T.C
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİMDALI
YÖNETİM BİLİMİ PROGRAMI
YÜKSEK LİSANS TEZİ

İŞLETMELERİN TAHMİNLEME SÜRECİNDE BULANIK DOĞRUSAL REGRESYON ANALİZİ VE LOJİSTİK REGRESYON ANALİZİNİN UYGULANMASI

Ayşe Cansu GÖK

Danışman
Doç. Dr. Ali ÖZDEMİR

2010

YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum "İşletmelerin Tahminleme Sürecinde Bulanık Doğrusal Regresyon Analizi ve Lojistik Regresyon Analizinin Uygulanması" adlı çalışmanın, tarafımdan, bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

Tarih

..../..../.......

Ayşe Cansu GÖK

ÖZET

Yüksek Lisans Tezi

(İşletmelerin Tahminleme Sürecinde Bulanık Doğrusal Regresyon Analizi ve Lojistik Regresyon Analizinin Uygulanması)

(Ayşe Cansu Gök)

Dokuz Eylül Üniversitesi
Sosyal Bilimler Enstitüsü
İşletme Anabilim Dalı
Yönetim Bilimi Programı

Son yıllarda artan rekabetçi ortamda ve küresel ekonominin yarattığı etkiler sonucunda, işletmeler için yaşamlarını devam ettirmek ve fark yaratabilmek adına en önemli araçlardan birisi de geleceğe yönelik tahminlerde bulunmak ve stratejilerini buna göre belirlemek olmuştur. Bu bağlamda işletmeler riskleri en aza indirgemek için, birçok uygulama alanında yer bulan tahminleme yöntemlerinin ve istatistiksel analizlerin kullanılmasına yönelmektedirler.

Sebep sonuç ilişkisine dayanan, bağımlı ve bağımsız değişkenler arasındaki ilişki biçimi regresyon olarak ifade edilmektedir ve istatistiksel analizlerde sıklıkla kullanılmaktadır. Bulanık Doğrusal Regresyon Analizi ise karar verme sürecinde, sistem yapısından kaynaklanan belirsizlikleri de dikkate alarak hem nitel hem de nicel değişkenlerin modele alınmasını sağlayan bir problem çözme tekniği kullanarak klasik regresyon analizine alternatif bir yöntem olmaktadır. Diğer bir regresyon tekniği olan Lojistik Regresyon Analizi ise sonuç değişkeninin iki veya çok düzeyli kategorik değişken olması, 0 ve 1 gibi kesikli değerler alması durumunda kullanılmaktadır. Bağımlı değişken üzerinde açıklayıcı değişkenlerin etkileri olasılık olarak elde edilerek, bu faktörlerin olasılık olarak belirlenmesi sağlanmaktadır.

Bu tez çalışmasında Bulanık Regresyon Analizi ve Lojistik Regresyon Analizi bir tahmin yöntemi olarak teorik olarak incelenmiş ve bu kapsamda bankaların sektör paylarının tahminlenmesine yönelik, her iki analiz yöntemi ile modeller oluşturularak bir uygulama yapılmıştır ve elde edilen sonuçlar yorumlanmıştır.

Anahtar Kelimeler: Bulanık Regresyon Analizi, Lojistik Regresyon Analizi, Tahminleme, Bulanık Mantık

ABSTRACT

Master Thesis

(Application of Fuzzy Linear Regression Analysis and Logistic Regression Analysis in Forecasting Process of Enterprises)

(Ayşe Cansu Gök)

Dokuz Eylül University
Institute of Social Sciences
Department of Business Administration
Management Science Program

In recent years, in increasingly competitive environment and as results of global economy, one of the most important tool for enterprises has become forecasting the future and determining their strategies in this way in order that enterprises maintain their life and create a difference. In this sense, for minimizing the risks, enterprises tend towards using of forecasting methods and statistical analysis that take part in many applications.

The relationship between dependent and independent variables that based on cause and effect relation, is expressed as a regression and oftenly used in statistical analysis. As for Fuzzy Linear Regression Analysis is being an alternative method to the classical regression analysis by using a problem solving tecnique which takes into account the fuzziness of system structure and include both qualitative and quantitative variables to the model in decision process. Another regression tecnique Logistic Regresson Analysis is used for the situation that the outcome variable is binomial or multinomial categorical variable, and take discrete values like 0 and 1. By the effects of explanatory variables on dependent variable are obtained as a probability, it provides to determine these factors as a probability.

In this thesis study, Fuzzy Regression Analysis and Logistic Regression Analysis is examined theoretically as a forecasting method and within this scope an application carried out for forecasting the sector portions of the banks with both analyses methods by setting models and consequently the obtained results are interpreted.

Keywords: Fuzzy Regression Analysis, Logistic Regression Analysis, Forecasting, Fuzzy Logic

İÇİNDEKİLER

İKİNCİ BÖLÜM

ÜÇÜNCÜ BÖLÜM

KISALTMALAR

SST: Total Sum of Squares = Tüm Hata Kareleri Toplamı

SSR: Regression Sum of Squares = Regresyon Kareleri Toplamı

SSE: Error Sum of Squares = Hata Kareleri Toplamı

p: Olasılık

O: Odds

SE: Standart Hata

TABLOLAR LİSTESİ

ŞEKİLLER LİSTESİ

GİRİŞ

Günümüzün küreselleşen dünyasında, işletmelerin varlıklarını sürdürebilmeleri ve rekabet ortamına ayak uydurabilmeleri açısından, firmalar için geleceğe yönelik tahminlerde bulunmak ve stratejilerini bu tahminlere göre belirmek önemli bir amaç haline gelmiştir. İşletmelerin vereceği kararların ve yapacağı planların temelini tahminleme süreci oluşturmaktadır. Dolayısıyla bu süreçte işletmelerin birçok problemde ve alanda uygulayabileceği tahminleme teknikleri ve analiz yöntemleri sıklıkla kullanılan bir araç olmaktadır. Bu yüzden, şirketler son yıllarda stratejik düşünce ve yönetime ağırlık vererek, geleceğe yönelik tahminlerde bulunabilmek için çeşitli analiz yöntemlerini yoğun olarak kullanmaya ve bir standart haline getirmeye başlamışlardır.

