanım aralığı ile bağımlı değişkenin tanım aralığı örtüşecektir.

Amaç yanıt değişkenin tanım aralığını $(-\infty , + \infty)$ yapmaktır. Olasılığın Odds oranına çevrilmesi, üst sınırı 1 ‘den $+\infty$ ’a çıkaracaktır. Odds oranının doğal logaritması alınır, alt sınır da 0’den $-\infty$ ’a inecektir. Bu dönüşüme lojit dönüşüm denmektedir.

⁴⁰ Tatlıdil, a.g.e., s.292.

Sonucun, doğrusal bir fonksiyonla açıklayıcı değişkenlere bağlanmasıyla da lojit model elde edilir. Lojit modelde P_i , $y_i=1$ olma olasılığıdır.

$$E(y_i) = L_i = \log (P_i / 1 - P_i) = \sum_{k=0}^p \beta_k x_{ik} + \varepsilon_j \quad (37)$$

Öncelikle burada tek bir bağımsız değişkenin olduğu basit gösterim verilmiştir. Ancak bundan sonra, bu durumu kapsayan, birden çok bağımsız değişkenin varlığını ifade eden genel gösterim kullanılacaktır. DOM'a alternatif olarak geliştirilen ve yaygın bir kullanımı olan ilk model lojit modeldir. Aslında, en baştan beri ilgilenilen P_i olasılığıdır. Modelden P_i olasılığı çekilirse lojistik fonksiyon elde edilir.

$$P_i = \left[\frac{\exp \sum_{k=0}^p B_k x_{ik}}{1 + \exp \sum_{k=0}^p B_k x_{ik}} \right] \quad (38)$$

Lojistik fonksiyonun bir diğer ifade ediliş şekli şöyledir:

$$P_i = 1 / [1 + \exp (- \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_k x_{ik})] \quad (39)$$

Lojistik fonksiyon süreklidir, x ve β değerleri ne olursa olsun, olasılık 0 ile 1 arasında herhangi bir değer alacaktır. Bağımsız değişkenin ikili kodlanması halinde lojistik regresyon modelinin alacağı değerler Tablo 2'de gösterilmiştir.

Tablo 2: Bağımsız değişken ikili kodlandığında lojistik regresyon modelinin değeri

Bağımlı değişken (Y)	Bağımsız değişken (X)	
	x=1	x=0
y=1	$P(1) = e^{\beta_0 + \beta_1} / 1 + e^{\beta_0 + \beta_1}$	$P(0) = e^{\beta_0} / 1 + e^{\beta_0}$
y=0	$1 - P(1) = 1 / 1 + e^{\beta_0 + \beta_1}$	$1 - P(0) = 1 / 1 + e^{\beta_0}$
Toplam	1	1

2.2.2 Lojistik Model Varsayımları

Lojistik model, basitçe şöyle tanımlanabilir: Lojistik model; başarı olasılığının başarısızlık olasılığına oranının logaritmasını, açıklayıcı değişkenlere doğrusal olarak bağlayan modeldir.⁴¹ Modelin çıkış noktası, esas olarak; bağımsız değişkenler kategorik ve sürekli değişkenlerin bir karışımı olduğunda, diskriminant analizinin bağımsız değişkenlerin çok değişkenli normal dağılıma uyması gerektiği varsayımının bozulması olarak kabul edilebilir.⁴² Bu durumda devreye giren lojistik regresyon analizinin açıklayıcı değişkenlerin dağılımına ilişkin bir kısıtı yoktur. Ancak bu, lojit modelin tamamıyla varsayımsız olduğu anlamına gelmez.

Bu modelde sonuç değişkenin ikili değerler alması nedeni ile hata terimi (24) denkleminde sıfır ortalamalı ve $p(1-p)$ varyanslıdır. Hata terimi bu parametrelerle binom dağılımlı olup, analiz bu teorik temele dayanmaktadır. İki grup lojistik modele ilişkin varsayımlar kısaca şöyledir:

i-) $Y \in (0,1)$

ii-) $P (y_i = 1 / x_i) = P_i$

iii-) $y_1, y_2, y_3, \dots, y_n$ değerleri istatistiksel olarak bağımsızdır,

iv-) Açıklayıcı değişkenler (x_k) birbirinden bağımsızdır⁴³,

Ayrıca, modelin sonuç değişkeninin sınırlarını genişletmek amacıyla uygulanan lojit $(P) = \log P/(1-P)$ lojit dönüşümünün bazı özellikleri de şöyle sıralanabilmektedir.

i-) P arttıkça lojit (P) 'de artar,

⁴¹ Allison, **a.g.e.**, s. 15.

⁴² Başarır, **a.g.e.**, s.10.

⁴³ Tatlıdil, **a.g.e.**, s.293.

ii-) P, 0-1 arasında değerler alırken lojit (P) tüm gerçel değerleri alır,

iii-) Eğer $P < 0.5$ ise lojit (P) < 0 ve eğer $P > 0.5$ ise lojit (P) > 0 'dır,

Bu özelliklerden üçüncüsü (gözlemlerin sınıflara atanmasında kullandığı için) çok önemlidir.

Açıklayıcı değişkenler üzerine herhangi bir kısıt getirmeyen lojistik modelde açıklayıcı değişkenlerin durumuna göre farklı modeller kullanılmaktadır. Bunlar;

i-) Açıklayıcı değişkenlerin tümü kesikli ise lojistik model (37) nolu denklemdeki eşitlikte tanımlandığı gibidir.

ii-) Açıklayıcı değişkenlerin tümü sürekli ise P ($x_1, x_2, x_3, \dots, x_p$) p açıklayıcı değişken üzerinde koşullu başarı olasılığı olmak üzere lojistik model,

$$\text{Log} \left(\frac{P.(x_1, x_2, x_3, \dots, x_p)}{1 - P.(x_1, x_2, x_3, \dots, x_p)} \right) = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} \quad (40)$$

biçiminde tanımlanmaktadır. Katsayı kestirimi daha sonra söz edilecek olan minimum lojit ki-kare yöntemi ile yapılmaktadır.

iii-) Açıklayıcı değişkenlerin bazılarının sürekli bazılarının kesikli olması durumunda, çok değişkenli frekans dağılımı; başarı durumu (olumlu durum) için $f_1(x_1, x_2, x_3, \dots, x_p)$ ve başarısızlık durumu için $f_0(x_1, x_2, x_3, \dots, x_p)$ biçiminde tanımlanmış iken lojistik model,

$$\log \left(\frac{(P f_1(x_1, x_2, x_3, \dots, x_p))}{(1 - P) f_0(x_1, x_2, x_3, \dots, x_p)} \right) = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} \quad (41)$$

olarak verilmektedir. Burada β katsayıları diskriminant fonksiyonunun katsayılarıdır ve gözlemleri f_0 ve f_1 fonksiyonlarına karşılık gelecek biçimde ayırmaktadır. Bu modelin katsayı kestirimleri ağırlıklandırılmış en küçük kareler yöntemi ile yapılmakta ve diskriminant fonksiyonunun katsayıları başlangıç değerleri olarak kullanılmaktadır.⁴⁴

Aslında lojistik model Genelleştirilmiş Lineer Modeller olarak bilinen çok geniş model ailesinin bir üyesidir. Modellerde sonuç değişkeni açıklayıcı değişkenlere doğrusal bir yapı ile bağlıdır. Bağ fonksiyonu olarak bilinen bu yapı, sonuç değişkeninin hangi fonksiyonunun açıklayıcı değişkenlerin doğrusal bir bağıntısı olduğunu verir. Çoklu bağlantı Regresyon Analizinde regresyon katsayıların yanlış tahmin edilmesine, katsayıların standart hataların artmasına ve modelin tahmin gücünün azalmasına sebebiyet verebilir. Lojistik Regresyon analizinde de benzer sorunlara yol açabilir. Bu yüzden eğer varsa, çoklu bağlantı durumunun tespit edilmesi ve gerekli düzeltme faaliyetlerinin yapılması gerekmektedir. Ancak araştırmacının amacı birimi ya da bireyi uygun sınıfa yerleştirmekse, çoklu bağlantı problemi ihmal edilebilir.⁴⁵

Problemin tespiti için değişkenler arasındaki korelasyonlara bakmak gerekir. % 99 ve üzeri korelasyon çoklu bağlantıya işaret eder. Ayrıca bağımsız değişkenlerin korelasyon matrisindeki öz değerlerine bakılarak çoklu bağlantı tespit edilebilir. Sıfıra yakın öz değerler çoklu bağlantının varlığını gösterir. Ayrıca Doğrusal Regresyonun varsayımları olan bütün ilişkili bağımsız değişkenlerin modele sokulması, ilişkisiz bağımsız değişkenlerin modelde bulunmaması, hata terimlerinin bağımsızlığı gibi varsayımlar Lojistik Regresyon için de geçerlidir.

⁴⁴ Başarır , a.g.e., s. 9.

⁴⁵ Anderson, Tatham, Black, **Multivariate Data Analysis**, s.127.

2.2.3 Lojistik Ayrımsama

Sonuç değişkenin 0, 1 değerlerine karşılık gelen H_1 ve H_2 gruplarının $(x_1, x_2, x_3, \dots, x_p)$ açıklayıcı değişkenlerine dayanılarak belirlenmesi, diskriminant analizinde olduğu gibi gözlemlerin $f_s(x_1, x_2, x_3, \dots, x_p)$ biçimindeki bir fonksiyondan (ya da fonksiyonlardan) yararlanılarak gruplara atanması ile yapılmaktadır.

Bu fonksiyona dayalı lojistik ayrımsama fonksiyonu; $s = 1, 2$ ve $B' = (\beta_0, \beta_1, \dots, \beta_p)$ iken,

$$\left(\frac{f_1(x)}{f_2(x)} \right) = \exp(B' x) \quad (42)$$

ya da logaritması alınarak,

$$\text{Log} \left(\frac{f_1(x)}{f_2(x)} \right) = (B' x) \quad (43)$$

biçiminde tanımlanmaktadır. Bu son eşitlik log olabilirlik oranı olup x' lerde doğrusaldır. Ayrıca $f_s(x)$ fonksiyonu üzerinde, tanım bölgesi içerisinde integralinin 1 olması dışında hiç bir kısıtı yoktur. Yukarıda verilen (41) eşitliğini daha genel biçimde,

$$\left(\frac{f_1(x)}{f_2(x)} \right) = \exp \{h(x, B)\} \quad (44)$$

olarak ifade etmek mümkündür. Burada h ifadesi β da doğrusal, x' lerde doğrusal olmayan bir fonksiyondur.

Lojistik ayrımsama fonksiyonu kullanılarak gözlemlerin gruplara atanması sonsal olasılıklara dayanılarak yapılmaktadır. Bu olasılıkların elde edilebilmesi için ise

β katsayılarının kestirilmesi gerekmektedir.⁴⁶ Yani, örneklemin olabilirlik fonksiyonunun belirlenmesi sorunu ile karşılaşmaktadır.

Olabilirlik fonksiyonu ise gözlemlerin örnekleme türünden etkilenmektedir. Çünkü örnekleme türleri ayırsama kuralının oluşturulmasında ve böylece gözlemleri H_1 ve H_2 gruplarından birine atamada etkili olmaktadır. Katsayı kestirimlerinin her biri için üç tür örneklemeden söz edilmektedir. Bunlar; karışık (mixture-prospective), koşullu (conditional) ve ayrı (separet-retrospective) örneklemedir⁴⁷.

Karışık örneklemede gözlemler (x, H) bileşik dağılımından örneklenmekte, yani gözlemler grup numarası bilinmeksizin rasgele çekilmektedir. Burada H grup üyeliğini gösteren değişken olup, iki grup olduğunda H_1 ve H_2 biçiminde gösterilmektedir. Koşullu örneklemede H' nin x koşulu altında dağılımı incelenmekte ve gözlemler bu koşullu dağılımdan örneklenmektedir. Biyolojik deney analizinde çok sık kullanılan bu örnekleme türüne ilişkin olabilirlik fonksiyonu öteki örnekleme türlerinin olabilirlik fonksiyonlarına temel olmaktadır.

Ayrı ya da ayırık örneklemede ise x' in H koşulu altındaki dağılımından örnekleme yapılmaktadır. Bu örnekleme türü genellikle geçmişe yönelik (retrospective) çalışmalarda kullanılmaktadır.

2.2.4 İki Grup Lojistik Modellerde Kestirim Yöntemleri

İki grup (küme-sınıf) olması durumunda lojistik modelin katsayı kestiriminde kullanılan yöntemler; en çok olabilirlik (Maximum Likelihood-ML), yeniden ağırlıklandırılmış iteratif en küçük kareler (Reweighted Iterative Least Square-RILS) ile tekrarlı veri durumunda kullanılan minimum lojit ki-kare (Minimum Logit Chi Square-MLCS) yöntemleridir.

⁴⁶ Başarır, a.g.e, s.16.

⁴⁷ Tatlıdil, a.g.e., s.295.

2.2.4.1 En Çok Olabilirlik Yöntemi

En çok olabilirlik (EÇO) tahmin yöntemi, modeldeki parametrelerin tahmini için geliştirilmiş çeşitli alternatif yöntemlerden biridir. En çok olabilirlik tahmin edicileri ve yeterli istatistikler arasında yakın bir ilişki vardır. Şöyle ki y parametresi için tek yeterli bir istatistik varsa, y 'nin en çok olabilirlik tahmin edicisi o yeterli istatistiğin bir fonksiyonu olacaktır.⁴⁸ Ayrıca EÇO tahminleme ile EKK tahminleme, bağımlı değişkenin normal dağıldığı varsayımı altında, klasik doğrusal regresyon analizinde, aynı sonuçları elde etmek için kullanılan iki farklı yöntemdir.⁴⁹

Doğrusal regresyonda, parametre tahminleri çoğunlukla EKK yöntemi ile yapılmaktadır. Doğrusal EKK, y 'nin tahmin edilen değerlerinin y 'nin gözlenen değerlerine en yakın parametre tahminleri bulmayı hedefler. Bu hedefine ulaşırken, bahsedilen yakınlığı hata kareler toplamı ile ölçer. EÇO, istatistiksel modellerin bütün çeşitlerinde kullanılabilen genel bir kestirim yöntemidir. EÇO ile EKK karşılaştırıldığında, EÇO' nun doğrusal olmayan karmaşık modellere uygulandığında doğrusal model kadar iyi sonuç verdiği görülür. Özellikle log-doğrusal modellerde, gizli değişken (gizli sınıf analizi) modellerinde kullanılır. Lojistik model de doğrusal olmayan bir model olduğundan, lojistik regresyon analizinde tahmin yöntemi olarak EÇO 'nun kullanılması tercih edilir.

Bir lojistik modelde parametrelerin tahmininde kullanılabilecek iki alternatif EÇO tahmin tekniği bulunmaktadır. Bunlar, koşulsuz yöntem ve koşullu yöntemdir. Modelde tahmin edilecek parametrelerin sayısı, gözlem sayısından küçükse koşulsuz EÇO tahmin yöntemi, büyükse koşullu EÇO tahmin yöntemi kullanılır. Uygulamada, genellikle gözlem sayısı tahmin edilmek istenen parametre sayısından büyük olmaktadır, bu sebeple koşulsuz EÇO tahmin metodu kullanılmaktadır. Koşullu EÇO tahmin yönteminin kullanılması gereken bir yerde koşulsuz EÇO tahmin yönteminin kullanılması yanlış sonuçlar verir. Koşullu EÇO ise her zaman uygundur ancak çok fazla

⁴⁸ H.Ceyhan İnal, Süleyman Günay, **a.g.e.**, s. 347.

⁴⁹ Kleinbaum, **a.g.e.**, s. 104.

matematiksel işlem gerektirir ve istatistik paket programlarında çalıştırılması uzun sürdüğünden pek kullanılmaz.

Bu sebeple en geniş kullanım sahası koşulsuz EÇO tahmin yöntemidir ve EÇO denilince bu yöntem akla gelir. Olasılık modellerinde denklemin sol tarafında bulunan ve P_i diye belirtilen olasılık, $P(y=1, x)$ 'tir. Y ' nin 0 ve 1 gibi iki değer alabileceği düşünülürse, her bir gözlemin P_i parametresiyle Bernoulli dağılacığı anlaşılır.

Ancak modelde 1 değil n gözlem vardır. Ve bu n gözlem birbirinden bağımsızdır. Bu n hacimlik örnekleme gözlemlene olasılığı gözlemlerin bağımsızlığı sebebiyle aşağıdaki gibi yazılabilir.

$$P(y / x) = \prod_{i=1}^n P_i^{Y_i} \cdot (1 - P_i)^{1-Y_i} \quad (45)$$

Bu ifade, olabirlik fonksiyonunun tanımıdır.