Sebep sonuç ilişkilerine dayanan bağımlı ve bağımsız değişkenler arasındaki ilişki biçimi regresyon olarak ifade edilmekte ve istatistiksel analizlerin çoğunda yer almaktadır. Regresyon modelleri nicel değişkenlerden faydalanarak karar verme ve tahmin problemlerinde yaygın olarak kullanılmaktadır. Regresyon analizi de, bu modellerde yer alan değişkenler arasındaki ilişkileri incelemekte ve buna bağlı tahminler yapılmasını sağlamaktadır. Bu yöntem; ekonomi, mühendislik, biyoloji ve sosyal bilimler gibi birçok alanda geniş uygulama alanı bulmaktadır.

İnsanlar tarafından yapılan tahminlerin etkili olduğu sistemleri modellemede, belirsiz bir sistem yapısıyla karşı karşıya kalınmaktadır. Bu tez çalışmasına konu olan Bulanık Doğrusal Regresyon Analizi de belirsiz bir ortamda, verilerin bulanık olması ve tamamının ya da bir kısmının kesin belirlenememesi durumlarında kullanılan ve klasik regresyon analizine alternatif olan bir yöntem olarak ortaya çıkmıştır. “Evet” veya “hayır” biçimindeki kesin verilere dayanmak yerine yaklaşık değer ve anlamlara, eksik veya kesin olmayan verilere olanak sağlayan bulanık mantık temellerine dayanan bulanık regresyon analizi, ilk olarak 1982 yılında Tanaka tarafından klasik regresyon prensiplerinin genişletilmesiyle ortaya koyulmuştur.

Bağımlı ve bağımsız değişkenler arasındaki ilişkiyi tanımlayan, tahminleme amacıyla geliştirilen alternatif yöntemlerden birisi de Lojistik Regresyon Analizidir. Son yıllarda lojistik regresyon analizi kullanım kolaylığının yanında rahat yorumlanabilmesiyle ön plana çıkmış ve sosyal bilimler alanında birçok uygulamada yaygın olarak kullanılmaya başlanmıştır. Lojistik regresyon analizi, sonuç değişkeninin kesikli olması, iki veya daha fazla değer alması durumunda kullanılan bir tahmin yöntemidir. Bağımlı değişken “başarılı-başarısız”, “az-orta-çok”, “olumlu-olumsuz” gibi kategorik verilerden oluştuğunda lojistik regresyon tercih edilmektedir.

Bu çalışmanın amacı da sosyal bilimler alanında yaygın olarak uygulanan regresyon tekniklerinden klasik regresyon yöntemine alternatif olan iki yöntem Bulanık Doğrusal Regresyon Analizi ve Lojistik Regresyon Analizlerini kullanarak modeller oluşturmak ve bu modellerle örnek bir uygulama yaparak bir tahminde bulunmaktır. Bu kapsamda, uygulamaya konu olan bankaların sektör paylarının tahminlenmesine yönelik her iki yöntemin uygulanması, elde edilen sonuçların kıyaslanması ve buna göre uygun yöntemin seçilmesi amaçlanmaktadır.

Çalışma üç bölümden oluşmaktadır ve birinci bölümünde Bulanık Doğrusal Regresyon Analizi yer almaktadır. Birinci bölümde, Bulanık Regresyon Analizinin çıkış noktası olan bulanık mantık ve bununla ilişkili kavramlar olan bulanık kümeler ve bulanık sayılara değinilmiştir. Daha sonra Bulanık doğrusal regresyon modeli verilerek temel model olan Tanaka modeli ve modelin bileşenleri anlatılmıştır. Bu doğrultuda bulanık regresyon yöntemlerinden uygulamalarda çoğunlukla karşılaşılan yöntemler Bulanık en küçük kareler Regresyonu ve Aralık Regresyonundan bahsedilmiştir.

Çalışmanın ikinci bölümünde Lojistik Regresyon Analizi yer almakta ve doğrusal regresyonla ilişkisine yer verilerek, Lojistik Regresyon Modeli anlatılmış, modelin anlaşılmasında faydalı olduğu düşünülen lojistik fonksiyon, odds oranı ve logit model kavramlarına yer verilmiştir. Çalışmanın uygulama kısmında, iki düzeyli bağımlı değişkenin yer aldığı model incelendiği için Lojistik Regresyon Analizi başlığı altında ikili model üzerinde durulmuştur. Daha sonra Lojistik Modellerde Parametrelerin Tahmini anlatılmış, parametre tahmin yöntemleri olan En çok Olabilirlik Yöntemi, Yeniden Ağırlıklandırılmış İteratif en küçük kareler Yöntemi ve Minumum Lojit Kikare Yönteminden bahsedilmiştir. Analizin devamında Modeldeki Katsayıların Anlamlılık Testi ve Yorumlanması anlatılarak, Olabilirlik Oranı Testi, Wald Testi ve Score Testine yer verilmiş, Modelin Uyum İyiliğinin Değerlendirilmesinde; Pearson Ki-kare(χ^2) ve Deviance İstatistiği, Model Ki-kare İstatistiği ve R_L^2 (Pseudo-R^2), Cox-Snell R^2 ve Nagelkerke R^2 İstatistiklerinden bahsedilmiştir.

Çalışmanın son bölümü olan üçüncü bölümde ise çalışmaya konu olan her iki analiz yönteminin kullanıldığı Sektör Paylarının Tahminlenmesinde Bulanık Doğrusal Regresyon Analizi ve Lojistik Regresyon Analizinin Uygulanması yer almaktadır. Uygulama kısmında çalışmada bu yöntemlerle ilgili daha önce yapılan çalışmaların bahsedildiği literatür taraması da bulunmaktadır. Bu doğrultuda araştırmanın amacı, modeli ve veri seti belirtilmiş, banka sektör paylarının tahmin edilmesine yönelik her iki tahmin yöntemi ile modeller oluşturulmuştur. Araştırmada Türk Bankacılık sektöründe faaliyet gösteren 45 bankaya ait veriler kullanılmış, veriler “Türkiye Bankalar Birliği” tarafından yayınlanan istatistiksel raporlardan elde edilmiştir, buna göre seçilen bağımlı ve bağımsız değişkenler arasındaki ilişkiler incelenmiş ve elde edilen sonuçlar değerlendirilerek karşılaştırılmıştır.

BİRİNCİ BÖLÜM

1. BULANIK DOĞRUSAL REGRESYON ANALİZİ

Bulanık doğrusal regresyon analizi, karar verme sürecinde hem nicel hem de nitel değişkenlerin dikkate alınmasına olanak sağlayan problemlerin çözümünde kullanılabilecek karar verme yöntemlerinden birisidir. Bulanık doğrusal regresyon, klasik regresyon modelini temel almakla birlikte, sistem yapısındaki bazı belirsizlikler nedeniyle klasik modele dahil edilemeyen açıklayıcı değişkenleri modele dahil edebilmesiyle klasik regresyon analizine göre olayları daha ayrıntılı ele alabilmektedir.

Bulanık regresyon analizi, ulaşılan verileri çok sınırlı olan ve kesin olmayan değişkenler arasındaki ilişkilerin tahmin edilmesinde, bu değişkenlerin belirsiz, nitel ve bulanık şekilde birbiriyle etkileşim içinde olduğu durumlarda kullanılmaktadır. Dolayısıyla, birçok işletme ve mühendislik probleminde uygulamaya elverişli bir yöntemdir. Bulanık doğrusal regresyon yöntemi ilk olarak Tanaka tarafından 1982 yılında klasik regresyondaki bazı katı varsayımların esnetilmesiyle ortaya koyulmuştur (Wang ve Tsaur, 2000a:355).

Bulanık doğrusal regresyon analizi, klasik regresyon analizinin bulanık bir şekli olup bulanık (belirsiz ve sınırları kesin tanımlanamayan) bir ortamda, bağımlı ve bağımsız değişkenler arasındaki fonksiyonel ilişkileri değerlendirmede kullanılır. Bulanık regresyon modelleri, tahminleme problemlerine çeşitli alanlarda uygulanabilmektedir. (Nasrabadi ve Nasrabadi, 2004:873)

Bulanık Doğrusal Regresyon analizi, Klasik Doğrusal Regresyon analizinin varsayımlarının sağlanmadığı durumlarda başarılı tahminler üretebilmektedir. Buna bağlı olarak da; normallik varsayımı, rastgelelik incelemeleri, durağanlık testleri, büyük örnekler, v.b. gerektirmez. Bulanık Doğrusal Regresyona konu olacak değişkenlerde aranan tek kriter doğrusal olmaları şartıdır (Yücel, 2005:1).

Bulanık regresyon yöntemi bulanık mantığın klasik regresyon yöntemlerine uygulanması sonucunda ortaya koyulmuş bir tahminleme yöntemidir. Dolayısıyla, yöntemin anlaşılabilmesi için öncelikle bulanık mantık, bulanık küme, bulanık sayı gibi kavramlara kısaca değinilecektir.

1.1. Bulanık Mantık

Bulanık mantık sistemleri, işletmelerde yapay zeka uygulamalarının küçük fakat ciddi ve gelişen kısmını temsil ederler. Bulanık mantık insan muhakemesine benzeyen muhakeme yöntemidir, çünkü evet ya da hayır seçimindeki gibi kesin verilere dayanmak yerine, yaklaşık değer ve anlamlara, eksik ya da belirsiz veriye olanak tanımaktadır. Bulanık mantık, eksik veriyi işlemden geçirerek diğer yöntemlerle çözülmesi zor sorunlara hızlı ve tahmini, fakat kabul edilebilir sonuçlar önerir (Şahin, 2005:223).

Bulanık mantığın ortaya çıkışı yıllar öncesine dayanmaktadır; uygulamaları geniş alanlara yayılarak çeşitlilik kazanmış, etkileri temel bilimlerde özellikle matematik ve fizik biliminde daha elle tutulur ve anlamlı hale gelmiştir. Bu doğrultuda bulanık mantığın asıl yaptığı, kelimelerle hesap yapmaktır. Genel olarak hesap yapmak rakamlarla ve sembollerle işlem yapmaya dayanmaktadır. Fakat insanlar hesaplama, nedenlendirme, sonuca varma gibi eylemlerde bulunurken kelimeleri kullanmaktadırlar. (Zadeh ve Kacprzyk, 1999:4)

Bilindiği gibi nitel değişkenler kesin değil görecelidirler. "Uzun-kısa", "zengin-fakir", "az-çok" gibi kavramlar kişiden kişiye farklı algılanır ve buna göre değerlendirilirler. Bulanık mantık da bu kavramlara nicel bir yaklaşım getirerek hesap yapmayı sağlamaktadır.

Klasik matematiksel yöntemlerle karmaşık sistemleri modellemek ve kontrol etmek zordur, çünkü veriler tam olmalıdır. Bulanık mantık kişiyi bu zorunluluktan kurtarır ve daha niteliksel bir tanımlama olanağı sağlar. Bir kişi için 38,5 yaşında demektense sadece orta yaşlı demek birçok uygulama için yeterli bir veridir. Böylece azımsanamayacak ölçüde bir bilgi indirgenmesi söz konusu olacak ve matematiksel bir tanımlama yerine daha kolay anlaşılabilen niteliksel bir tanımlama yapılabilecektir. (http://www.yapay-zeka.org/modules/icontent/index.php?page=33, Erişim:10.03.2010).

Aristo mantığı olarak bilinen iki değerli klasik mantık, 1920'lerden itibaren filozof ve teorik matematikçilerin ürettikleri paradoksları açıklamakta yetersiz kalmıştır. Çünkü Aristo mantığı gerçek dünyayı bütünüyle tasvir etmekten uzaktır. Herhangi bir önermenin yalnızca doğru ya da yalnızca yanlış olması gerekliliği, ikili mantığın gelişerek çok değerli mantığa dönüşmesine sebep olmuştur. Çok değerli mantığın en ilkel hali olan üç değerli mantık, önermelerin {0,1} değerlerinin yanında, {0.5} değerini de almasını sağlamıştır ve böylece değer kümesi {0, 0.5, 1} olarak geliştirilmiştir. Değer kümesindeki {0} öğesi önermenin kesinlikle yanlış olduğunu, {0.5} öğesi belirsiz olduğunu ve {1} öğesi de kesinlikle doğru olduğunu ifade etmektedir (Yücel, 2005:4).