EÇO tahmin yöntemi, y ' yi gözlemlene olabirliğini maksimum yapan bir parametre tahmini seçmeye dayanır. Bu sebeple \hat{B} 'nın her bir değeri için olabirlik fonksiyon değeri elde edilir, bu fonksiyonu maksimum yapan β değeri, \tilde{B} olarak seçilir. Doğrusal EKK tahmin yönteminde, hata kareler toplamı hesaplanmakta ve her bir parametre tahminine göre birinci dereceden türev alınıp sifra eşitlenmektedir. Ayrıca k normal eşitlik yardımıyla \hat{B} 'lar bulunmaktadır.

EÇO tahmin yönteminde üstünde çalışılacak olan ifade olabirlik fonksiyonudur. Lojistik modelin olabirlik fonksiyonu, biraz önce verilen genel ifade P_i 'nin yerine koyulmasıyla bulunur.

$$L[(Y/x, \beta)] = \prod_{i=1}^n \left[\frac{\exp(\sum_{k=0}^p \hat{\beta}_k x_{ik})}{1 + \exp(\sum_{k=0}^p \hat{\beta}_k x_{ik})} \right]^{Y_i} \left[\frac{1}{1 + \exp(\sum_{k=0}^p \hat{\beta}_k x_{ik})} \right]^{1-Y_i} \quad (46)$$

Daha sonra, işlem kolaylığı sağlaması amacıyla logaritma alınır. Sonuçta lojistik modelin olabilirlik fonksiyonu aşağıdaki gibi çıkar:

$$\text{Log L} [(Y / x, \beta)] = \sum_{i=1}^n [Y_i \log P_i + (1 - Y_i) \cdot \log (1 - P_i)] \quad (47)$$

Bulunmaya çalışılan olabilirlik fonksiyonunu maksimize eden bu değer, bu fonksiyonun logaritmasını da maksimize edecektir. İfadenin β 'ya göre türevlerinin alınmasıyla olabilirlik denklemleri elde edilir, bu denklemlerin sıfıra eşitlenmesiyle $\hat{\beta}$ kestirim değerleri bulunur.⁵⁰

Olabilirlik eşitlikleri $\hat{\beta}$ 'da doğrusal olmadığından, tek adımda kesin çözüm elde edilmez, iteratif çözümleme gerekir.⁵¹ İteratif çözümlemede öncelikle β 'lara başlangıç değerleri verilir, ilk kestirimler bulunur. Elde edilen ilk kestirimlerde δ kadar ayarlamalar yapılır, türevler alınır ve EÇÖ kestirimleri elde edilir. δ düzeltme terimleri iterasyon değerlerini değiştirmediği zaman yakınsama sağlanmıştır, işlemler durdurulur. Az sayıda iterasyonla en iyi çözüme ulaşmak istenir. İterasyon sayısını azaltmanın en iyi yolu ise başlangıç değerinin isabetli seçimidir.⁵²

EÇÖ tahmin yönteminin popülaritesinin sebebi, büyük örnekleme tahmincinin iyi özelliklere sahip olmasıdır. EÇÖ tahmin edicileri yeterli, asimtotik etkin, asimtotik yansız ve asimtotik normaldir. Tahminlerin örnekleme dağılımının, büyük örneklerde normale yakın olması, güven aralıkları oluşturulması konusunda normal ve χ^2 dağılımlarının kullanılmasına olanak verir.⁵³

⁵⁰ Başarır, **a.g.e.**, s. 18.

⁵¹ Başarır, **a.g.e.**, s. 15.

⁵² Tatlıdil, **a.g.e.**, s.296.

⁵³ Allison, **a.g.e.**, s. 16.

2.2.4.2 Yeniden Ağırlıklandırılmış İteratif En Küçük Kareler Yöntemi

Doğrusal regresyonda bilinmeyen parametreleri bulmak için sıklıkla kullanılan yöntem en küçük karelerdir. Bu yöntemle modele göre tahmin edilen Y değerlerinin gözlemlenen değerlerden sapmalarının karesini minimize edecek β_0 ve β_1 elde edilir. En Küçük Kareler Yöntemi bilinen varsayımlar altında oldukça iyi sonuçlar verir. Ancak iş bağımlı değişkenin kesikli olması durumuna gelince En Küçük Kareler aynı varsayımları sağlamaktan uzak kalır.⁵⁴

Gruplandırılmış verilerde J grubun her birinde n_j denemeden r. başarı elde edildiğinde başarı oranı $P_j = r_j / n_j$ olarak tanımlanabilir. $\text{Var} (r_j / n_j) = P_j (1 - P_j) / n_j$ olduğundan, her binom dağılımlı gözlem için varyans değişmektedir.⁵⁵

Bu durumda lojit (r_j / n_j) 'nin açıklayıcı değişkenler üzerinde $w_j = n_j / P_j (1 - P_j)$ ağırlığı ile ağırlıklandırılmış regresyona uygulanmalıdır. Ancak w_j ağırlık değerleri de P_j 'nin bir fonksiyonu olduğu için en küçük kareler yöntemi iteratif olarak uygulanacak ve ağırlık değerleri her adımda (kestirim değerlerine bağlı olarak) yeniden elde edilecektir.

2.2.4.3 Minimum Lojit Ki-Kare Yöntemi

Ağırlıklı en küçük kareler kestirim yönteminin özel bir biçimi olan ve Berkson tarafından geliştirilen bu yöntemde, $2 \times J$ çapraz tablolarındaki beklenen ve gözlenen lojit değerleri arasındaki farktan yararlanılmaktadır. Yöntem tekrarlı veriler olması durumlarında kullanılmaktadır. Bir önceki yöntemde verilen P_j olasılığı üzerinden yapılan lojit dönüşümü, bu yöntemde sonuç değişkenini oluşturmaktadır.

⁵⁴ Menard, a.g.e, s.13

⁵⁵ Tatlıdil, a.g.e, s.296

Kestirimde kullanılan ağırlık deęerleri $n_j \cdot P_j (1 - P_j)$ olarak elde edilmektedir. Bu bilgiler ışığında yöntem, lojit deęeri olarak tanımlanan sonuç deęişkeninin, açıklayıcı deęişkenler ile (tanımlanan ağırlık deęerleri ile ağırlıklandırılmış) regresyonundan en küçük kareler kestirimlerini elde etmeye dayanmaktadır. Buradan tek adımda bulunan ağırlıklı en küçük kareler kestirimleri minimum lojit ki-kare kestirimleri adını almaktadır.⁵⁶

Kısaca deęinilen üç kestirim yöntemi dışında kullanılan bazı kestirim yöntemlerinden de söz edilebilir. Bunlardan en çok bilinenleri, iteratif olmayan en küçük kareler yöntemi ile diskriminant fonksiyonuna dayalı kestirim yöntemidir. Ancak çok özel durumlarda kullanılmaları nedeniyle çalışmada deęinilmemiştir.

Vermiş olduğumuz kestirim yöntemlerinin karşılaştırılması yapılacak olunur ise, kısaca şunlar söylenebilir. En çok olabilirlik yöntemi her zaman tutarlı, etkin ve yeterli kestirimler vermekte, ancak bu kestirimler her zaman yansız olmamaktadır. Yansızlık ve normal dağılımlılık asimptotik bir özelliktir. Doğrusal olasılık modelinin ağırlıklı en küçük kareler kestirimi ile lojistik modelin en çok olabilirlik kestirimi, varsayımlar sağlandığı sürece benzer istatistiksel özelliklere sahiptir. Tek farklılık, en çok olabilirlik yönteminde fonksiyonun doğrusal olmaması nedeniyle iteratif çözümün gerekli olmasıdır. Öte yandan minimum lojit ki-kare yönteminden de asimptotik olarak etkin ve yeterli kestirimler elde edilmektedir.

Sonuç olarak, nokta kestirimi için minimum lojit ki-kare yönteminin, çıkarsama için ise en çok olabilirlik yönteminin kullanılması önerilmektedir. Bu arada bazı Robust kestirim yöntemleri ile Bayes kestirim yöntemleri ve Kernel kestirim yöntemlerinin de bu amaçla kullanıldığını görmek mümkündür.

⁵⁶ Tatlıdil, a.g.e., s.297.

2.2.5 Uyum İyiliği Ve Sapma Ölçütleri

Modelde bulunması gereken tüm değişkenler modele alındıktan sonra, modelin yanıt değişkenini açıklamadaki etkinliğini araştırmaya uyum iyiliği kontrolü denir.⁵⁷

Dört ana durumda modelde problem oluşur:

i-) Modele logaritmik dönüşüm uygulanmıştır ve böylece lojit modele gidilmiştir. Ancak belki de logaritmik dönüşüm yerine başka bir dönüşüm kullanılmalıdır.

ii-) Model tam olarak doğru tanımlanmamış olabilir. Örneğin önemli değişkenler ya da gerekli etkileşim terimleri modele alınmamıştır. Bu noktada logaritmik dönüşüm doğru olsa bile, baştan bir hata mevcut olacaktır.

iii-) Gerekli bütün değişkenler modeldedir ancak ölçek yanlıştır.

iv-) Modelde aykırı değer vardır.

Bu dört durum varsayım bozulumu olarak adlandırılır ve bunları araştırmak için uyum iyiliği incelenir.

i-) Tüm değişkenleri içeren model ile kestirilen modele ilişkin olabilirlik oran değerlerinin farkına dayanan (hata kareler toplamına benzer) ölçütlerin ki-kare dağılacağı düşüncesinden hareketle, kurulan modelin geçerliliği sınanmaktadır. Bu yolla modele girecek açıklayıcı değişkenlere ve eklenecek karesel terimlere karar verilmektedir.

ii-) Hata terimlerinin, x değerlerine ya da olasılık değerlerine karşı çizimi ile aykırı değer (outlier) araştırması yapılmaktadır.

⁵⁷ Başarır, a.g.e., s. 23.

ii-) Hata kareler toplamı ve olabilirlik oranına dayalı R^2 türü ölçütler de model uyumunu test etmede kullanılmaktadır.

iii-) Lojistik model ayırimsama amacıyla kullanıldığında modelin doğru sınıflandırma oranı da bir uyum iyiliği ölçütüdür.

Yukarıda da belirtildiği gibi lojistik modelde normallik varsayımı kısıt olmadığı için, uyum iyiliği testlerinde öteki çok değişkenli testlerin birçoğunda olduğu gibi t ve F tablo değerleri karşılaştırma amacıyla kullanılamamakta bunlar yerine ki-kare, G^2 gibi parametrik olmayan ölçütlerden yararlanılmaktadır.⁵⁸

Ki-kare ve G^2 bilinen en basit parametrik olmayan ölçütlerdir. Çünkü, O-gözlenen, E-beklenen değerleri, OlogO ve OlogE sırasıyla gözlenen ve beklenen olabilirlikleri göstermek üzere bu ölçütler,

$$\chi^2 = \sum (O-E)^2 / E \quad (48)$$

$$G^2 = -2 \sum O \log (E / O) = 2 \sum O \log (O/E)$$

$$G^2 = -2 [\sum O \log (E) - \sum O \log (O)] = -2 [\log (E) - \log (O)] \quad (49)$$

biçiminde tanımlanmaktadır.

Bu ölçüt basittir ve çok kullanılır. Ancak çok sayıda açıklayıcı değişken olduğunda ya da açıklayıcı değişkenler arasında sürekli ölçekle ölçülmüş değişkenler olduğunda kullanılamazlar.⁵⁹

⁵⁸ Tatlıdil, a.g.e, s.298.

⁵⁹ Allison, a.g.e, s.54.

Tekrarlı veriler durumunda χ^2 ölçütü;

$$\chi^2 = \sum_{i=1}^J \frac{(y_i - n_i \cdot \hat{P}_i)^2}{n_i \cdot P_i (1 - \hat{P}_i)} \quad (50)$$

Lojistik regresyon analizinde, kurulan modelin önemliliğini test etmede ve bir anlamda modele girmesi gereken açıklayıcı değişkenleri belirlemede de yine G^2 yaklaşımı kullanılmaktadır. Bu amaçla önerilen sapma ölçütü, doymuş model; değişken sayısı kadar parametre içeren model, kestirilmiş model; sadece önemli olduğu düşünülen değişkenleri içeren model olmak üzere,

$$D = -2 \log (\text{kestirilen modelin olabilirliği}) / (\text{doymuş modelin olabilirliği}) \quad (51)$$

biçiminde tanımlanmaktadır.

C istatistiği (χ^2 istatistiği)

Temel F testine karşılık gelir. Sabit katsayı (β_0) dışında diğer bütün katsayıların sifira eşit olup olmadığını test eder.

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots \dots \dots \beta_k = 0$$

$$H_1: \text{En az biri farklıdır.}$$

Bu test istatistiği, temelde benzerlik oranı prensibine dayanır.

$$C = -2 \log \left(\frac{L_0}{L_1} \right) \quad (52)$$

L_0 : Sabit katsayı dışındaki bütün katsayıların 0 olması durumunda olabilirlik değeri.

L_1 : Verilere uydurulan tam modelin olabilirlik fonksiyonu değeri.

Temel hipotez doğru olduğunda c istatistikleri χ^2 dağılımlıdır. Dağılımın serbestlik derecesi $(k-1)$ 'dir, çünkü sabit katsayının sifira eşitliği sınanamaz.⁶⁰

Determinasyon Katsayısı (R^2):

R^2 , regresyon modeli yorumlamasında oldukça önemli bir yere sahiptir, çünkü bağımsız değişkenler tarafından açıklanan varyans oranını verir. Ortada bir açıklanamayan varyans yani değişim vardır. Açıklanamayandan yola çıkıp açıklanana varmak, mantıklıdır.

Nitel bağımlı değişkenli modellerde, hata varyansının minimizasyonuna bakılamaz. Çünkü nitel bağımlı değişkenlerde, ortalama ve varyans ayrı parametreler değildir. Örneğin Bernoulli dağılan bir değişkende, ortalama p , varyans $(p(1-p))$ 'dir. Varyans, p 'nin 0 ve 1 'e çok yakın değerlerinde minimize olur. Ancak bunun hiçbir akılcı açıklaması yoktur. Bu sebeple, hata varyansı minimizasyonu, uygun bir ölçüt değildir.⁶¹

Pseudo R^2 :

Regresyondaki R^2 'nin yerine kullanılabilir bir takım uyum iyiliği ölçütleri geliştirilmiştir. Bunlardan biri doğru sınıflandırma yüzdesidir.

$$\text{Pseudo } R^2 = \frac{c}{(N + c)} \quad (53)$$

Bu ölçütün 2 avantajı vardır: Kolay hesaplanır ve sınırları mevcuttur. (alt sınır: 0, üst sınır: 1) Fakat bu ölçüt, regresyondaki R^2 gibi, bağımsız değişken sayısından etkilenir. Serbestlik derecesi ile düzeltilmesi gerekir.⁶²

⁶⁰ John H.Aldrich, Forrest D.Nelson, **Linear Probability, Logit and Probit Models**, 1.Edition, California: Sage, 1984, s.56.

⁶¹ Aldrich ve Nelson, **a.g.e**, s.57.

⁶² Aldrich ve Nelson, **a.g.e**, s.57.

2.3 Çoklu Grup Lojistik Modeller

Bağımlı değişkenin ikiden çok düzeyli kesikli değişken olması durumunda bu değişkenin açıklayıcı değişkenler üzerindeki regresyon modeli, çoklu grup (multigroup-polycotomous) lojistik model adını almaktadır.⁶³

Çoklu grup lojistik regresyon modellerinde bağımlı değişkenin sınıflayıcı ölçme düzeyinde ölçülmüş ve en az üç kategoriye sahip olması gerekmektedir.⁶⁴ Genelde iki grup lojistik modellerin çoklu grup durumunda da kullanılabilmesi mümkündür. Örnek olarak bağımlı değişken $y = 0, 1, 2$ gibi üç düzeyli olsun. Bu durumda iki tane farklı iki grup lojistik model söz konusudur. Biri $y = 1$ ' e karşı $y = 0$ için, ötekisi ise $y = 2$ ' ye karşı $y = 0$ içindir. Yani $y = 0$ temel (referance - base) grup iken $y = 2$ ' ye karşı $y = 1$ ' i karşılaştıran lojistik fonksiyon yukarıda tanımlı iki karşılaştırmaya ilişkin lojistik fonksiyonların farklarından elde edilmektedir. Bu fonksiyonlar;

$$g_1(x) = \log \left[\frac{p(y=1/x)}{p(y=0/x)} \right] = \beta_{10} + \beta_{11}x_p + \dots + \beta_{1p}x_p \quad (54)$$

$$g_2(x) = \log \left[\frac{p(y=2/x)}{p(y=0/x)} \right] = \beta_{20} + \beta_{21}x_p + \dots + \beta_{2p}x_p \quad (55)$$

biçiminde olmaktadır.⁶⁵

Sonuç değerleri için koşullu olasılıklar üç grup durumunda $j=0,1,2$ için ,

$$P_j(x) = P(y=j/x) = \frac{\exp(g_j(x))}{\sum_{k=0}^2 \exp(g_k(x))} \quad (56)$$

genel formu ile gösterilir.