1930'ların başında Polonyalı mantık bilimcisi Lukasiewicz, üç değerli mantıktan yola çıkarak çok değerli mantığı bütünüyle ele almıştır ve sonsuz değerli mantığı geliştirmiştir. 1965'te A. Lotfi Zadeh, o zaman kadar yapılan tüm mantıksal yaklaşımları toplu bir şekilde ele alarak yorumlamış ve ulaştığı çıkarımlarla bulanık mantığı keşfeden kişi olmuştur. Bulanık mantıkta önermeler [0,1] aralığında sonsuz değer alabilirler. Klasik mantıkta doğru veya yanlış olma durumu dışında başka herhangi bir durumun gerçekleşmesi olanaksız olarak varsayılır ve genellikle böyle durumların paradoks oldukları kabul edilir. Hâlbuki böyle durumlar bulanık mantık açısından son derece doğaldır. Hatta {0,1} dışında sadece üçüncü bir durumla yetinilmez, bunun dışında [0,1] aralığında değer alabilecek sonsuz durum gerçekleşebilmektedir (Yücel, 2005:4).

Karar vericiler hangi şartlarda ve boyutlarda karar verirlerse versinler, bir belirsizlik ortamı içinde bu işlevlerini yerine getirmek zorundadırlar. Verilen kararların doğruluğu ise, söz konusu belirsizliğin riske dönüştürülebildiği ölçüde sağlanacaktır. Ancak karar vericiler karar sürecinde klasik bilimsel yaklaşım ve bu yaklaşımın içerdiği yöntemleri kullanıyorlarsa, sonuçta verilen kararlar, iyi-kötü, güzel-çirkin, doğru-yanlış, evet-hayır, siyah-beyaz ya da 0-1 gibi yönlü kararlar olacaktır. Oysa gerçek yaşam mutlak ayrım üzerine kurulu değildir. Diğer bir deyişle karar ortamlarında mutlak siyah ve mutlak beyazın yanında binlerce gri tonunun varlığı unutulmamalıdır. Bu noktada genel anlamda karar süreçlerinde belirsizliğin nasıl öngörüleceği ve nasıl karar süreçlerinin bir parçası haline getirilebileceği yolunda çalışmalar başlamış ve bu çalışmaların sonunda alternatif bilimsel yaklaşım düşüncesi ortaya atılmıştır. Bu süreçteki son nokta ise Loutfi Zadeh' in Bulanık Mantık Teorisi olmuştur. Klasik mantık ve bulanık mantık arasındaki temel farklılıklar aşağıda Tablo 1'de gösterildiği gibi açıklanabilir. (www.deu.edu.tr/userweb/k.yaralioglu/dosyalar/bul_man.doc, Erişim:10.03.2010)

Tablo 1: Klasik Mantık ve Bulanık Mantık Arasındaki Temel Farklılıklar

Klasik Mantık	**Bulanık Mantık**
A veya A değil	A ve A değil
Kesin	Kısmi
Hepsi veya hiçbiri	Belirli derecelerde
0 veya 1 {0,1}	0 ve 1 arasında süreklilik [0,1]
İkili birimler	Bulanık(dereceli) birimler

Kaynak: www.deu.edu.tr/userweb/k.yaralioglu/dosyalar/bul_man.doc, Erişim:10.03.2010

Bulanık mantık konusunun temel elemanı "*bulanık kümeler*" dir. Bulanık mantıkla oluşan önermelerde yer alan terimleri kapsayan kümeler, bulanık küme olarak adlandırılırlar. Bulanık kümelerde, kümede bulunan her elemanın bir ait olma derecesi bulunmaktadır. Çalışmanın bir sonraki alt bölümünde bulanık kümelere kısaca değinilmiştir.

1.2. Bulanık Kümeler

Konuşma dilinde ifade edilen ve üzerinde çalıştığımız çoğu sınıflandırmalarda kullandığımız, kesin sınırlarla tanımlanamayan ve kişiden kişiye farklı yorumlanan "çok güzel", "fazla uzun", "aşırı sıcak", "hafif pahalı", "biraz tatlı" gibi bulanık kavramlar klasik mantığın öngördüğü şekilde incelenemezler. Bu tür terimlerle ifade edilen "Ayşe çok güzel.", "Hava aşırı sıcak.", "Amcam epeyce yaşlı." gibi ifadeleri, kesin hüküm belirtmediğinden, klasik mantık önerme olarak kabul etmez ve bu kavramlarla da klasik manada küme tanımlanamaz. İşte, bu tür önermelere *bulanık önermeler* ve bunlarla uğraşan mantığa da *bulanık mantık* denir. Bulanık önermeleri oluşturan bulanık terimlerin her biri bir *bulanık küme* ile modellenir. O halde, bir bulanık önermenin oluşturduğu bir bulanık küme, çalışma yapılan alana ait her bir bireye matematiksel olarak kümedeki aitlik derecesini temsil eden [0,1] aralığındaki gerçel sayılardan bir değer atayarak tanımlanır. Bu değer, söz konusu elemanın bulanık küme tarafından ifade edilen kavrama uygunluk derecesini belirtir. Tam üye olma ve üye olmama durumu, bulanık kümede de sırasıyla 1 ve 0 değerleriyle karşılanır. Dolayısıyla, klasik küme kavramı bulanık küme kavramının bu iki değere kısıtlanmış özel bir halidir. Bu nedenle, bulanık kümelerin matematiksel olarak ifadesi, klasik kümelerin karakteristik fonksiyonunun {0,1} değer kümesinin, [0,1] gerçel sayılar aralığına genelleştirilmesiyle yapılır. Buradan, bulanık kümelerin klasik kümelere bir alternatif değil, onların genelleştirilmişi olduğu görülür (Çağman, 2006:2-3).