⁶³ Tatlıdil, **a.g.e**, s.304.

⁶⁴ Nuran Bayram, **Multinomial Lojistik Regresyon Analizinin İstihdamdaki İşgücüne Uygulanması**, s.61.

⁶⁵ D.W.Hosmer, Jr.S.Lemeshow, **Applied Logistic Regression**, John Wiley&Sons, s.1.

Olabilirlik fonksiyonunu oluşturmak için grup üyeliğini belirlemede üç tane iki değer alan değişkenden yararlanılmaktadır. Bu değişkenler; $y = 0$ için, $y_0 = 1, y_1 = 0, y_2 = 0$; $y = 1$ için $y_0 = 0, y_1 = 1, y_2 = 0$; $y = 2$ için $y_0 = 0, y_1 = 0, y_2 = 1$ olmak üzere gösterge (dummy) değişkenler biçiminde kodlanmaktadır. Yani y ' nin tüm değerleri için

$\sum_j y_i = 1$ iken n bağımsız gözlemler için koşullu olabilirlik fonksiyonu,

$$L(\beta) = \prod_{i=1}^n (P_0(x_i)^{y_{0i}} \cdot P_1(x_i)^{y_{1i}} \cdot P_2(x_i)^{y_{2i}}) \quad (57)$$

biçimindedir. Sonuç olarak, bağımlı (sonuç) değişken g kategorili olduğunda, grupları ikiye karşılaştıran ($g-1$) tane lojistik modele ihtiyaç duyulmaktadır. Bu durumda (56) nolu eşitlikle elde edilen olasılıkları, $j = 1, 2, 3, \dots, g$ olmak üzere,

$$P_j(x) = P(y = j / x) = \frac{\exp(\beta'_j(x))}{\sum_{k=0}^{g-1} \exp(\beta'_k(x))} \quad (58)$$

olarak g grup için yazmak mümkündür.

Bir önceki alt bölümdeki ifade biçimine paralel olarak, sonuç değişkenin g kategorisinde gözlemler, J grupta tekrar olduğunda lojistik modeli olasılıkları çok değişkenli normal dağılımlıdır. Bu olasılıklar $j = 1, 2, 3, \dots, g$ ve $i = 1, 2, 3, \dots, j$ için,

$$P_{ji}(x) = P(y = j / x_i) = \frac{\exp(\beta'_j(x_i))}{\sum_{j=1}^g \exp(\beta'_j(x_i))} \quad (59)$$

biçimindeki eşitlikten bulunmaktadır. Bu eşitliğin kullanımı ile gözlemlerin (çoklu grup durumunda) sınıflanması yapılmaktadır. Bu amaçla, H_s grup üyeliği göstergesi, $f_s(x)$ inci grup için olasılık yoğunluk fonksiyonu, H_g temel sınıf (grup) ve $f_g(x)$ de temel sınıf için olasılık yoğunluk fonksiyonu ise tüm sınıfların temel sınıfla karşılaştırılmasında $s = 1, 2, 3, \dots, g-1$ iken,

$$\frac{f_s(x)}{f_g(x)} = \exp(\beta'_s x) \quad (60)$$

eşitliği ya da bu eşitliğin logaritması alınarak elde edilen lojistik model kullanılmaktadır.

$$\text{Log} \left(\frac{f_s(x)}{f_g(x)} \right) = (\beta'_s x) \quad (61)$$

Temel sınıf olarak H_g değil de H_t gibi bir sınıfın seçilmesi durumunda da eşitlik benzer biçimde yazılmaktadır.

$$\text{Log} \left(\frac{f_s(x)}{f_g(x)} \right) = (\beta_s - \beta_t)' x ; s \neq t = 1, 2, 3, \dots, g-1 \quad (62)$$

Yukarıda da belirtildiği gibi lojistik ayımsamada, verilen son eşitliklerden yararlanılmaktadır. $H_1, H_2, H_3, \dots, H_g$ sınıflarında sıra ile $n_1, n_2, n_3, \dots, n_g$ gözlem olması durumunda ($n = \sum_{s=1}^g n_s$) bu gözlemlerin ayırsama fonksiyonunu oluşturmak için ilk olarak $\beta_s (s=1, 2, 3, \dots, g-1)$ katsayı terimleri elde edilmelidir. Diskriminant analizinde olduğu gibi g grubu birbirinden ayırmak için $(g-1)$ tane lojistik fonksiyon gerektiğinden, toplam $(g-1)$ $(p+1)$ tane katsayı kestirim değeri elde edilecektir.⁶⁶ Olabilirlik fonksiyonlarından elde edilen kestirimler yine örnekleme yöntemlerinden etkilenmektedir. Örneğin; koşullu ve karışık örnekleme türleri için sonsal olasılıklar,

$$P_s = P(H_s / x) = \frac{\exp(\beta'_s x)}{\sum_{t=1}^g \exp(\beta'_t x)} \quad (63)$$

⁶⁶ Tatlıdil, a.g.e., s.306.

eşitliğinden bulunmaktadır. Burada $B_s = (B_{os}^*, B_{1s}, \dots, B_{ps})$ biçiminde olup sabit terim için,

$$\exp(B_{os}^*) = \exp(B_{os}) \left(\frac{\tilde{P}_s}{\tilde{P}_g} \right) \quad (64)$$

eşitliğinden yararlanılmaktadır. Burada n_s ve n_g sırasıyla s ve g 'inci gruplardaki gözlem sayılan ($n = n_s + n_g$) iken $\tilde{P}_s = \frac{n_s}{n}$, $\tilde{P}_g = \frac{n_g}{n}$ bu gruplar için önsel olasılıklardır.

Sonsal olasılıklar kullanılarak bulunan olabirlik fonksiyonları yine örnekleme durumlarından etkilenmekte olup, iki grup durumundaki modelin g gruba genellemesi biçimindedir. Örneğin; koşullu örnekleme durumunda n gözlemlilik örneklemin olabirlik fonksiyonu, $s_{(i)}$, s 'inci gruptaki i 'inci gözlemi göstermek üzere,

$$L = L(X, \beta) = \prod_{i=1}^n P(H_{s(i)} / x_i) = \frac{\prod_{i=1}^n \exp(\beta'_{s(i)} x_i)}{\sum_{i=1}^g \exp(\beta'_i x)} \quad (65)$$

biçiminde tanımlanmaktadır. Bu fonksiyonu,

$$L(X, \beta) = \prod_{i=1}^n \exp \left(\sum_{t=1}^g y_{it} \log P(H_t / x_i) \right) \quad (66)$$

olarak yazmak da mümkündür. Burada i 'inci gözlemin H grubunda olup olmamasına göre 1 ya da 0 değerini almaktadır.

Verilen (65) ya da (66) eşitliklerinden yararlanılarak β_{sj} en çok olabirlik kestirim değerleri bulunmaktadır. Genel olarak, $\hat{\beta}$, β 'nın en çok olabirlik kestirimleri olduğunda çoklu grup lojistik modele dayalı ayırmsamada $s \neq t = 1, 2, 3, \dots, g$ için;

$$P(H_s / x) = \max P(H_t / x) \quad 1 \leq t \leq g \quad (67)$$

$$\text{Ya da,} \quad (\beta_s - \beta_t)' x_i > 0 \quad (68)$$

ölçütlerinden biri kullanılmaktadır. Bu koşullardan herhangi birinin denenmesi ve sağlanması durumunda i ' inci gözlem H_s grubuna atanmaktadır. Burada $s = 1, 2, 3, \dots, g-1$ için $\hat{\beta}_s' x$ vektörüne “ayırimsama skor vektörü” adı verilmektedir ⁶⁷

2.3.1 Çoklu Grup Lojistik Model Varsayımları

Çoklu grup lojit model, dikotom lojit modelin uzantısı olduğundan yeni ve farklı bir varsayım gerektirmez. Çoklu grup lojit modelin varsayımları şunlardır:

i-) k bağımsız değişkenlerin üzerinde M gözlem vardır. Bu gözlemlere göze ya da hücre terimi de denir. Bu k değişkenin 1 tanesi genellikle sabit olur. (model için sabit katsayının tahmini). K değişken, doğrusal olarak bağımlı değildir. Sonuç olarak $M \geq K$ olması istenir.

ii-) Bağımlı değişken yanıtlarının sayısı kadar değerle ölçümlenir. Her gözeye j mümkün kategoriden biri düşmektedir. J , birden büyük pozitif bir tamsayıdır.

iii-) Y' deki yanıtlar (tekrarlar) hem içlerinde hem de aralarında bağımsızdır.

iv-) X_i' nin verilen bir gözlem değeri için, Y' nin bir yanıtının olasılığı (J kategorisi için) P_{ji} için verilen denklem yardımıyla hesaplanır. ⁶⁸

⁶⁷ Tatlıdil, a.g.e. s.307.

⁶⁸ Aldrich ve Nelson, a.g.e. s.74.

2.3.2 Çoklu Grup Lojistik Modellerde Kestirim Yöntemleri

Tekrarlı olmayan veriler, tekrarlı verilerin özel bir durumu olarak düşünülebilir. Bu özel durum, her gözede sadece bir tekrar sayısının bulunmasıdır. Bu sebeple burada odak noktası tekrarlı veriler olacaktır. K bağımsız değişkenin aldığı değerleri yine X_i temsil etsin. Gözlemlenen değerlerin M tane farklı kombinasyonu olduğu varsayalım ve $i=1, 2, 3, \dots, M$ olsun. Dikotom durumda, Y 'nin alabileceği iki değer varken, politom ($y=1, 2, 3, \dots, j$) durumda, Y 'nin alabileceği J değer vardır. Dikotom durumda tahmin edilecek K parametre vardır. Politom durumda ise, daha önceden açıklandığı gibi, tahmin edilecek $K(J-1)$ parametre vardır. İstatistiksel sonuçları kanıtlamak için örneklem büyüklüğüne ihtiyaç vardır. Çünkü büyük örneklem özellikleri söz konusudur. Büyük örneklem özelliklerinin sağlanabilmesi olasılığı bağımsız değişken sayısı olan K ile arttığı gibi kategori sayısı olan J ile de artar. $K (J-1)$ 'in 50 civarını aşması pratiklik sağlar.⁶⁹

N_i , bağımlı değişkenin i . gözesindeki tekrar (gözlem) sayısıdır. Y 'nin J değerinde iken ($j=1, 2, 3, \dots, J$) N_i dışındaki gözlemlerinin sayısını N_{ji} temsil etsin.

$$N_i = \sum_{j=1}^J N_{ji} \quad (69)$$

Çoklu normal lojistik olasılıklar şu denklemler ile verilir.

$$P_{ji} \equiv P (Y=j / X_i) = \exp \frac{(\hat{\beta}'_j X_i)}{D_i} \quad (70)$$

$$D_i = \sum_{j=1}^J \exp(\hat{\beta}'_j X_i) \quad (71)$$

$$\hat{\beta}'_j X_i = \sum \hat{\beta}_{kj} X_{ik} \quad (72)$$

⁶⁹ Aldrich ve Nelson, **a.g.e.** s.74.

Tahmin edilecek bilinmeyen parametreler $\hat{\beta}_j$ 'lerdir. Modelde parametrelerin J kümesi vardır ve her küme K katsayı içerir. Bu katsayılar kümesinin hepsi tahmin edilemez. Bu J kümesinin elemanlarının sadece (J-1)'i tahmin edilir. Bu durum, dikotom durumuyla da örtüşür. Dikotom durumda aslında J=2 küme vardır. Bu kümelerin sadece biri tahmin edilir. İki düzeyli lojit modelde olduğu gibi, çok düzeyli lojit modelde de en çok kullanılan tahminleme tekniği En Çok Olabilirlik tahminleme tekniğidir.⁷⁰

2.3.2.1 En Çok Olabilirlik Yöntemi

Sonuç değişkeninin g düzeyli olduğu çoklu grup lojistik modelde açıklayıcı değişkenlerin oluşturduğu katsayılar matrisi, $n(g-1) * (g-1) (p+1)$ boyutludur ve $X=(x_1, \dots, x_n)$ asil matrisinde i'inci gözleme karşılık gelen alt matris de $(g-1)*(g-1)(p+1)$ boyutlu olacaktır. Ayrıca P, (63) denkleminde P_{si} olasılık değerinin i'inci gözlem için tanımlanmış kestirimine karşılık gelmek üzere, $r_i = y_{si} - \hat{P}_{si}$ artıkları gösterdiğinde, parametrelerin EÇO kestirimi, (64) nolu denkleminde logaritması alınarak elde edilen log olabilirlik fonksiyonunun β 'ya göre türevinin alınıp sıfıra eşitlenmesinden bulunmaktadır. Bununla birlikte, olabilirlik fonksiyonunun ikinci türev matrisi en çok olabilirlik tahmincisinin kovaryans matrisinin tahmincisini ve bilgi matrisini bulmak için kullanılır.⁷¹ Bu tahminciyi bulmak için

$$\frac{\partial \text{Log}L(x, \beta)}{\partial \beta} = X_r' \quad (73)$$

denkleminin çözümü bulunur.

⁷⁰ Allison, a.g.e, s.114.

⁷¹ Nuran Bayram, a.g.e., s.69.

2.3.2.2 Yeniden Ağırlıklandırılmış İteratif En Küçük Kareler Yöntemi

Ağırlıklandırılmış yöntemler serisi olarak ele alınan Newton-Raphson ve Fisher-skorlama (Fisher-scoring) yöntemleri genel olarak yeniden ağırlıklandırılmış en küçük kareler yöntemleri başlığı altında incelenmektedir. Newton-Raphson yönteminde, olabilirlik fonksiyonunun t 'inci adımdaki EÇO kestirimi $\beta(t)$ kullanılarak $t+1$ 'inci adımdaki kestirim değeri,

$$\beta(t+1) = \beta(t) - \left[\frac{\partial^2 \log L(X, \beta)}{\partial \beta' \partial \beta} \right]^{-1} \Bigg|_{\beta=\beta(t)} \left[\frac{\partial \log L(X, \beta)}{\partial \beta' \partial \beta} \right] \Bigg|_{\beta=\beta(t)} \quad (74)$$

eşitliğinden elde edilmektedir. Yöntemde iterasyon sayısı arttıkça ıraksama olabileceği gibi hızlı bir biçimde yakınsama da olabilmektedir. Yakınsama süresi (iterasyon sayısı) büyük ölçüde β 'nin başlangıç değerlerinin iyi kestirilmesine bağlıdır.⁷²

İkinci yöntem olan Fisher-skorlama yönteminin Newton-Raphson yönteminden farkı, yukarıdaki eşitlikte ikinci derece türevler matrisi yerine beklenen değerlerin kullanılmasıdır. Sonuç olarak kestirilen katsayılara ilişkin kovaryans matrisinin beklenen değerinin kendisine eşit olması nedeniyle aynı olan bu iki yöntemde birinci ve ikinci türev değerleri kullanılarak $t+1$ 'inci adımdaki kestirim,

$$\beta(t+1) = \beta(t) + (XV(t)X')^{-1} Xr'(t) \quad (75)$$

olarak elde edilmektedir. Burada $V(t)=\text{Var}(\beta(t))$, $r(t)$ ise $\beta(t)$ 'de bulunan artık vektörüdür.

⁷² Tatlıdil, a.g.e., s.308.

2.3.2.3 Minimum Lojit Ki-Kare Yöntemi

Veriler J grupta tekrar edildiğinde ve her grupta tekrar sayısı çok olduğunda katsayı kestirimleri, ağırlıklı EKK ya da MLKK yöntemleri ile elde edilebilmektedir.

Çoklu grup lojistik modelde, g-1 tane iki grup lojistik modelin her biri için 2.2.4.3 alt bölümünde verilen minimum lojit ki-kare yöntemi ayrı ayrı uygulanmaktadır. Ancak burada, modellerin P_{ij} olasılıklarına dayanması ve bu olasılıkların toplamının 1 olma zorunluluğu gibi nedenlerden dolayı incelenen modellerin hata terimleri ilişkili olmakta ve bu ilişkileri dikkate alan kestirim yöntemleri daha iyi sonuçlar vermekle birlikte daha karmaşık yapı göstermektedir.