Bulanık küme teorisinin bulucusu olarak kabul edilen A. Lotfi Zadeh'in 1965 yılında yayınladığı makalesinde bulanık küme kavramı şöyle tanımlanmaktadır: Bulanık bir küme, farklı üyelik yani ait olma dereceleri olan elemanlara sahip bir küme türüdür. Böyle bir küme, elemanlarının her birine 0 ile 1 arasında üyelik değeri atayabilen bir üyelik fonksiyonu ile karakterize edilebilir (Zadeh, 1965:338).

Kümeye dahil olmayan elemanların üyelik değerleri 0, kümeye tam dahil olanların üyelik değerleri de 1 olarak atanmaktadır. Kümeye dahil olup olmadıkları belirsiz olan elemanlara ise belirsizlik durumuna göre 0 ile 1 arasında değerler atanır. Oysa kesin küme teorisinde belirsiz eleman diye bir şey söz konusu değildir. Bir eleman ya kümeye dahildir ya da tamamı ile kümenin dışındadır. Dolayısıyla kesin kümelerde bir elemanın alabileceği üyelik değeri ya 0 ya da 1'dir (Altaş, 1999:83)

Klasik bir kümenin karakteristik fonksiyonu şu şekilde tanımlanır: A kümesi X evrenselinde klasik bir kümeyi temsil etmek üzere A kümesinin karakteristik fonksiyonu X_A ile aşağıdaki gibi ifade edilir.

$$X_A : X \rightarrow \{0,1\} \tag{1.1}$$

$$X_A(x) = \begin{cases} 1 & x \in X \\ 0 & x \notin X \end{cases} \tag{1.2}$$

Yukarıda verilen ifadeye göre, eğer bir x elemanı A kümesinin elemanı ise A kümesinin karakteristik fonksiyonu olan $X_A(x) = 1$'dir. Eğer bu kümenin elemanı değilse, $X_A(x) = 0$'dır (Tanaka, 1997:9).

Karakteristik fonksiyonlar klasik kümeler için nadiren kullanılmaktadır. Bulanık kümeler, klasik kümelerin genişletilmiş bir hali olarak düşünüldüğünde, karakteristik fonksiyonlar bulanık kümeler için daha anlamlı bir tanımlama sağlarlar. Klasik kümeler karakteristik fonksiyonlarla tanımlanırken, bulanık kümeler üyelik fonksiyonları ile tanımlanırlar (Tanaka, 1997:9). Buna göre X evrenselinde A bulanık kümesinin üyelik fonksiyonu "μ_A" olarak ifade edilmek üzere:

$$\mu_A : X \rightarrow [0,1] \tag{1.3}$$

A bulanık kümesi için $\mu_A(x)$ değeri x elemanının üyelik değerini göstermektedir. Bu üyelik değeri x elemanının A bulanık kümesine ait olma derecesini ifade eder. $\mu_A(x)$ değeri ne kadar 1'e yakınsa x elemanı da A bulanık kümesine o kadar aittir. Eğer $\mu_A(x) = 1$ ise, x elemanı A kümesine tam olarak aittir. $\mu_A(x) = 0$ ise, x elemanı A kümesine ait değildir (Tanaka, 1997:10).

1.3. Bulanık Sayı

$\tilde{A}$ ile gösterilen bir bulanık sayı, R reel sayılar kümesinin normalize edilmiş bulanık bir alt kümesidir. Dubois ve Prade tarafından ortaya koyulan, *L-R* tipi bulanık sayılar olarak adlandırılan simetrik bulanık sayıların genel gösterimi aşağıdaki gibidir (Modarres, Nasrabadi ve Nasrabadi, 2005:978).

$\tilde{A}=(\alpha, c)_L$ ile gösterilen $\tilde{A}$ bulanık sayısı için;

$$\tilde{A}(t) = L((t - c) / \alpha), \quad \alpha \geq 0$$

α değeri $\tilde{A}$ sayısının merkezi değerini, c ise yayılım değerini göstermektedir. Buna göre, referans fonksiyonu $L(x)$ ve aynı koşulları sağlayan $R(x)$ aşağıda gösterilen özelliklere sahiptirler:

i. $L(x) = L(-x)$
ii. $L(0) = 1, L(1) = 0$
iii. L, $[0,\infty)$ aralığında azalandır.
iv. L, $[0,1]$ aralığında tersi alınabilir. (Modarres ve diğer, 2005:978)

Bulanık doğrusal regresyon analizine temel oluşturan bulanık mantık, bulanık küme ve bulanık sayı kavramlarından kısaca bahsedildikten sonra, bu çalışmanın devamında bulanık doğrusal regresyon modeli anlatılacak ve model ile ilgili yaklaşımlara yer verilecektir.

1.4. Bulanık Doğrusal Regresyon

Regresyon analizi bağımlı değişken olan cevap değişkeni ve bağımsız değişken olan bir ya da daha fazla açıklayıcı değişken arasındaki ilişkileri incelemek için etkili ve ayrıntılı bir yöntemdir. Regresyon modelleri ile ilişkilendirilmiş çıkarımsal problemler, model parametrelerinin tahminini ve açıklayıcı değişkenlerin bilgisi ışığında cevap değişkenin tahminini içermektedir. Bu yöntem; ekonomi, mühendislik, biyoloji ve fizik gibi birçok alanda geniş uygulama alanı bulmaktadır. Klasik istatistik teknikleriyle cevap değişkeni veya açıklayıcı değişkenler için yapılan gözlemlerin genellikle normal dağılım olmak üzere, kesin bir olasılık dağılımına uyması gerekmektedir. Fakat uygulamada, doğası gereği gözlemlerin bulanık olduğu durumlar söz konusudur. Örneğin, "geniş", "ağır" ya da "yaklaşık olarak" gibi sözel ifadeler içeren gözlemler bulanıktırlar. Bu durumda, klasik regresyon analizi için regresyon katsayılarının ve sonraki aşamaların belirsiz bir ortamda tahmin edilmesi oldukça zordur (Kao ve Chyu, 2001:401)