Çoklu grup lojistik modellerin EÇO kestirimleri iteratif olarak hesaplanırken veri konfigürasyonları nedeniyle bazen tek bir maksimum değere ulaşamamaktadır. Bu duruma neden olan veri konfigürasyonlarının belirlenmesi, kestirim sürecinde iterasyon sayısının azaltılması açısından önemlidir. EÇO kestirimlerinin varlığını da gösteren veri konfigürasyonlarını genelde üç grupta toplamak mümkündür.

a-)Tam ve Yarı Tam Bölünme: Lojistik varsayıma dayalı sınıflandırma kuralına göre eğer $j, k=1, 2, 3, \dots, g$ ve $j \neq k$ için,

$$(\beta_j - \beta_k)' x_i > 0 \quad ; \quad i=1, 2, 3, \dots, n \quad (76)$$

ise $H_1, H_2, H_3, \dots, H_g$ gruplarını oluşturan gözlemler “tam bölünmüştür” denir ve bu durumda $\text{SuplogL}(X, \beta) = 0$ olduğu için β' nın EÇO kestirimi sonlu değildir. Eğer (76) nolu denklemi en az bir i, j, k için eşitlik biçiminde ise örneklemdaki gözlemlere yarı-tam bölünmüştür denmekte ve yine $\text{Suplog L}(X, \beta) < 0$ olduğu için EÇO kestirimi sonlu olmamaktadır.⁷³

⁷³ Tatlıdil, a.g.e., s.309.

b-) Kısmi ve Yarı-kısmi Bölünme : Örneklemde g grupta yer alan gözlemler (76) nolu eşitsizlikle ifade edildiği biçimde tam bölünmüş q tane $C_1, C_2, C_3, \dots, C_q$ kümelerine gruplanabiliyorsa bu kümelerin yanı sıra, $D_1, D_2, D_3, \dots, D_q$ gibi kümeler de bulunabilmektedir ve $q < g$ ise gözlemlere kısmi bölünmüştür denmekte ve $v = (p+1)(q-1)$ olmak üzere her bir C grubundaki katsayı kestirimleri $\gamma \in R^v$ olup, $s, t = 1, 2, 3, \dots, q, s \neq t$ ve $i = 1, 2, 3, \dots, n$ için,

$$(\gamma_s - \gamma_t)' x_i > 0 \quad (77)$$

özelliği sağlanmaktadır.

Kısmi bölünme tam bölünmenin genellemesidir. Yani tam bölünme kısmi bölünmenin $q = g$ olan özel bir durumudur. Bu nedenle ne tam ne de yarı-tam bölünme olmayan ve ara bir durumu ifade eden kısmi bölünmede β 'nin EÇÖ kestirimi sonlu değildir. Yukarıda verilen (77) eşitsizliğin en az bir i, s, t üçlüsü için eşitlik olması durumunda gözlemlere yarı kısmi bölünmüştür denmekte ve bu bölünme türü için de EÇÖ kestirimi sonlu olmamaktadır.

c-) Taşma Durumu: Veri kümesinde ne tam (yarı-tam) ne de kısmi (yarı kısmi) bölünme yoksa muhtemelen taşma durumu vardır. Taşma durumunda $j, k=1,2,3, \dots, g$ ($j \neq k$) ve $i=1, 2, 3, \dots, n$ için,

$$(\beta_j - \beta_k)' x_i < 0 \quad ; \quad i=1, 2, 3, \dots, n \quad (78)$$

eşitsizliği söz konusudur. Grupların tümünün birbiri üzerine taşmasının zorunlu olmadığı bu durumda β 'nin EÇÖ kestirimi vardır ve tektir. β 'nin EÇÖ kestiriminin sonlu (ya da tek) olabilmesi verilerin konfigürasyonlarına (bölünme durumlarına) bağlı olduğundan ve pek çok bölünme durumlarında da sonlu kestirimi olmaz. Ancak kestirimin sonlu olmaması, elde edilen sonsuz sayıda kestirimin doğru sınıflandırma yüzdesinin yüksek olması durumunda istenmeyen bir özellik değildir. Ayrıca böylesi durumlarda, çözüm kümesini daraltmak amacıyla yönelik EÇÖ kestirimlerine Begg-

Gray yaklaşımının uygulanması önerilmektedir. Bu yaklaşımın temeli; sonuç değişkeninin her bir düzeyini, temel alınan sınıfla bireysel olarak iki grup lojistik modellerle karşılaştıran ikili lojistik regresyon analiz yönteminin uygulanmasıdır.

2.3.3 Uyum İyiliği Testleri

İki grup lojistik modellerde kullanılan ki-kare ve sapma ölçütleri çoklu grup durumunda da uyum iyiliği ölçütleri olarak kullanılmaktadır. Sırasıyla;

$$\chi^2 = \sum_{i=1}^n \chi^2 = \frac{\sum_{i=1}^n \sum_{t=1}^g (y_{it} - \hat{P}_{it})^2}{\hat{P}_{it}(1 - \hat{P}_{it})} \quad (79)$$

$$D = -2 \sum_{i=1}^n d^2 = -2 \sum_{i=1}^n \sum_{t=1}^g y_{it} \log \hat{P}_{it} \quad (80)$$

biçiminde tanımlanan bu ölçütlerde $\hat{P}_{it} = (H_t / x_i)$ olup (63) nolu denklemde s yerine t konarak i'nci gözlem için elde edilen olasılık değerinin kestirimidir. Ayrıca burada y_{it} , i'nci gözlem t'inci grupta ise 1 değilse 0 değeri alan gösterge değişkendir. Çoklu grup için uyum iyiliği test ölçütleri, iki grup için önerilenlerin genellenmeleri olduğundan, çoklu grup lojistik modelde yer alan gerek tüm değişkenlerin, gerekse bunların bir alt kümesinin önemliliğine ilişkin kurulan hipotezler ve bulunan ölçüt değerlerinin yorumlanması (2.2.5) alt bölümünde anlatıldığı gibidir. Tek farklılık, buradaki D sapma ölçütünün n-p(g-1) serbestlik dereceli ki-kare dağılımı göstermesidir.

2.3.4 Değişken Ve Temel Sınıf Seçimi

Değişken Seçimi; Çoklu grup lojistik modellerde değişken seçimi konusunda genelde iki farklı grupta toplanan çok sayıda yöntemden yararlanılmaktadır. Bu gruplardan ilkinde, bir anlamda modelin uyum iyiliğinin de bir göstergesi olan; olabilirlik oran ölçütü, Wald ölçütü ve skor ölçütü gibi yöntemler yer almaktadır. İkinci grupta ise doğrusal regresyon analizinde karşılaşılan; ileriye doğru seçim, geriye doğru çıkarma, adımsal seçim, tüm olası alt kümelerden seçim gibi klasik değişken seçimi

yöntemleri bulunmaktadır. Ancak, özellikle ikinci grupta yer alan yöntemlerin genelde küçük örneklem için daha kullanışlı olması nedeniyle, büyük örneklem olması durumunda ilk grupta yer alan yöntemlerin kullanılması ya da örneklem olası alt gruplara bölünmesi veya Begg-Grayin ikili grup karşılaştırma yaklaşımından yararlanılması daha sağlıklı sonuçlar vermektedir.⁷⁴

Temel Sınıfın Seçimi; Çoklu grup lojistik regresyonda temel sınıfın seçimi önemli bir konudur. Çünkü temel sınıfın doğru belirlenmesi, daha iyi lojistik modelin elde edilmesine ve daha iyi lojistik model de yüksek performanslı ayırmsamaya neden olmaktadır. Ayrıca temel sınıfın doğru belirlenmesi genellikle modelde minimum sayıda parametre bulunmasına neden olduğu için minimal temel sınıf olarak adlandırılan temel sınıfın belirlenmesi sorunu elde edilen modelin uyum iyiliği ve değişken seçimi ile de yakından ilgilidir.

2.4 Lojistik Regresyon İle Diskriminat Analizinin Karşılaştırılması

Bir önceki bölümde, gözlemleri veri kümesinin yapısında var olan gruplardan birisine sınıflandırma amacına yönelik çok değişkenli istatistiksel analiz yöntemlerinden birinin de diskriminant analizi olduğu ve yaygın uygulandığı belirtilmiştir. Diskriminant analizinin iki temel varsayımı olan; verilerin çok değişkenli normal dağılması ve grupların ortak varyans-kovaryans matrisli olmaları, pratikte genellikle sağlanmadığı için bu yöntemin kullanımı kısıtlı olmaktadır. Varsayımların bozulmuş durumda kullanılan bazı özel diskriminant fonksiyonları olmakla birlikte yorum kolaylığı ve kestirim gücüllüğü gibi nedenlerden dolayı lojistik regresyon analizi son yıllarda daha çok kullanılmaya başlanmıştır.

$$y = b_0 + b_1 x_1 + \dots + b_p x_p \quad (81)$$

biçimindeki diskriminant fonksiyonuna ilişkin olasılık değeri,

$$P(y) = \frac{1}{(1 + e^{-y})} \quad (82)$$

⁷⁴ Tatlıdil, a.g.e., s.312.

şeklinde yazılabilmektedir ve bu değer aynı zamanda lojistik modele ilişkin olasılık değeridir. Bu nedenle birbirine çok yakın olan bu iki yöntem arasındaki ilişkinin gösterilmesi de mümkündür. Şöyle ki diskriminant analizi, y koşulu altında $P(x/y = 1)$ olasılığını x 'lerin dağılım fonksiyonu yardımı ile tanımlamaktadır.

Bayes teoreminden,

$$P(y = 1/x) = P(y = 1) \frac{P(x/y = 1)}{P(x/y = 1) + P(x/y = 0)} \quad (83)$$

yazılabilmektedir. Burada paydadaki olasılıklar, ortak varyans-kovaryans matrisli ve normal dağılımlı verilerden elde edildiğinde diskriminant analizine karşılık gelmektedir. $P(y=1/x)$ koşullu olasılığı ise lojistik model olasılıklarını tanımlamaktadır. Yani lojistik regresyon analizinde (diskriminant analizinin aksine) x koşulu altında y 'lerin dağılım fonksiyonu kullanılarak bu olasılıklar elde edilmektedir.

Diskriminant analizi için gerekli varsayımların sağlanması durumunda dahi lojistik regresyon analizinin tercih edilmesi gerektiği son yıllarda sıkça söylenmektedir. Nitekim böylesi durumlarda hangi analizin üstün olacağına ilişkin birçok çalışma yapılmıştır. Bu çalışmalardan bir tanesinde de iki grup durumu için önceki bölümde verilen (74) denkleminin benzer biçimindeki,

$$D^2 = (\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)}) \quad (84)$$

Mahalanobis uzaklığı kullanılmaktadır. Elde edilen D^2 değerine göre:

- i-) $D^2 < 2$ ise lojistik modelin diskriminant fonksiyonuna göre etkinliğinin % 90 olduğu,
- ii-) $D^2 \geq 3$ ise; $D^2 = 3$ için bu etkinliğin % 65, $D^2 = 3.5$ için % 48 ve $D^2 = 4$ için % 34 olduğu, bilgilerinden yararlanılarak analiz seçimi yapılmalıdır. ⁷⁵

⁷⁵ Tatlıdil, a.g.e., s.314.

ÜÇÜNCÜ BÖLÜM

3. SİGORTACILIK SEKTÖRÜNDE LOJİSTİK REGRESYON ANALİZİNİN UYGULANMASI

İstatistiki analiz yöntemlerinin en önemlilerinden biri şirketleri ve şirketlerin sahip oldukları ürünleri iyi tanımayı, bu ürünleri iyi sınıflandırmayı sağlayan çok değişkenli istatistiksel analiz yöntemleridir. Biz de çalışmamızda son yıllarda yıldızı parlayan ve gelişen teknikler arasında çok önemli bir yer edinmeye başlayan lojistik regresyon analizini, Türk sigorta şirketlerine uygulayarak kabul gören bazı sınıflandırmaların doğruluğunu test etmek için kullandık. Böylece şirketlerin kendilerini ve güvendikleri ürünlerini aynada daha gerçekçi olarak görmemiz sağlanabilir.

3.1 Uygulamanın Amacı ve Kapsamı

Bu uygulamada, sektörde bulunan sigorta şirketlerinin 2005 yılı üretimleri, 2004-2005 yılı pazar payları ve 2005 yılı ürün bazında ortalama satışları alınarak söz konusu şirketlerin sınıflandırılması test edilmeye çalışılmıştır. Bu sınıflandırmalar; belirli özellikleri bakımından şirketleri, tercih edilebilirlik olarak iki sınıfa ayırma, bir önceki yıla göre gelişmişlikleri bakımından ikiye ayırma ve sektör ortalama değerlerinin üzerinde veya altında kalma olarak iki sınıfa ayırma olarak tanımlanabilir. Bu sınıflandırmalarda her şeyden önce analiz sonucunda çıkacak olan sonuçların objektif ve diğer araştırmacılar tarafından da kullanılabilir olmasına dikkat edilmiştir. Lojistik regresyon analizi, bağımsız değişkenlerin kesikli değer aldığı durumlarda kullanılır ki bu çalışmada kullanılan bağımsız değişkenler poliçe üretimleri, satışları, şirketlerin pazar payları olduğundan lojistik regresyon için uygun yapı oluşturmaktadır. Ayrıca seçilen değişkenlerin bir kısmı sürekli bir kısmı kesikli olduğu için de lojistik regresyon yöntemi en uygun analiz yöntemi olarak öne çıkmaktadır. Sözü edilen bu avantajından dolayı özellikle gözlemlerin (sigorta şirketlerinin) gruplara ayrılmasında ve yeni şirketlerin uygun gruba atanmasında lojistik regresyon oldukça uygun bir yöntem olarak ortaya çıkmaktadır. Uygulamada yer alan 40 sigorta şirketinin 2005 yılı üretim, ortalama satış değerleri ve Pazar payları aşağıda Tablo 3'de gösterilmiştir.

3.2 Uygulamada Kullanılan Veriler

Tablo 3. Üretim, satış ve Pazar Payları Bazında Türk Sigorta Sektöründeki Şirketler

Şirket Adı	Toplam Police Adedi	Toplam Üretim (YTL)	Pazar Payı (2004)	Pazar Payı 2005	Satış Artışına Göre	Pazar Payı Artışına Göre	Ortalama Satışa Göre
ACIBADEM SAGLIK ve HAYAT	33718	65.039.530	1	1	1	1	0
AIG	86094	94.864.237	1	1	1	1	0
AK EMEKLILIK	68420	54.805.796	1	1	1	1	0
AK SIGORTA	1352561	516.416.648	6	7	1	1	1
AMERICAN LIFE	47842	62.207.644	1	1	1	0	0
ANADOLU	1619046	825.931.968	10	11	1	1	1
ANADOLU HAYAT ve EM.	637028	348.475.919	6	4	0	0	1
ANKARA	652802	172.418.025	2	2	1	1	1
ANKARA EMEKLILIK	20750	13.011.970	0	0	0	0	0
AVIVA	619430	142.864.405	2	2	1	1	0
AVIVA HAYAT ve EMEKLILIK	127899	157.051.091	3	2	0	0	0
AXA OYAK	2844606	759.853.945	10	10	1	0	1
BASAK	1597485	374.677.103	6	5	0	0	1
BASAK EMEKLILIK	103583	111.078.572	1	1	1	1	0
BIRLIK	357534	65.333.002	1	1	1	1	0
BIRLIK HAYAT	258614	13.777.556	0	0	1	0	0
DEMIR HAYAT	12559	20.848.000	0	0	0	0	0
FINANS	667617	161.297.781	2	2	1	1	0
FORTIS	71292	17.625.202	0	0	0	0	0
GARANTI	981585	299.720.260	3	4	1	1	1
GARANTI EMEKLILIK	82261	54.643.081	1	1	1	1	0
GENEL YASAM	27642	64.668.670	1	1	1	1	0
GENERALI	217300	59.812.976	1	1	1	0	0
GLOBAL HAYAT	2758	2.552.141	0	0	1	0	0
GUNES	2112961	446.803.780	6	6	1	1	1
GUVEN	714647	166.100.567	2	2	1	1	1
GUVEN HAYAT	10031	17.357.541	0	0	1	0	0
HUR	273075	34.281.394	0	0	1	0	0
IHLAS	456991	65.349.745	1	1	1	1	0
ISIK	248493	61.846.382	1	1	1	0	0
ISVICRE	1823184	418.324.816	6	5	1	0	1
ISVICRE HAYAT	39102	62.439.557	1	1	1	0	0
KOC ALLIANZ	1748515	636.338.556	7	8	1	1	1
KOC ALLIANZ HAYAT ve EM.	1132640	89.273.293	3	1	0	0	0
MAGDEBURGER	8	79	0	0	0	0	0
RAY	949281	200.745.994	2	3	1	1	1
RUMELI HAYAT	3497	1.241.790	0	0	1	0	0
SEKER	327953	87.215.590	1	1	0	0	0
T.GENEL	844026	206.768.492	3	3	1	0	1
T.NIPPON	716	32.588	0	0	0	0	0
TEB	169799	73.321.629	1	1	1	1	0
TICARET	18012	5.086.660	0	0	0	0	0
TOPRAK	27878	4.880.167	0	0	0	0	0
VAKIF EMEKLILIK	969241	74.967.330	1	1	1	0	0
YAPI KREDI	744343	469.737.522	6	6	1	0	1
YAPI KREDI EMEKLILIK	296541	117.037.124	2	1	0	0	0

Şirket Adı	Hazinenin Belirlediği İlk 20 Şirket	Yangın	Nakliyat	Mühendislik	Tarım	Kasko	Trafik
ACIBADEM SAGLIK ve HAYAT	0	0	0	0	0	0	0
AIG	0	11357	20109	196	0	10588	24546
AK EMEKLILIK	0	0	0	0	0	0	0
AK SIGORTA	1	176826	116085	22752	3532	265175	716502
AMERICAN LIFE	0	0	0	0	0	0	0
ANADOLU	1	248066	4440	105173	1	425219	609482
ANADOLU HAYAT ve EMEKLILIK	1	0	0	0	0	0	0
ANKARA	1	29089	26043	12208	0	123161	344267
ANKARA EMEKLILIK	0	0	0	0	0	0	0
AVIVA	1	49341	55994	15557	0	63921	240751
AVIVA HAYAT ve EMEKLILIK	1	0	0	0	0	0	0
AXA OYAK	1	214900	139576	23984	833	472538	1076938
BASAK	1	168912	67648	73156	99434	201479	409095
BASAK EMEKLILIK	1	0	0	0	0	0	0
BIRLIK	0	42498	5261	22389	0	41096	156763
BIRLIK HAYAT	0	0	0	0	0	0	0
DEMIR HAYAT	0	0	0	0	0	0	0
FINANS	1	58738	29821	47679	0	93822	333028
FORTIS	0	0	0	0	0	0	0
GARANTI	1	65806	60971	38948	211	134853	44538
GARANTI EMEKLILIK	0	0	0	0	0	0	0
GENEL YASAM	0	0	0	0	0	0	0
GENERALI	0	22911	24026	1404	0	38505	114173
GLOBAL HAYAT	0	0	0	0	0	0	0
GUNES	1	165182	108995	29289	14846	200322	751728
GUVEN	1	33702	31225	2326	85807	108176	436436
GUVEN HAYAT	0	0	0	0	0	0	0
HUR	0	4450	1887	2022	0	10768	236645
IHLAS	0	14323	10628	184541	0	37305	206798
ISIK	0	22488	19935	4427	0	30009	156491
ISVICRE	1	97598	85447	37452	10436	157743	886323
ISVICRE HAYAT	0	0	0	0	0	0	0
KOC ALLIANZ	1	123480	176551	12879	9212	216143	447065
KOC ALLIANZ HAYAT ve E.	0	0	0	0	0	0	0
MAGDEBURGER	0	1	1	1	0	0	1
RAY	1	57855	43075	16620	0	125921	331028
RUMELI HAYAT	0	0	0	0	0	0	0
SEKER	0	20853	11522	1938	4389	46449	212584
T.GENEL	1	38048	57646	4007	0	60412	616777
T.NIPPON	0	1	0	0	0	0	715
TEB	0	19128	31267	10903	0	20954	18249
TICARET	0	2992	1311	115	3	4324	8832
TOPRAK	0	1150	659	84	0	2314	23524
VAKIF EMEKLILIK	1	0	0	0	0	0	0
YAPI KREDI	1	99620	71152	5883	1	144085	331342
YAPI KREDI EMEKLILIK	1	0	0	0	0	0	0

Şirket Adı	Hazinenin Belirlediği İlk 20 Şirket	Zorunlu Karayolu Taşımacılığı	Zorunlu Koltuk Taşıma	Ferdi Kaza	Diğer Kaza	Sağlık	Hayat
ACIBADEM SAGLIK ve HAYAT	0	0	0	3050	0	19178	11490
AIG	0	0	75	7812	4551	6860	0
AK EMEKLİLİK	0	0	0	26	0	0	68394
AK SIGORTA	1	2308	2348	18128	16332	6090	6483
AMERICAN LIFE	0	0	0	17488	0	1104	29250
ANADOLU	1	2736	2525	47003	107440	66961	0
ANADOLU HAYAT ve E.	1	0	0	14974	0	0	622054
ANKARA	1	4649	7877	2362	103040	106	0
ANKARA EMEKLİLİK	0	0	0	7181	0	0	13569
AVIVA	1	455	1287	4220	179464	4220	4220
AVIVA HAYAT ve EMEKLİLİK	1	0	0	21669	0	0	106230
AXA OYAK	1	2606	2033	642680	14629	9225	244664
BASAK	1	504	319	209459	353655	13824	0
BASAK EMEKLİLİK	1	0	0	44019	0	0	59564
BIRLIK	0	1131	898	43541	43460	497	0
BIRLIK HAYAT	0	0	0	0	0	0	258614
DEMİR HAYAT	0	0	0	1258	0	2521	8780
FINANS	1	981	1094	99655	2799	0	0
FORTIS	0	0	0	25079	0	0	46213
GARANTI	1	550	156	249959	384682	911	0
GARANTI EMEKLİLİK	0	0	0	82261	0	0	0
GENEL YASAM	0	0	0	507	0	21772	5363
GENERALI	0	409	160	2943	12649	39	81
GLOBAL HAYAT	0	0	0	1520	0	233	1005
GUNES	1	2280	2657	376132	436404	21895	3231
GUVEN	1	4284	2928	5529	4234	0	0
GUVEN HAYAT	0	0	0	91	0	1906	8034
HUR	0	117	62	687	16437	0	0
IHLAS	0	377	1277	896	846	0	0
ISIK	0	0	423	13770	808	142	0
ISVICRE	1	4136	10906	158657	374486	0	0
ISVICRE HAYAT	0	0	0	2379	0	19643	17080
KOC ALLIANZ	1	900	590	295765	412386	53544	0
KOC ALLIANZ HAYAT ve EM.	0	0	0	1285	0	0	1131355
MAGDEBURGER	0	0	0	0	1	2	0
RAY	1	0	529	139071	235182	0	0
RUMELI HAYAT	0	0	0	0	0	0	3497
SEKER	0	422	494	19168	614	652	8868
T.GENEL	1	565	1307	2418	62845	1	0
T.NIPPON	0	0	0	0	0	0	0
TEB	0	45	7	21592	47652	2	0
TICARET	0	3	1	183	245	3	0
TOPRAK	0	0	24	55	68	0	0
VAKIF EMEKLİLİK	1	0	0	1326	0	0	967915
YAPI KREDİ	1	1076	1396	3215	3415	83158	0
YAPI KREDİ EMEKLİLİK	1	0	0	64815	0	0	231726

KAYNAK: MİLLİ REASÜRANS T.A.Ş. YILLIKLARI

3.3 Şirketlerin 2005 Yılı Üretimleri Baz Alınarak Uygulanan Lojistik Regresyon Analizi

Aşağıda kurulan modelle Türkiye’de faaliyet gösteren 46 şirketin 2005 yılında yapmış oldukları poliçe üretim sayılarını lojistik regresyon analiziyle sınıflandırarak şirketlerin kendi aralarındaki performanslarını karşılaştırmaya çalışacağız.

Lojistik regresyon analizinin giriş yöntemini kullanarak adımsal olarak doğru çözüme ulaşmaya çalışacağız. Başlangıç çözümü aşağıdaki gibidir.

Tablo 4. Seçilen Gözlemler

Model Özeti		N	Yüzde
Ağırlıklandırma ^a			
Modele Seçilen	Analize Giren	46	100,0
	Kayıp Gözlem	0	,0
	Toplam	46	100,0
Modele Girmeyen		0	,0
Toplam		46	100,0

a. Modele giren gözlemler

Yukarıdaki tablo incelendiğinde eldeki 46 şirket verisinin hepsinin gözlemlenmiş olduğu görülür. Analizde kayıp gözlem yoktur. Bağımlı değişken olarak tanımlanan hazinenin belirlemiş olduğu en iyi 20 şirket (1) diğer şirketler (0) olarak kodlanmıştır.

Tablo 5. Başlangıç Adımı Sınıflandırma Tablosu

Gerçek			Tahmini		Doğru Sınıflandırma
			HAZINE		
			0	1	
Adım 0	HAZINE	0	26	0	100,0
		1	20	0	,0
Toplam yüzde					56,5

a. Sabit değişken modele alındı

b. Modele değişkenlerin alım oranı 0,5 olarak belirlendi

Başlangıç çözümüne bakarsak başarısız gözükten şirketlerin tamamının doğru atandığı fakat buna karşın başarılı gözükten şirketlerin tamamının da yanlış atandığı gözükmektedir. Bunun ana sebebi gerekli değişkenlerin modele alınmaması olarak gösterilebilir ve diğer aşamaya geçmeye karar verilir.

Tablo 6. Başlangıç Adımında Denkleme Giren Değişkenler

Denkleme Alınan Değişkenler						
	B	S.E.	Wald	df	Sig.	Exp(B)
Adım 0 sabit	-,262	,297	,778	1	,378	,769

Görüldüğü üzere başlangıç çözümünde oluşan denkleme sadece sabit katsayı alınmıştır.

Tablo 7. Başlangıç Adımında Denkleme Girmeyen Değişkenler

Denkleme Alınmayan Değişkenler			Skor	df	Sig.
Adım 0	Değerler	YANGIN	16,222	1	,000
		NAKLIYAT	15,755	1	,000
		MÜHENDİS	1,961	1	,161
		TARIM	3,832	1	,050
		KASKO	16,572	1	,000
		TRAFİK	17,124	1	,000
		Z.K.YOLU	13,143	1	,000
		Z.KOLTUK	8,739	1	,003
		FERDI.K	9,449	1	,002
		DIĞER.K	12,142	1	,000
		SAĞLIK	3,830	1	,050
		HAYAT	,530	1	,467
	Toplam istatistik		24,953	12	,015

Tablo 7' den görülebileceği gibi başlangıç adımında oluşan denkleme hiçbir değişken alınmamıştır. Buradan da anlaşılacağı üzere başlangıç çözümü yetersizdir. Doğru çözüme ulaşmak için diğer adıma geçip analize devam edilmelidir. Birinci adımın çözümü aşağıdaki gibidir.

Tablo 8. Anlamlılık Testi

		Model Katsayıları		
		ki-kare	df	Sig.
adım	adım	43,450	12	,000
1	Blok	43,450	12	,000
	Model	43,450	12	,000

Yukarıdaki tablo incelendiğinde adım 1'deki anlamlılık(sig.) değerine bakarsak $0 < 0,05$ olduğundan kurulan lojistik model anlamlıdır.

Tablo 9. Model Özeti

Model Özeti			
adım	-2 Log likelihood	Cox & Snell R kare	Nagelkerke R kare
1	19,534	,611	,820

Modele ait Nagelkerke $R^2 = 0,82$ olarak bulunmuştur. Yani kurulan lojistik modelin kullanılan değişkenlerce açıklanma oranı %82'dir. Yine Cox ve Snell R^2 de 0,611 olarak bulunmuştur. Bu açıklama oranlarının yeterince yüksek ve modelin anlamlı olduğu, ayrıca dışarıda kalan değişkenlerin çok fazla önemli olmadığı, modele alınan değişkenlerin yeterli olduğu söylenebilir.

Tablo 10. Sonuç Sınıflandırma Tablosu

Gerçek Değerler			Tahmini Değerler		Doğru sınıflandırma
			HAZINE		
			0	1	
adım	HAZINE	0	24	2	92,3
1		1	3	17	85,0
	toplam istatistik				89,1

a. Modele değişkenlerin alım oranı 0,5 olarak belirlendi

1. adımdan sonra kurulan denkleme göre ilk 20 ye giremeyen 26 şirketin 24 ünün doğru atandığını fakat 2 sinin yanlış atandığını görebiliriz. Yine ilk 20 şirket içerisinde gösterilen 3 şirketin yanlış atandığını, söz konusu şirketlere ilişkin istatistikler incelendiğinde en iyi diyebileceğimiz şirket sayısının sadece 17 olduğunu söyleyebiliriz. Tahmini değerlere de bakarak hazinenin ilk 20 şirket içerisinde görmüş olduğu Aviva Hayat Emeklilik şirketi, Anadolu Hayat Emeklilik şirketi ve Başak Emeklilik şirketi analiz sonucunda en iyi şirketler arasından çıkarılmıştır. Garanti Emeklilik ve Koç Allianz Hayat Emeklilik şirketleri de en iyi şirketler sınıfına dahi edilmiştir. Bu değerlendirme şirketlerin 2005 yılı poliçe üretimi ve satışlarına bakarak yorumlanmıştır. Sonuç olarak ta kurulan bu lojistik analizin %90 a yakın bir doğrulukla sınıflandırmaları doğru yaptığını söyleyebiliriz.

Tablo 11. Denkleme Giren ve Girmeyen Değişkenler

Denkleme Alınan Değişkenler

	B	S.E.	Wald	df	Sig.	Exp(B)
Adım # YANGIN	-,000155	,000202	,588090	1	,44316	,999845
NAKLIYAT	,000041	,000080	,266367	1	,60578	1,000041
MÜHENDİS	-,000058	,000094	,384406	1	,53525	,999942
TARIM	-,000021	,000135	,023440	1	,87832	,999979
KASKO	,000099	,000129	,583289	1	,44503	1,000099
TRAFİK	-,000010	,000026	,148944	1	,69955	,999990
Z.K.YOLU	-,003287	,006315	,270849	1	,60276	,996719
Z.KOLTUK	,007632	,011613	,431941	1	,51104	1,007661
FERDİ.K	,000060	,000027	5,065321	1	,02441	,999845
DIĞER.K	,000024	,000056	,186629	1	,66574	1,000041
SAĞLIK	,000021	,000083	,066966	1	,79581	,999942
HAYAT	,000004	,000002	5,687184	1	,01709	,999979
Constant	-3,810457	1,205345	9,993826	1	,00157	,221380

a. 1. Adımdan sonra denkleme giren değişkenler: YANGIN, NAKLIYAT, MÜHENDİS, TARIM, KASKO, TRAFİK, Z.K.YOLU, Z.KOLTUK, FERDİ.K, DIĞER.K, SAĞLIK, HAYAT.

1.adımdan sonra kurulan denkleme tüm değişkenler alınmıştır ki buda kurulan lojistik modelin yeterliliğini gösterir. Analiz sonucunda çıkan denklem aşağıdaki gibidir.

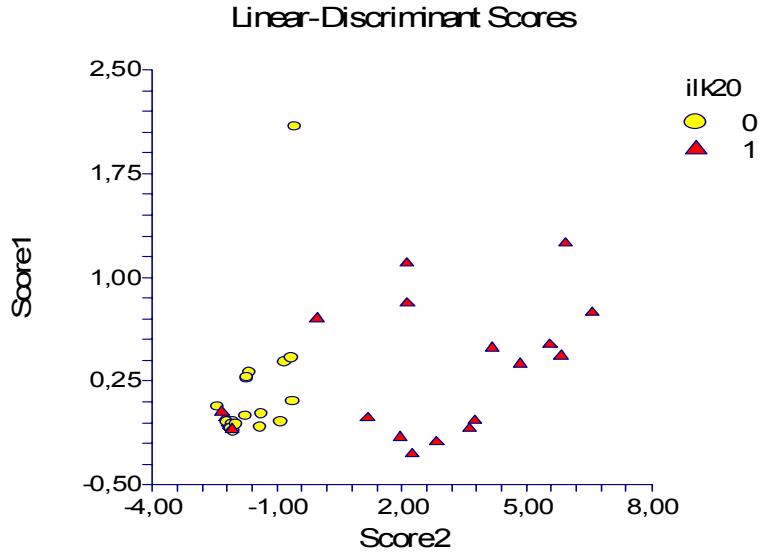
$$Y = -3,810457 - 0,000155X_1 + 0,000041X_2 - 0,000058X_3 - 0,000021X_4 + 0,000099X_5 - 0,000010X_6 - 0,003287X_7 + 0,007632X_8 + 0,000006X_9 + 0,0000243X_{10} + 0,0000214X_{11} + 0,0000044X_{12}$$

Diğer bir ifade ile lojistik fonksiyonun %90 doğruluk oranı vermesi ve seçilen değişkenlerde bir şart aramaması nedeniyle lojistik regresyon analizinin diskriminant analizine tercih edilmesi doğru bir karar olmuştur.