Regresyon analizlerinin hesaplamaları bilgisayar programları tarafından kolaylıkla yapılabilmektedir. Çoğu bilgisayar kesin verilerle çalıştığı için, sözel verileri tanımlamak için bu verilere sembolik rakamlar atanmaktadır. Örneğin, 4: Mükemmel 3: Çok iyi 2: İyi 1: Fena değil gibi derecelendirmeler yapılır. Gerçek dünyadaki birçok problemde, verilerin fazla basitleştirilmesi regresyon modelleri için önemli bilgilerin atlanmasına sebep olmaktadır. Bazı bilgiler, "fena değil", "iyi" ve "mükemmel" gibi ifadeler sadece sözel terimlerle tanımlanabilirler. Böyle veriler için, bulanık küme teorisi, sözel değişkenlerin bulanık üyelik fonksiyonları yardımıyla modellenmesini sağlamaktadır. Klasik regresyon olasılık teorisine dayanırken, bulanık regresyon hem olasılık hem de bulanık küme teorisine dayanmaktadır (Chang ve Ayyub, 2001:187).

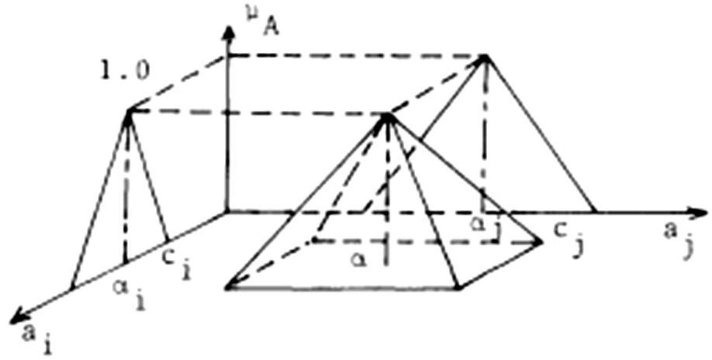
Bulanık regresyon modeli bağımlı ve bağımsız değişkenler arasındaki fonksiyonel ilişkilerin belirsiz bir ortamda değerlendirilmesinde kullanılmaktadır. Birçok bulanık regresyon sorununda, kararların kesin olmayan ya da kısmen ulaşılabilir verilerle alınmak zorunda kalındığı durumlarda doğrusal regresyon uygulanması önerilir. Bulanık regresyon modelleri tahminlemek için birçok yöntem bulunmaktadır. Bulanık doğrusal regresyon, bulanık doğrusal modeli veri ikilileri arasındaki ilişkileri tanımlamak için kullanır. Doğrusal olmayan durumda, bu varsayım modelleme hatalarına yol açmaktadır (Wang, 2006:208).

1.4.1. Bulanık Doğrusal Regresyon Modeli

İnsanlar tarafından yapılan tahminlerin etkili olduğu sistemleri modellemede, belirsiz bir sistem yapısıyla karşı karşıya kalınmaktadır. Bu yapı, parametreleri bulanık kümeler tarafından verilen bir bulanık doğrusal fonksiyondan oluşmaktadır. Bulanık doğrusal fonksiyonlar Zadeh'in genişletme prensibine dayanmaktadır. Bulanık regresyon analizi; bulanık doğrusal regresyon fonksiyonu, sistemin belirsiz yapısının bir modeli olarak düşünülerek formüle edilmektedir. Klasik regresyon modelinde, gözlenen ve tahmin edilen değerler arasındaki sapmalar ölçüm hatası olarak ele alınmaktadır. Bulanık regresyon modelinde ise, bu sapmaların sistem yapısının belirsizliğine yani sistem parametrelerinin kesin olmayışına dayandığı varsayılmaktadır. Bu nedenle bu sapmalar, doğrusal bir fonksiyonda bulanık parametrelerle tanımlanmaktadır. Modelin bulanık parametreleri, olasılık dağılımları ile ifade edilirler. (Tanaka, Uejima ve Asai, 1982:903)

Doğrusal modelden elde edilen bulanık parametreler, sistemin bulanıklığı (belirsizliği) ile örtüşen olasılık dağılımlarına karşılık gelirler. Bu bağlamda bulanık parametreler, üçgen üyelik fonksiyonları ile sınırlandırılmaktadırlar. Şekil 1'de bulanık parametrelere ait üçgensel üyelik fonksiyonu gösterilmektedir (Tanaka ve diğer, 1982:903).

Şekil 1: Bulanık Parametre Üçgensel Üyelik Fonksiyonu



Kaynak: Tanaka ve diğer, 1982:903

Bağımlı değişkenin karar vericinin öznel düşüncelerine göre belirlendiği durumlarda, bağımlı değişken için bir olasılık dağılımı bulmak zor olmaktadır. Tanaka tarafından sunulan bulanık regresyon modeli, bağımlı değişkenin üyelik değerlerini yansıtan bir olasılık dağılımına dayanmaktadır (Kim ve Bishu, 1998:344).

Girdisi kesin, çıktısı bulanık verilerden oluşan problemlerin bulanık regresyonla çözümlenmesini ilk ortaya koyan kişi H. Tanaka'dır. Tanaka'nın modelinde bulanık çıktı verisinin, üçgensel üyelik fonksiyonlarına sahip bulanık sayılardan meydana geldiği varsayılmaktadır (Wang ve Tsaur, 2000b:638).

Tanaka'nın geliştirdiği model üzerinden farklı ölçütler ele alınarak başka bulanık regresyon yöntemleri de geliştirilmiştir. Celmins, Diamond, Tanaka ve Isibuchi, Savic ve Pedrycz bu alanda katkıda bulunan bazı isimlerdir (Chang ve Ayyub, 2001:188). Bu çalışmada, öncelikle temel bulanık doğrusal regresyon modeli olan Tanaka modeli anlatılacak, daha sonra bulanık regresyon yöntemlerine ilişkin farklı yaklaşımlara da kısaca değinilecektir.

1.4.1.1. Tanaka Modeli

Tanaka modelinde bulanık bağımlı değişken ve kesin bağımsız değişkenin yer aldığı regresyon problemi bir matematiksel programlama sorunu olarak ele alınmıştır. Programlamadaki amaç; bulanık regresyon katsayılarının toplam yayılımını, regresyon modelinin bulanık cevapları tahmin etmek için önceden belirlenmiş olan üyelik değerlerine uygun olma kısıtını sağlayacak şekilde minimize etmektir (Kao ve Chyu, 2002:402).

X ve *Y* olarak iki küme ve *X*'ten *Y*'ye eşleşen bir $f(x,\ \mathrm{a})$ fonksiyonu düşünülsün. Eğer parametreler bulanık *A* kümesi tarafından verilirse, fonksiyon bulanık fonksiyon olarak adlandırılır ve $f(x, A)$ ile gösterilir. *x* değeri verildiğinde, $Y=f(x,\ A)$ bulanık fonksiyonu aşağıdaki gibi tanımlanır:

$$f : X \rightarrow F(y);\ Y = f(x, A) \tag{1.4}$$

F (y), *Y* kümesindeki bütün bulanık alt kümelerden oluşmaktadır. A bulanık kümesinin üyelik fonksiyonu $\mu_A(a)$ ile gösterilirken, *Y* bulanık kümesinin üyelik fonksiyonu $\mu_Y(y)$ aşağıdaki gibi ifade edilir:

$$\mu_Y(y) = \begin{cases} \max \mu_A(a), & \{a | y = f(x, a)\} \neq \emptyset \\ 0, & \text{aksi halde} \end{cases} \tag{1.5}$$

Bulanık doğrusal regresyonda, bulanık parametreler daha önce Şekil 1'de de gösterildiği gibi üçgensel üyelik fonksiyonları ile ifade edilen bulanık kümelerle sınırlandırılmışlardır. Bu bulanık kümeler aşağıdaki üyelik fonksiyonu ile tanımlanmaktadır: (Tanaka ve diğer, 1982:903)

$$\mu_A(a) = \min [\mu_{Aj}(a_j)] \tag{1.6}$$

$$\mu_{Aj}(a_j) = \begin{cases} 1 - \frac{|\alpha_j - a_j|}{c_j}, & \alpha_j - c_j \leq a_j \leq \alpha_j + c_j \\ 0, & \text{aksi halde} \end{cases} \tag{1.7}$$

Üçgensel üyelik fonksiyonları ile tanımlanan $\tilde{A}_j$ ile ifade edilen bulanık parametreler için $\tilde{A}_j = (\alpha_j, c_j)$, α_j merkezi değeri ve c_j yayılım değerini göstermektedir. $X = [X_0, X_1, \dots, X_N]^T$ bağımsız değişkenlerin vektörü, $\tilde{A} = [\tilde{A}_0, \tilde{A}_1, \dots, \tilde{A}_N]^T$ bulanık katsayıların vektörü olmak üzere temel modele dayanan bulanık doğrusal fonksiyon aşağıdaki gibidir: (Wang ve Tsaur, 2000a:355)

$$\tilde{Y} = \tilde{A}_0 + \tilde{A}_1 X_1 + \cdots + \tilde{A}_N X_N = \tilde{A} X \tag{1.8}$$

$\tilde{A}_j = (\alpha_j, c_j)$ üçgensel bulanık parametreleriyle tanımlanan bulanık regresyon fonksiyonu aşağıdaki şekilde bu parametreler gösterilerek de yazılabilir:

$$\tilde{Y}_i = (\alpha_0, c_0) + (\alpha_1, c_1)X_1 + (\alpha_2, c_2)X_2 + \cdots + (\alpha_N, c_N)X_N \tag{1.9}$$

Yukarıdaki bulanık regresyon yöntemi, girdi ve çıktı arasındaki ilişkinin, parametreleri olasılık dağılımına uyan bulanık fonksiyonlar tarafından tanımlandığı kesin girdi ve çıktı verileri için analiz yapar. Genişletme prensipleri uygulanarak, her bağımlı değişkenin değeri bulanık sayı olarak tahminlenir ve $\tilde{Y}_i$ bulanık sayısının üyelik fonksiyonu aşağıdaki gibi tanımlanır. Bulanık doğrusal fonksiyon, aşağıdaki üyelik fonksiyonu olarak elde edilir.

$$\mu\ (Y_i) = \begin{cases} 1 - \frac{|Y_i - X^t\alpha|}{c^t|X|}, & X \neq 0 \\ 1, & X=0,\ Y \neq 0 \\ 0, & X=0,\ Y=0 \end{cases} \tag{1.10}$$

$c^t = (c_0, c_1, \dots, c_N)$ ve $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_N)$'dir.

Buna göre bulanık sayı olarak tahminlenen $\tilde{Y}_i$ bağımlı değişkeninin alt ve üst sınırları aşağıdaki gibi hesaplanır: (Wang ve Tsaur, 2000a:356)

Alt sınır: $$\tilde{Y}_i = \sum_{j=0}^{N} (\alpha_j - c_j) X_{ij} \tag{1.11}$$

Merkezi değer: $$\tilde{Y}_i = \sum_{j=0}^{N} \alpha_j X_{ij} \tag{1.12}$$

Üst sınır: $$\tilde{Y}_i = \sum_{j=0}^{N} (\alpha_j + c_j) X_{ij} \tag{1.13}$$

Bulanık regresyon analizinde bahsedilen bulanık çıktı $\tilde{Y}_i$, $\tilde{Y}_i\ (y_i,\ e_i\ )$ şeklinde tanımlanmaktadır. y_i merkezi değeri, e_i ise yayılım değerini göstermektedir. Buna göre, bulanık çıktı $\tilde{Y}_i$ için aşağıdaki üyelik fonksiyonu tanımlanmaktadır: (Tanaka ve diğer, 1982:903)

$$\mu_{Yi}\,(Y) = 1 - \frac{|y_i - y|}{e_i} \qquad (1.14)$$

Burada amaç e_i hata değerinin yani buna karşılık gelen toplam yayılım değerini ifade eden $\sum c_j \left| x_{ij} \right|$ değerinin minimize edilmesidir. Bulanık regresyon analizinin bulanıklığı minimize edecek şekilde uygulanabilmesi için, $\tilde{Y}_i$ bulanık sayısının toplam yayılımı aşağıda (1.15) eşitliğinde gösterildiği gibi minimize edilir.