Tablo 12. Korelasyon Matrisi

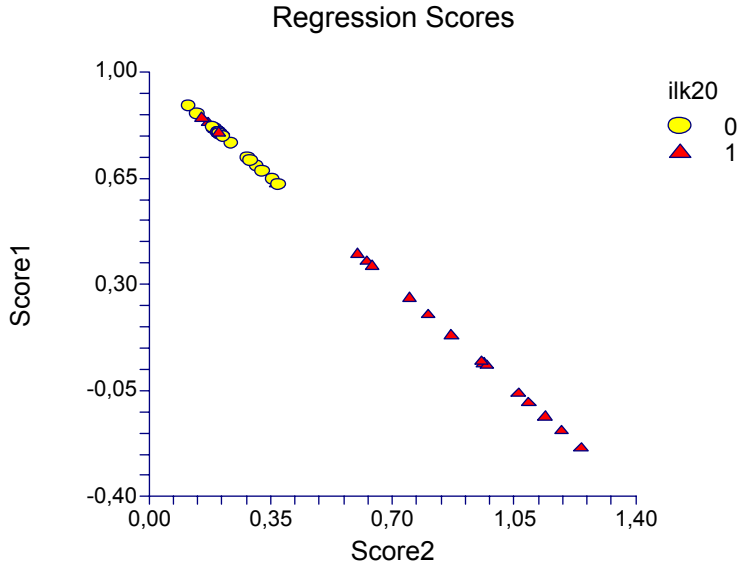
		Korelasyon Matrisi												
		SA BİT	YA NGI N	NA KLI YA T	MÜ HE NDİ S	TA Rİ M	KA SK O	TR AFI K	Z. K.Y OL U	Z.K OLT UK	FER Dİ.K	DIĞ ER. K	SA ĞLI K	HA YAT
Adım 1	SABİT	1,00	,158	-,207	,092	,165	-,101	,013	,007	-,065	-,708	-,102	-,205	-,672
	YANGIN	,158	1,00	-,481	,099	,310	-,944	,591	,219	-,428	-,218	-,688	,105	-,088
	NAKLIYAT	-,207	-,481	1,0	,244	-,07	,325	-,29	,111	,038	,170	,119	,009	,136
	MÜHENDİSLİK	,092	,099	,244	1,00	,181	-,258	,389	,309	-,414	-,271	-,452	,278	-,027
	TARIM	,165	,310	-,068	,181	1,0	-,321	,088	-,18	,047	-,223	-,560	,001	-,093
	KASKO	-,101	-,944	,325	-,258	-,32	1,0	-,66	-,36	,494	,183	,708	-,264	,048
	TRAFİK	,013	,591	-,290	,389	,088	-,659	1,0	,707	-,908	-,114	-,645	,463	,007
	Z.K.YOLU	,007	,219	,111	,309	-,18	-,362	,707	1,0	-,872	-,071	-,303	,406	,005
	Z.KOLTUK	-,065	-,428	,038	-,414	,047	,494	-,91	-,87	1,000	,135	,536	-,420	,030
	FERDİK	-,708	-,218	,170	-,271	-,22	,183	-,11	-,07	,135	1,000	,150	,133	,390
	DIĞER.K	-,102	-,688	,119	-,452	-,56	,708	-,64	-,30	,536	,150	1,00	-,209	,056
	SAĞLIK	-,205	,105	,009	,278	,001	-,264	,463	,406	-,420	,133	-,209	1,0	,138
	HAYAT	-,672	-,088	,136	-,027	-,09	,048	,007	,005	,030	,390	,056	,138	1,00

Ayrıca korelasyon matrisine bakarak değişkenlerin kendi aralarında ilişkiye sahip olup olmadıklarını inceleyebiliriz. Örneğin tabloya bakarak yangın sigortası poliçe satışıyla kasko sigortası poliçe satışı arasında negatif yönde çok güçlü bir ilişki (-0,994), fakat yangın sigortası poliçe satışıyla trafik sigortası poliçe satışı arasında pozitif yönlü vasat bir ilişki (0,591) , yine kasko sigortasıyla trafik sigortası arasında negatif yönlü güçlü bir ilişki (-0,66) olduğunu görebiliriz. Benzeri şekilde sağlık sigortası poliçe satışıyla hayat sigortası poliçe satışı ve tarım sigortası poliçe satışı arasında pozitif yönde çok zayıf bir ilişki (sırasıyla 0,138 ve 0,001) olduğunu görebiliriz. Zorunlu sigortalara bakarsak ta trafik sigortası poliçe satışı arasında zorunlu karayolu taşımacılığı sigortası poliçe satışıyla pozitif yönlü güçlü bir ilişki (0,707) olduğu fakat zorunlu koltuk taşıma sigortası poliçe satışıyla çok güçlü negatif yönlü bir ilişkinin (-0,908) var olduğunu görebiliriz.



Şekil 5. Lineer Diskriminat Analizi ile Sınıflandırma

Diskriminant analizi ile oluşturulan grafikte şirketlerin kendi aralarında nasıl gruplandıklarını çok rahat görebiliyoruz. Kırmızı nokta ile gördüğümüz şirketler ilk 20 ye girebilen şirketlerdir.



Şekil 6. Regresyon Analizi ile Sınıflandırma

Regresyon analiziyle oluşturulan trend ile şirketlerin nasıl bir eğilim gösterdiğini açıkça görebiliyoruz. Kırmızı nokta ile gördüğümüz şirketler ilk 20 ye girebilen ve diğer şirketlerden ayrılan şirketlerdir.

3.4 Şirketlerin 2004 – 2005 Yılı Pazar Payları Baz Alınarak Uygulanan Lojistik Regresyon Analizi

Aşağıda kurulan modelle Türkiye’de faaliyet gösteren 46 şirketin 2004 yılında pazarda buldukları yerle 2005 yılındaki pazarda buldukları yeri kıyaslayarak şirketlerin bir yıl içerisinde gösterdikleri gelişmişliği lojistik regresyon analiziyle sınıflandırarak şirketlerin kendi aralarındaki performanslarını karşılaştırmaya ve lojistik regresyon analizinin giriş yöntemini kullanarak adimsal olarak doğru çözüme ulaşmaya çalışacağız. Başlangıç çözümü aşağıdaki gibidir.

Tablo13. Seçilen Gözlemler

Model Özeti			N	Yüzde
Ağırlıklandırma ^a				
Modele Seçilen	Analize Giren		46	100,0
	Kayıp Gözlem		0	,0
	Toplam		46	100,0
Modele Girmeyen			0	,0
Toplam			46	100,0

a. Modele giren gözlemler

Yukarıdaki tablo incelendiğinde eldeki 46 şirket verisinin hepsinin gözlemlenmiş olduğu görülür. Analizde kayıp gözlem yoktur. Bağımlı değişken, bir önceki yıla göre Pazar payını artırmış olan şirketler (1) Pazar payı aynı kalan ya da azalan şirketler ise (0) olarak kodlanmıştır.

Tablo 14. Başlangıç Adımı Sınıflandırma Tablosu

Gerçek			Sınıflandırma Tablosu ^{a,b}		
			Tahmini		Doğru Sınıflandırma
			PAZAR.P		
			0	1	
Adım 0	PAZAR.P	0	27	0	100,0
		1	19	0	,0
Toplam yüzde					58,7

a. Sabit değişken modele alındı

b. Modele değişkenlerin alım oranı 0,5 olarak belirlendi

Başlangıç çözümüne bakarsak başarısız gözükten şirketlerin tamamının doğru atandığı fakat buna karşın başarılı gözükten şirketlerin tamamının da yanlış atandığı gözükmektedir. Bunun ana sebebi gerekli değişkenlerin modele alınmaması olarak gösterilebilir. Doğru çözüme ulaşmak için diğer adıma geçip analize devam edilmelidir. Birinci adımın çözümü aşağıdaki gibidir.

Tablo 15. Başlangıç Adımında Denkleme Giren Değişkenler

Denkleme Alınan Değişkenler						
	B	S.E.	Wald	df	Sig.	Exp(B)
Adım 0 Sabit	-,351	,299	1,377	1	,241	,704

Görüldüğü üzere başlangıç çözümünde oluşan denkleme sadece sabit katsayı alınmıştır.

Tablo 16. Başlangıç Adımında Denkleme Giren Değişkenler

Denkleme Alınmayan Değişkenler					
			Skor	df	Sig.
Adım 0	Değerler	YANGIN	2,895	1	,089
		NAKLIYAT	2,633	1	,105
		MÜHENDIS	4,919	1	,027
		TARIM	,091	1	,763
		KASKO	2,908	1	,088
		TRAFİK	1,350	1	,245
		Z.K.YOLU	3,999	1	,046
		Z.KOLTUK	1,139	1	,286
		FERDI.K	,735	1	,391
		DIĞER.K	3,800	1	,051
		SAĞLIK	1,208	1	,272
		HAYAT	3,418	1	,065
	Toplam istatistik		20,843	12	,053

Tablo 16'dan görüleceği gibi başlangıç adımında oluşan denkleme hiçbir değişken alınmamıştır. Buradan da anlaşılacağı üzere başlangıç çözümü yetersizdir. Doğru çözüme ulaşmak için diğer adıma geçip analize devam edilmelidir. Birinci adımın çözümü aşağıdaki gibidir.

Tablo 17. Anlamlılık Testi

		Model Katsayıları		
		ki-kare	df	Sig.
Adım	Adım	36,158	12	,000
1	Blok	36,158	12	,000
	Model	36,158	12	,000

Yukarıdaki tablo incelendiğinde adım 1'deki anlamlılık değerine (sig.) bakarsak $0 < 0,05$ olduğundan kurulan lojistik model anlamlıdır.

Tablo 18. Model Özeti

Model Özeti			
Adım	-2 Log likelihood	Cox & Snell R kare	Nagelkerke R kare
1	26,213	,544	,733

Modele ait Nagelkerke $R^2 = 0,73$ olarak bulunmuştur. Yani kurulan lojistik modelin kullanılan değişkenlerce açıklanma oranı %73'dir. Ayrıca Cox ve Snell R^2 de 0,54 olarak bulunmuştur ki bu açıklama oranlarının yeterince yüksek olduğu ve kurulan modelin anlamlı olduğu söylenebilir.

Tablo 19. Sonuç Sınıflandırma Tablosu

Gerçek Değerler			Tahmini Değerler		
			PAZAR.P		Doğru sınıflandırma
			0	1	
adım 1	PAZAR.P	0	26	1	96,3
		1	4	15	78,9
toplam istatistik					89,1

a. Modele değişkenlerin alım oranı 0,5 olarak belirlendi

1. adımdan sonra kurulan denkleme göre pazar payı azalmış ya da aynı kalmış olarak atadığımız bir şirketin analize göre yanlış atandığı aslında bu şirketin pazar payının arttığı söylenebilir. Yine pazar payı artmış olarak ayrılan 4 şirketin de aslında pazar payının artmadığı söylenebilir. Buna göre Acıbadem Sağlık ve Hayat, Başak Emeklilik, Güven Hayat, Ak Emeklilik şirketlerinin pazar paylarının bir önceki yıla göre artış göstermediği, Generali şirketinin pazar payının ise bir önceki yıla göre artış gösterdiği söylenebilir. Açıktır ki bu yorumlar şirketlerin 2004–2005 yılı Pazar payına bakarak söylenmektedir. Sonuç olarak kurulan bu lojistik analizin %89 civarı bir doğruluk içerdiği söylenebilir.

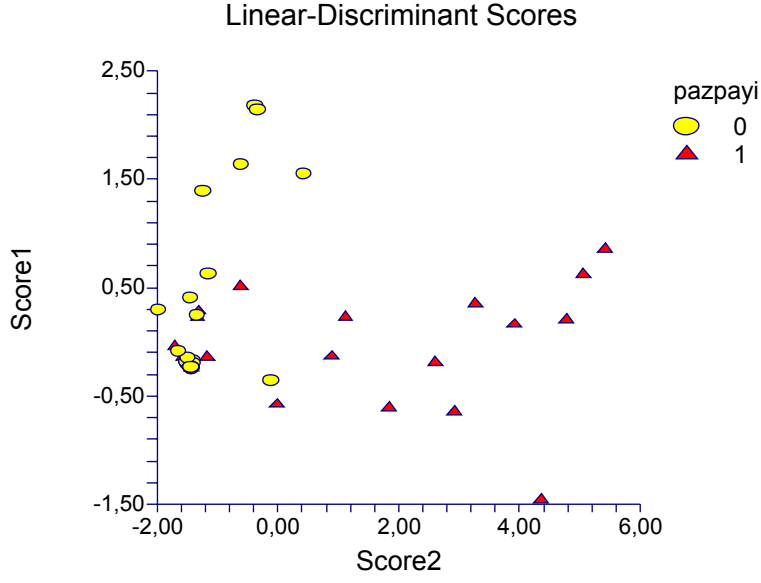
Tablo 20. Denkleme Giren Değişkenler

Denkleme Alınan Değişkenler						
	B	S.E.	Wald	df	Sig.	Exp(B)
Adım 1						
YANGIN	-,000030	,000289	,011115	1	,916038	,999970
NAKLIYAT	,000182	,000210	,748426	1	,386975	1,000182
MÜHENDİS	,000559	,000734	,579208	1	,446623	1,000559
TARIM	-,000463	,000291	2,530304	1	,111679	,999537
KASKO	-,000151	,000129	1,357473	1	,243976	,999849
TRAFİK	-,000017	,000024	,481037	1	,487953	,999983
Z.K.YOLU	,020714	,011224	3,405800	1	,064968	1,020930
Z.KOLTUK	-,009416	,005029	3,505028	1	,061183	,990628
FERDI.K	,000034	,000027	1,590207	1	,207297	1,000034
DIĞER.K	,000053	,000038	1,971762	1	,160261	1,000053
SAĞLIK	,000060	,000067	,807531	1	,368851	1,000060
HAYAT	-,000009	,000011	,603900	1	,437094	,999991
Constant	1,602278	,836985	3,664719	1	,055576	,201437

a. 1. Adımdan sonra denkleme giren değişkenler: YANGIN, NAKLIYAT, MÜHENDİS, T/KASKO, TRAFİK, Z.K.YOLU, Z.KOLTUK, FERDI.K, DIĞER.K, SAĞLIK, HAYAT.

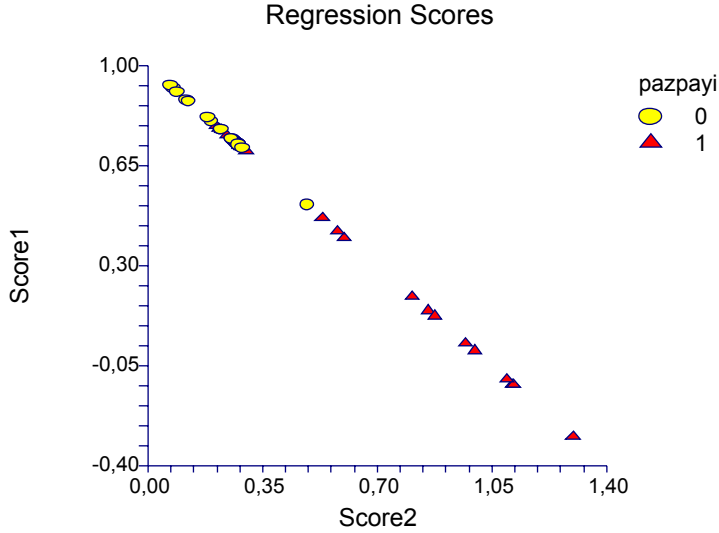
1. adımdan sonra kurulan denkleme tüm değişkenler alınmıştır ki buda kurulan lojistik modelin yeterliliğini gösterir. Analiz sonucunda çıkan denklem aşağıdaki gibidir.

$$Y = -1,602278 - 0,000305X_1 + 0,0001816X_2 - 0,0005585X_3 - 0,00046X_4 - 0,00015X_5 - 0,00017X_6 + 0,0207X_7 - 0,009416X_8 + 0,0000339X_9 + 0,0000529X_{10} + 0,0000602X_{11} - 0,000009X_{12}$$



Şekil 8. Linear Diskriminat Analizi ile Sınıflandırma

Diskriminant analizi ile oluşturulan grafikte şirketlerin kendi aralarında nasıl gruplandıklarını çok rahat görebiliyoruz. Kırmızı nokta ile gördüğümüz şirketler Pazar paylarını bir önceki yıla göre arttırabilen şirketlerdir.



Şekil 9. Regresyon Analizi ile Sınıflandırma

Regresyon analiziyle oluşturulan trend ile şirketlerin nasıl bir eğilim gösterdiğini açıkça görebiliyoruz. Kırmızı nokta ile gördüğümüz şirketler Pazar payı bir önceki yıla göre artış gösteren şirketlerdir.

3.5 Şirketlerin 2005 Yılı Ortalama Satışları Baz Alınarak Uygulanan Lojistik Regresyon Analizi

Aşağıda kurulan modelle Türkiye’de faaliyet gösteren 46 şirketin 2005 yılında yapmış oldukları ortalama satışlarını lojistik regresyon analiziyle sınıflandırarak şirketlerin kendi aralarındaki performanslarını karşılaştırmaya çalışacağız. Lojistik regresyon analizinin giriş yöntemini kullanarak adımsal olarak doğru çözüme ulaşmaya çalışacağız. Başlangıç çözümü aşağıdaki gibidir.

Tablo 21. Seçilen Gözlemler

Model Özeti

Ağırlıklandırma ^a	N	Yüzde
Modele Seçilen		
Analyze Giren	46	100,0
Kayıp Gözlem	0	,0
Toplam	46	100,0
Modele Girmeyen	0	,0
Toplam	46	100,0

a. Modele giren gözlemler

Yukarıdaki tablo incelendiğinde eldeki 46 şirket verisinin hepsinin gözlemlenmiş olduğu görülür. Analizde kayıp gözlem yoktur. Bağımlı değişken, şirketlerin 2005 yılı satışlarını ortalaması olup bu ortalamanın üzerinde yer alan şirketler (1) ortalamanın altında kalan şirketler ise (0) olarak kodlanmıştır

Tablo 22. Başlangıç Adımı Sınıflandırma Tablosu

Sınıflandırma Tablosu^{a,b}

Gerçek			Tahmini		
			ORT.SATŞ		Doğru Sınıflandırma
			0	1	
Adım 0	ORT.SATŞ	0	32	0	100,0
		1	14	0	,0
Toplam yüzde					69,6

a. Sabit değişken modele alındı

b. Modele değişkenlerin alım oranı 0,5 olarak belirlendi

Başlangıç çözümüne bakarsak başarısız gözükten şirketlerin tamamının doğru atandığı fakat buna karşın başarılı gözükten şirketlerin tamamının da yanlış atandığı gözükmektedir. Bunun ana sebebi gerekli değişkenlerin modele alınmaması olarak gösterilebilir. Doğru çözüme ulaşmak için diğer adıma geçip analize devam edilmelidir. Birinci adımın çözümü aşağıdaki gibidir

Tablo 23. Başlangıç Adımında Denklem Giren Değişkenler

Denkleme Alınan Değişkenler							
		B	S.E.	Wald	df	Sig.	Exp(B)
Adım 0	Sabit	-,827	,320	6,656	1	,010	,438

Görüldüğü üzere başlangıç çözümünde oluşan denklem sadece sabit katsayı alınmıştır.

Tablo 24. Başlangıç Adımında Denklem Girmeyen Değişkenler

Denkleme Alınmayan Değişkenler					
			Skor	df	Sig.
Adım 0	Değerler	YANGIN	24,797	1	,000001
		NAKLIYAT	23,281	1	,000001
		MÜHENDİS	3,073	1	,079586
		TARIM	6,827	1	,008981
		KASKO	25,957	1	,000000
		Z.KARA.Y	20,975	1	,000005
		Z.KOLTUK	13,438	1	,000247
		FERDI.K	12,996	1	,000312
		DIĞER.K	17,884	1	,000023
		SAĞLIK	8,031	1	,004597
		HAYAT	,168	1	,681498
Toplam istatistik			35,591	11	,000198

Tablo 23'den görüleceği gibi başlangıç adımında oluşan denklem hiçbir değişken alınmamıştır. Buradan da anlaşılacağı üzere başlangıç çözümü yetersizdir. Doğru çözüme ulaşmak için diğer adıma geçip analize devam edilmelidir. Birinci adımın çözümü aşağıdaki gibidir.

Tablo 25. Anlamlılık Testi

		Model Katsayıları		
		ki-kare	df	Sig.
Adım 1	Adım	47,993	11	,000001
	Blok	47,993	11	,000001
	Model	47,993	11	,000001

Yukarıdaki tablo incelendiğinde adım 1'deki anlamlılık değerine (sig.) bakarsak $0 < 0,05$ olduğundan kurulan lojistik model anlamlıdır.

Tablo 26. Model Özeti

Model Özeti			
Adım	-2 Log likelihood	Cox & Snell R kare	Nagelkerke R kare
1	8,541	,648	,916

Modele ait Nagelkerke $R^2 = 0,916$ olarak bulunmuştur. Yani kurulan lojistik modelin kullanılan değişkenlerce açıklanma oranı %91,6'dır. Ayrıca Cox ve Snell R^2 de 0,648 olarak bulunmuştur ki bu açıklama oranlarının yeterince yüksek olduğu ve kurulan modelin anlamlı olduğu söylenebilir.

Tablo 27. Sonuç Sınıflandırma Tablosu

Gerçek Değerler		Sınıflandırma Tablosu ^a		
		Tahmini Değerler		Doğru sınıflandırma
		ORT.SATŞ		
		0	1	
adım 1	ORT.SATŞ 0	32	0	100,0
	1	1	13	92,9
toplam istatistik				97,8

a. Modele değişkenlerin alım oranı 0,5 olarak belirlendi

1. adımdan sonra kurulan denkleme göre ortalamanın altında kalan 32 şirketin doğru atandığı fakat ortalamanın üstünde bulunan 1 şirketin yanlış atandığı aslında ortalamanın altında hesaplanması gerektiği görülmüştür. Diğer bir deyişle Türkiye'deki tüm sigorta şirketlerinin satış ortalaması hesaplandığında Aviva Hayat Emeklilik şirketinin satış ortalaması ülke ortalamasının üstünde kalmadığı söylenebilir. Sonuç olarak kurulan bu lojistik analizin %98 e yakın bir doğruluk içerdiği görülmektedir.

Tablo 28. Denkleme Giren ve Girmeyen Değişkenler

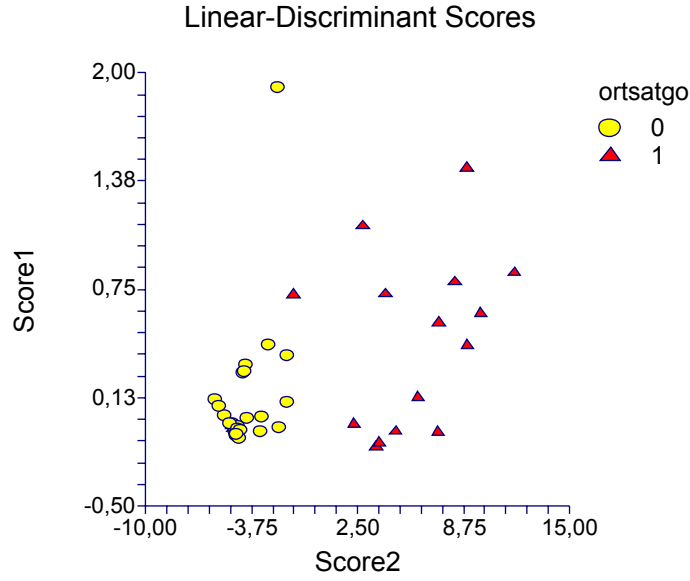
Denkleme Alınan Değişkenler

	B	S.E.	Wald	df	Sig.	Exp(B)
Adım # YANGIN	-,000050	,000476	,011214	1	,915663	,999950
NAKLIYAT	,000083	,000166	,248141	1	,618387	1,000083
MÜHENDIS	-,000219	,000348	,397368	1	,528452	,999781
TARIM	,000063	,000219	,081969	1	,774646	1,000063
KASKO	,000093	,000239	,150455	1	,698101	1,000093
Z.KARA.Y	-,002301	,011153	,042547	1	,836581	,997702
Z.KOLTUK	,002398	,007627	,098822	1	,753248	1,002400
FERDI.K	,000050	,000059	,725352	1	,394394	1,000050
DIĞER.K	-,000015	,000034	,200060	1	,654672	,999985
SAĞLIK	,000094	,000309	,092346	1	,761216	1,000094
HAYAT	,000005	,000004	2,426256	1	,119318	1,000005
Constant	-6,896261	3,396259	4,123120	1	,042301	,001012

a. 1. Adımdan sonra denkleme giren değişkenler: YANGIN, NAKLIYAT, MÜHENDIS, TARIM KASKO, Z.K.YOLU, Z.KOLTUK, FERDI.K, DIĞER.K, SAĞLIK, HAYAT.

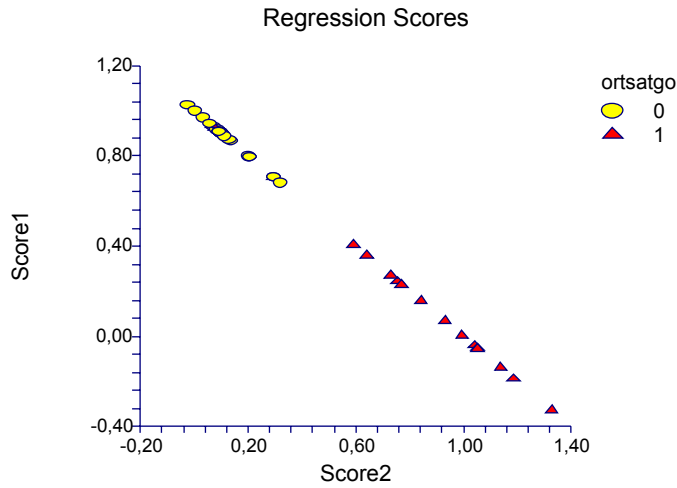
1.adımdan sonra kurulan denkleme aşağıdaki değişkenler alınmıştır. Analiz sonucunda çıkan denklem de aşağıdaki gibidir.

$$Y=6,896261-0,00005X_1+0,000083X_2-0,000219X_3+0,000063X_4+0,000093X_5-0,002301X_6+0,002398X_7+0,00005X_8-0,000015X_9+0,000094X_{10}+0,000005X_{11}$$



Şekil 11. Lineer Diskriminat Analizi ile Sınıflandırma

Diskriminant analizi ile oluşturulan grafikte şirketlerin kendi aralarında nasıl gruplandıklarını çok rahat görebiliyoruz. Kırmızı nokta ile gördüğümüz şirketler sektör satış ortalamasının üzerinde kalan şirketlerdir.



Şekil 12. Regresyon Analizi ile Sınıflandırma

Regresyon analiziyle oluşturulan trend ile şirketlerin nasıl bir eğilim gösterdiğini açıkça görebiliyoruz. Kırmızı nokta ile gördüğümüz şirketler, 2005 yılında Türkiye’de yapılan tüm sigorta poliçe satışı sonucunda ortalamanın üzerinde kalan şirketlerdir.

3.6 Şirketlerin 2005 Yılı Pazar Payları Baz Alınarak Uygulanan Lojistik Regresyon Analizi İle Diskriminat Analizinin Karşılaştırılması

Aşağıda önce Türkiye’de faaliyet gösteren 46 şirketin 2004 yılında pazarda buldukları yerle 2005 yılında pazarda buldukları yer kıyaslanarak şirketlerin bir yıl içerisinde gösterdikleri gelişmişlik diskriminant analiziyle sınıflanacak ve şirketlerin kendi aralarındaki performanslarını karşılaştırılacak sonra lojistik regresyon analiziyle arasındaki fark ortaya konmaya çalışılacaktır.

Tablo 29. Seçilen Gözlemler

Model Özeti		N	Yüzde
Ağırlıklandırma ^a			
Modele Seçilen	Analize Giren	46	100,0
	Kayıp Gözlem	0	,0
	Toplam	46	100,0
Modele Girmeyen		0	,0
Toplam		46	100,0

a. Modele giren gözlemler

Araştırmada kullanılan 46 şirkete ait verilerin hepsi analize katılmıştır, kayıp gözlem bulunmamaktadır.

Tablo 30. Anlamlılık Testi

	Wilks' Lambda	F	df1	df2	Sig.
YANGIN	,937	2,955	1	44	,093
NAKLIYAT	,943	2,671	1	44	,109
MÜHENDİS	,893	5,268	1	44	,027
TARIM	,998	,087	1	44	,769
KASKO	,937	2,969	1	44	,092
TRAFİK	,971	1,330	1	44	,255
Z.KARA.Y	,913	4,190	1	44	,047
Z.KOLTUK	,975	1,117	1	44	,296
FERDI.K	,984	,714	1	44	,403
DİĞER.K	,917	3,962	1	44	,053
SAĞLIK	,974	1,187	1	44	,282
HAYAT	,926	3,531	1	44	,067

Yukarıdaki tabloda değişkenlerin anlamlılık değerlerine bakıldığında aralarında 0,05 den büyük olanlar göze çarpmaktadır. Buda ortalamaların ortalamaları eşit hipotezinin reddine yol açar ki böylece analizin anlamlılığı belirir ve analize devam kararı alınır.

Tablo 31. Bağımlı Değişkenlerin Sınıflandırılması

		Grup İstatistikleri		
PAZAR.P		Ortalama	Standart Hata	Geçerli
				Gözlemler
0	YANGIN	25700,89	55216,888	27
	NAKLIYAT	17807,78	35243,950	27
	MÜHENDIS	5721,22	15842,992	27
	TARIM	4262,81	19140,528	27
	KASKO	43282,44	101390,069	27
	TRAFİK	150868,15	286068,933	27
	Z.KARA.Y	364,37	928,256	27
	Z.KOLTUK	634,26	2116,093	27
	FERDİ.K	44900,04	129217,585	27
	DIĞER.K	31105,63	96833,308	27
	SAĞLIK	4905,67	16336,888	27
	HAYAT	136997,59	296936,510	27
1	YANGIN	57652,16	70803,062	19
	NAKLIYAT	37919,21	48304,330	19
	MÜHENDIS	27445,26	45600,686	19
	TARIM	5979,42	19722,110	19
	KASKO	98245,05	113533,708	19
	TRAFİK	245325,32	254296,788	19
	Z.KARA.Y	1089,26	1474,574	19
	Z.KOLTUK	1276,21	1894,017	19
	FERDİ.K	75869,89	111819,506	19
	DIĞER.K	104130,11	152137,422	19
	SAĞLIK	10633,47	19187,151	19
	HAYAT	8355,00	19903,034	19
Toplam	YANGIN	38898,15	63402,054	46
	NAKLIYAT	26114,67	41847,820	46
	MÜHENDIS	14694,20	33071,877	46
	TARIM	4971,85	19183,052	46
	KASKO	65984,39	108830,788	46
	TRAFİK	189883,07	274518,817	46
	Z.KARA.Y	663,78	1223,854	46
	Z.KOLTUK	899,41	2030,825	46
	FERDİ.K	57691,93	122009,839	46
	DIĞER.K	61267,91	126481,271	46
	SAĞLIK	7271,50	17595,304	46
	HAYAT	83862,61	234953,438	46

Her grupta 12 tane değişken bulunmaktadır. Toplam 2 grup vardır. Şirketlerin pazar paylarındaki değişim temel alınarak ayırım 2 grupta oluşmuş, birinci grupta 27, ikinci grupta 19 gözlem sınıflanmıştır.

Tablo 32. Modelin Açıklanma Oranı

Özdeğerler

fonksiyon	özdeğer	% varyans	kümülatif %	korelasyon
1	,829 ^a	100,0	100,0	,673

a. canonical diskriminat analizi fonksiyonun oluşumunda kullanılmıştır.

Analizde sadece 1. fonksiyon kullanılmış ve 1. fonksiyonun varyansı açıklama yüzdesi % 100 olarak hesaplanmıştır.

Tablo 33. Anlamlılık Testi

Wilks' Lambda

test fonksiyonu	Wilks' Lambda	ki-kare	df	Sig.
1	,547	22,933	12	,028

1. fonksiyon anlamlılık değeri $0,028 < 0,05$ olduğundan diskriminant fonksiyonu anlamlıdır.

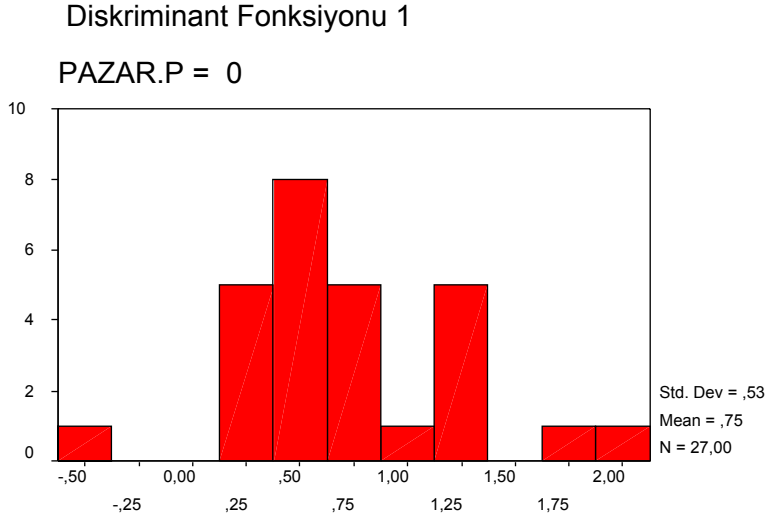
Tablo 34. Kurulan Modelin Standartlaştırılmış Fonksiyonu

Standardize edilmiş diskriminat fonksiyonun katsayıları

	Fonksiyon
	1
YANGIN	-,449
NAKLIYAT	-,630
MÜHENDİS	-,720
TARIM	,845
KASKO	,844
TRAFİK	,901
Z.KARA.Y	-2,878
Z.KOLTUK	2,083
FERDİ.K	,340
DIĞER.K	-,773
SAĞLIK	,225
HAYAT	,336

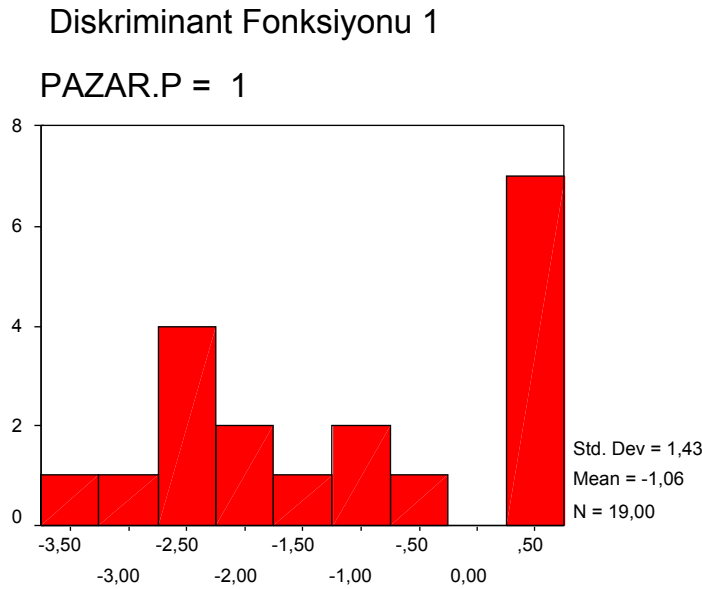
Kullanılan 1. fonksiyon aşağıdaki gibidir.

$$Y = -0,4499X_1 - 0,63X_2 - 0,72X_3 + 0,845X_4 + 0,844X_5 + 0,901X_6 - 2,878X_7 + 2,083X_8 + 0,34X_9 - 0,773X_{10} + 0,225X_{11} + 0,336X_{12}$$



Şekil 14. Pazar Payları Artmayan Şirketlerin Analizi

Elde ettiğimiz diskriminant fonksiyonunda pazar payına (y bağımlı değişkenine) 0 değeri verdiğimizde elde edilen modelin standart sapması 0,53 ortalaması ise 0,75 olmaktadır.



Şekil 15. Pazar Payları Artan Şirketlerin Analizi

Elde ettiğimiz diskriminant fonksiyonunda pazar payına (y bağımlı değişkenine) 1 değeri verdiğimizde elde edilen modelin ortalaması -1,06 standart hatası ise 1,43 olmaktadır.

Tablo 35. Sonuç Sınıflandırma Tablosu

		tahmini		Toplam
		0	1	
Diskriminant analizi gerçek	PAZAR.P			
	0	26	1	27
	1	7	12	19
	%	0	96,3	3,7
	1	36,8	63,2	100,0
Lojistik regresyon analizi gerçek ^a	PAZAR.P			
	0	26	1	27
	1	4	15	19
	%	0	96,3	3,7
	1	,1	,8	100,0

a. Lojistik regresyon analizi sonucu

b. 82,6% Diskriminant analizin doğru sınıflandırma oranı.

c. 89,1% lojistik regresyon analizinin doğru sınıflandırma oranı

Yapılan analizden sonra kurulan denkleme göre pazar payı azalmış ya da aynı kalmış olarak belirtilen Generali Sigortanın pazar payının yapmış olduğumuz diskriminant analizine göre arttığı söylenebilir. Ancak Pazar payı artmış olarak ayrılan Acıbadem Sağlık ve Hayat, AIG, Ak Emeklilik, Başak Emeklilik, Garanti Emeklilik, Genel Yaşam ve Ray sigorta şirketlerinin analize göre pazar paylarının artmadığını sadece 13 şirketin pazar payının artmış olduğu yorumu yapılabilir. Ancak bu yorumu şirketlerin 2004–2005 yılı pazar payına bakarak söylemekteyiz. Sonuç olarak, kurulan bu modelin % 83 civarı bir doğruluk içerdiği görülmektedir.

Yukarıdaki %83' lik doğru sınıflandırma oranının anlamlılığının sınanması amacıyla aşağıda Q test istatistiği hesaplanmıştır.

$$Q = (n-n'k)^2 / n(k-1) \quad (81)$$

$$n' = 26+12$$

$$Q = (46 - 38 \cdot 2)^2 / 46 (2 - 1)$$

$$Q = 19,565 > 3,841 \chi^2(1;0,05)$$

Elde edilen 19,565 Q test istatistiği, 0,05 anlamlılık düzeyinde 1 serbestlik dereceli χ^2 değerinden (3,841) büyük olduğu için fonksiyonun ayırma özelliğinin önemli olduğu ve sonuçların tatmin edici düzeyde çıktığını söyleyebiliriz.

SONUÇ

Bu çalışmada, ülkemizdeki sigorta şirketlerinin aynada kendi gerçek büyüklüklerini görüp, geleceğe yönelik stratejiler oluştururken şirket büyüklüklerine ilişkin ön yorum yapmaları hedeflenmiştir. Bu amaç doğrultusunda geleceğe yönelik tahminlerin oluşturulmasında en önemli matematiksel olgu olan ve çalışmada ana öge olarak ele alınan lojistik regresyon analizi anlatılarak bu analize temel teşkil eden doğrusal regresyon analizinden ve karşılaştırma yapılmasına olanak veren alternatif birçok değişkenli sınıflandırma modeli olan diskriminant analizine değinilmiştir. Ayrıca doğrusal regresyon analizinin ve diskriminant analizinin gerekliliği ve yetersiz kaldığı aşamalar anlatılarak lojistik regresyon analizinin uygun çözüm yolları ürettiği gösterilmiştir.

Verilerin sınıflandırılmasında kullanılan modellemelerin oluşturulması için seçilen verilerin de çok önemli olduğu, öyle ki seçilen veri kümesine göre uygulanacak analizlerin dahi değiştiği gösterilmiştir. Klasik doğrusal regresyon analizinde de, diğer çok değişkenli analizlerde de bağımlı değişken sürekli iken, lojistik regresyon analizinde bağımlı değişken kategoriktir. Bu çalışmada kullanılacak verilerde sürekli ve kategorik değişkenlerin bir karması olarak seçilmiştir. Dolayısıyla kullanacağımız analiz türü olarak da söz konusu veri kümesinden en iyi sonucu elde edebileceğimiz lojistik regresyon analizi seçilmiştir.

Ülkemizde sigortacılık sektörü incelendiğinde özellikle son zamanlarda globalleşmenin neticesinde ve Avrupa Birliği'ne giriş sürecinde, sektörde oluşan uygulamalarda ve sürekli gelişen bir rekabet ortamında şirketlerin çok zor durumlara düştüğü, özellikle sermaye yeterliliği konusunda bir hayli gerilerde kaldığı gözükmektedir. Ülkeye yerli ortaklıklarla giren ve yakın gelecekte direk giriş hamlelerinde bulunacak olan dünyanın bu dev şirketleri çok büyük sermayeleri ve gelişmiş analiz teknikleriyle birlikte sektörde çok yıkıcı bir rekabet ortamı oluşturacaklardır. Bu nedenlerle ülkemizdeki sigorta şirketlerinin bu rekabet ortamında

ayakta durmak için gerekli olan bilimsel sınıflandırma ve istatistikî analizlere daha fazla yönelmeleri gerekmektedir.

Türk sigorta şirketlerinin en büyük eksikliği olan sermaye zaafı, pazarı ve ülke insanını iyi tanıma avantajından yararlanıp oluşturulacak istatistiksel analiz teknikleri şirket politikalarında sahaya yansıtılması gererken en önemli noktalardan biri gibi gözükmektedir. Bu çalışma ile ana hedefte ki kararların alınmasında bir ön analiz oluşturacak performans sınıflamalarının yapılabileceği gösterilmiştir.

Bundan sonra yapılacak çalışmalarda ise lojistik regresyon analizine daha fazla önem verilerek şirketlerin bir ön analizi olmaktan çıkarılıp daha fazla zaman ve para ayrılarak güçlü alansal araştırmalarla daha kaliteli veriler elde edilip, şirketlerin ve idari mercilerin karar almasını sağlayan somut ve kullanılabilirliği yaygın analizler elde edilmelidir. Böylece, sigorta şirketleri gerek kendi bünyelerinde gerek sektörel çaplı kararlar alırken somut ve güvenilir kaynaklardan yararlanma şansı elde edebileceklerdir.

KAYNAKÇA

Akdeniz, Fikri.; **Olasılık ve İstatistik** , Adana: Ç.Ü. Basımevi, 1998.

Aldrich, John H. ve Forrest D.Nelson.; **Linear Probability, Logit and Probit Models**, 1. Edition, California: Sage, 1984.

Allison, D. Paul.; **Logistic Regression Using The SAS System**, 2. Edition, Cary: SAS Institute, 2000.

Başarır, Gülay.; **Uygulamalı Parametrik Olmayan İstatistiksel Yöntemler**, İstanbul: Pamaş, 2001.

Başarır, Gülay.; **Çok Değişkenli Verilerde Ayrımsama Sorunu ve Lojistik Regresyon Analizi**, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Yayınlanmamış Doktora Tezi, 1990.

Bayram, Nuran.; **Multinomial Lojistik Regresyon Analizinin İstihdamdaki İşgücüne Uygulanması**, 2004.

Bekiroğlu, N.; **Lojistik Regresyon ve Multinomial Lojistik Regresyonla Bir Uygulama**, Doktora Tezi, İstanbul Üniversitesi, 1996.

Çil, Burhan.; **İstatistik**, 4.Baskı, Ankara: Detay Yayınları, 2004.

Ediz, B.; **Lojistik Regresyon Ayırma Analizi**, U.Ü., Biyoistatistik Anabilim Dalı, Doktora Tezi, 1997.

Evans, James R.; **Statistics, Data Analysis, And Decision Modeling**, 1.Baskı, Upper Saddle River NJ: Prentice, 2003.

Genceli , Mehmet.; **Ekonometride İstatistik Bilgiler**, Filiz Kitabevi , İstanbul, 1989.

Gujarati Damodar N.; **Temel Ekonometri** (Çev. :Ümit Şenesen, Gülay Günlük Şenesen) 1. Basım, İstanbul: Çevik, 1999.

Gürsakal, Necmi.; **Bilgisayar Uygulamalı İstatistik 2** , 1. Baskı, İstanbul : Alfa, 2001.

Hair, J.F., Anderson, R.E., Tatham, R.L & Black, W.C .; **Multivariate Data Analysis**, Prentice Hall International, 1995.

Hosmer, D.W. ve Lemeshow, Jr.S.; **Applied Logistic Regression**, John Wiley&Sons, 1995.

İnal, H.Ceyhan ve Günay, Süleyman.; **Olasılık ve Matematiksel İstatistik**, Ankara: H.Ü.Fen Fakültesi Basımevi, 1999.

İşyar, Y.; **Ekonometrik Modeller**, Uludağ Üni. Güçlendirme Vakfı Yayını 92. 2000.

Kleinbaum, David G.; **Logistic Regression: A Self-Learning Text**, 1.Edition, NewYork: SpringerVerlag, 1994.

Menard, Scott.; **Applied Logistic Regression Analysis**, 2.Baskı, London: Sage Publications 2002.

Milli Reasürans T.A.Ş Yıllıkları, 2005.

Newbold, Paul., William, L.Carlson. & Betty, Thorne.; **Statistics For Business And Economics**, 5.Baskı, Upper Saddle River NJ: Prentice Hall, 2003.

Özdamar, Kazım.; **Paket Programlar ile İstatistik Veri Analizi 2.** 4.Baskı, Eskişehir, Kaan Yayınevi, 2002.

Serper, Özer; **Uygulamalı İstatistik 2.** 3.Baskı, İstanbul, Filiz,1996.

Sharma, Subhash.; **Applied Multivariate Techniques**, NewYork: John Wiley&Sons, mc., 1996.

Sigorta Sektörü Almanığı, **Milli Reasürans T.A.Ş.**, 2004.

Tatlıdil, Hüseyin.; **Uygulamalı Çok Değişkenli İstatistiksel Analiz**, Ankara: Cem Web Ofset, 1996.

Uralcan, G. Şebnem.; **Temel Sigorta Bilgileri ve Sigorta Sektörünün Yapısal Analizi**, İstanbul: Beta Yayınevi, 2004.

Walpole, E.Ronald.; **Introduction To Statics**, 3.Baskı, New York: Mac Millan, 1982.

Agresti, D.; **Logistic Regression**, <http://userwww.sfsu.edu/~efc/classes/biol710/logistic/logisticreg.htm>, 2006.

Hilbe, Joseph.; **Logistic Regression**, <http://www.statistics.com/content/courses/logistic/>, 2006.

Aldrich, John. & Nelson, Forrest.; **The Logistic Regression Model**, <http://www.nd.edu/~rwilliam/stats2/182.pdf>, 2006.

King, Gary. & Zeng, Langche.; **Logistic Regression in Rare Events Data**, <http://gking.harvard.edu/files/0s.pdf>, 2006.

Handcock, Mark S., Rendall, Michael S. & Cheadle, Jacob E.; **Improved regression estimation of a multivariate relationship with population data on the bivariate relationship**, <http://www.csss.washington.edu/Papers/wp36.pdf>. 2006.

Fisher, Robin. & Campbell, Jennifer.; **Health Insurance Estimates for States**, <http://www.Census.gov/hhes/www/saife/asapaper/fishercampbellasa02.pdf>, 2006.

Galiano, Marcos Marin. & Christmann, Andreas.; **An R-Program to Model Insurance Data**, <http://www.sfb475.uni-dortmund.de/berichte/tr49-04.pdf>, 2006.

Temple, Jeromey.; **Explaining the private health insurance coverage for older Australians**, <http://www.elecpres.monash.edu.au/pnp/cart/download/free.php?paper=178>, 2006.

Cho, Moon J. & Eltinge, John L.; **Modeling of Survey Response Rates and Reporting Rates in the U.S. Consumer Expenditure Interview Survey**, <http://www.bls.gov/ore/pdf/st040040.pdf>, 2006.

Weissert, William G., Elston, Jennifer M. & Koch, Gary G.; **Risk of Institutionalization**, <http://aspe.hhs.gov/daltcp/reports/instrkes.pdf>, 2006.

Williamson, John B. & McNamara Tay K.; **Work Beyond the Typical Retirement Age**, http://www.bc.edu/centers/crr/papers/wp_2001-09.pdf, 2006.

Sullivan, Kerri.; **Transportation & Work**, http://www.ncsall.net/fileadmin/resources/research/op_sullivan.pdf, 2006.

Álvarez, Montserrat.; **Wholly-Owned Subsidiaries Versus Joint Ventures: The Determinant Factors in the Catalan Multinational Manufacturing Case**. <http://www.pcb.ub.es/ieb/serie/doc2003-5.pdf>. 2006.

Lee ,Tsun-Siou., Yeh,Yin-Hua. & Liu, Rong-Tze.; **Center for Economic Institutions**, <http://cei.ier.hit-u.ac.jp / working/2003/2003WorkingPapers /wp2003-14.pdf>, 2006.

Capella, Michele E.; **Predictors of Competitive Employment for Blind and Visually Impaired Consumers of Vocational Rehabilitation Services** http://eric.ed.gov/ERICDocs/ data ericdocs2/ content_storage_01/ 0000000b/ 80/ 2c/ 7d/ 1a.pdf, 2006.

Braniv, Ran., Agarwal, Anurag. & Leach, Robert.; **Predicting Bankruptcy Resolution**, http://_bear.cba.ufl.edu/_agarwal/_vita/PredictingBRResolutionJBFA.pdf, 2006.