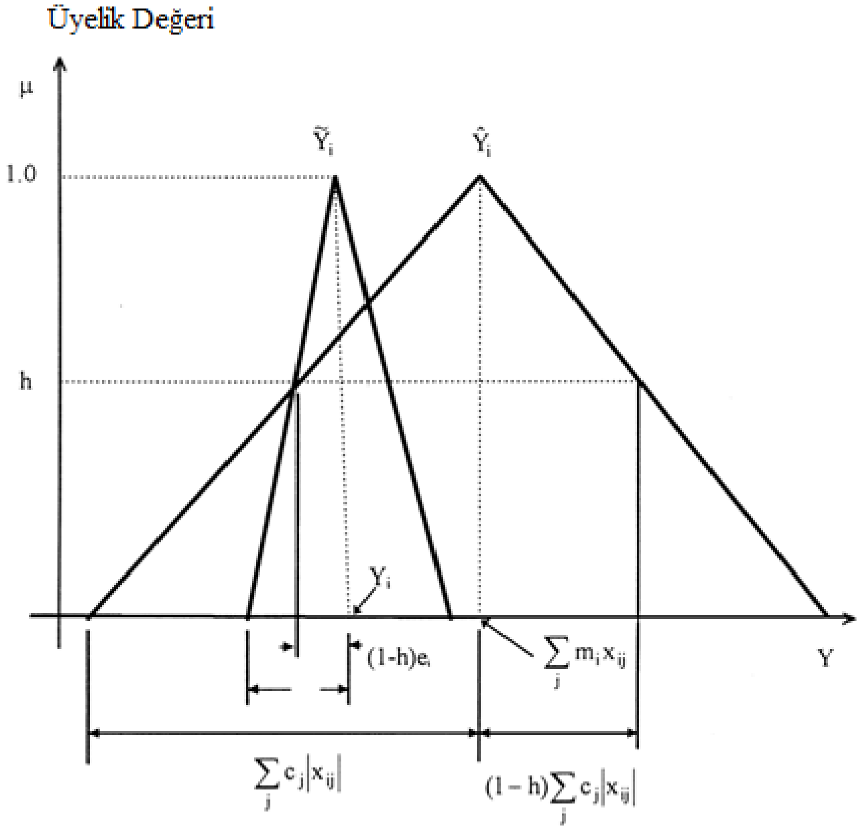
$$MINc^t \left| X \right| = MIN \sum_{j=0}^{N} \left(c_j \sum_{i=1}^{M} \left| x_{ij} \right| \right) \qquad (1.15)$$

Buna göre kısıtlar her Y_i gözleminin $\tilde{Y}_i$'ye en az h derecesi ile ait olmasını gerektirmektedir. $\mu\ (Y_i) \geq h$ olmalıdır ve $i= 1, 2,\ldots,M$'dir. (Wang ve Tsaur, 2000a:356).

$$1 - \frac{\left| Y_i - X^t \alpha \right|}{c^t \left| X \right|} \geq h \qquad (1.16)$$

"h terimi", bir uyum iyiliği ölçütü ya da regresyon modeli ile veri arasındaki uyumluluğun ölçütü olarak tanımlanmaktadır. Aşağıda Şekil 2'de gösterildiği gibi, bulanık $\tilde{Y}_i$ ya da kesin veri Y_i olabilen gözlenen her veri kümesi, h düzeyinde tahmin edilen $\hat{Y}$ değeri içinde olmak zorundadır. (Chang ve Ayyub, 2001:188).

Şekil 2: Bulanık Veri $\tilde{Y}_i$'ye göre $\hat{Y}_i$'nin Uygunluk Derecesi



Kaynak: Chang ve Ayyub, 2001:189

Bulanık doğrusal regresyonda hata; tüm model katsayılarına dağıtılır. Bu durumda her bir parametre belli bir bulanıklık seviyesinde tahmin edilir. Söz konusu bulanıklık seviyesi "*h*" olarak adlandırılır ve [0, 1] aralığında değer almaktadır. Bu değer, çalışmanın başında veri kümesinin eksik, yarım veya tam olma durumuna bakılarak analist tarafından belirlenir ve hesaplamalara sabit bir girdi olarak dahil edilir. *h* seviyesinin en ideal değerinin (gerçeğe en yakın tahminler üretebilen değerinin) ne olması gerektiği hala bir tartışma konusudur. Konuyla ilgili çalışmış olan bazı bilim adamları, *h*'ın optimum değerlerini belirleyerek bu değerler üzerinden bulanık tahmin yapılırsa başarılı olunacağını savunmuşlardır. Örneğin; Tanaka, Uejima ve Asai $h = 0.5$ olarak, Gharpuray, Fan ve Lai ise $h = 0.9$ olarak belirlenmesini savunmuşlardır (Moskowitz ve Kim, 1993:304-305).

Belirlenen bulanıklık seviyesi ile, aslında klasik doğrusal regresyon analizi ile tahmin edilen regresyon doğrusunu alttan ve üstten simetrik bir şekilde kuşatan bulanık regresyon aralığının ne kadar geniş olacağına karar verilmiş olunur. Bulanıklık seviyesinin baştan belirlenmesi, sistem parametrelerinin de bu bulanıklık düzeyinde tahmin edilmesine neden olur. Dolayısıyla Bulanık Doğrusal Regresyon modelinde, klasik modeldeki gibi ayrıca bir hata terimi (ε) yoktur. Modelin hatası, parametrelerin toplam yayılımlarına eşittir. Her bir parametre belli bir ölçüde hata ile (bulanıklıkla) tahmin edilmektedir. Yani hata, sistem parametrelerine dağıtılmış durumdadır (Yücel, 2005:1).

Bulanık doğrusal regresyon analizi, bu doğrultuda bulanık katsayıları belirlemek için aşağıdaki doğrusal programlama modeli ile formüle edilir:

$$MIN\sum_{j=0}^{N}\left(c_j\sum_{i=1}^{M}\left|x_{ij}\right|\right) \tag{1.17}$$

$$\sum_{j=0}^{N}\alpha_j x_{ij}+(1-h)\sum_{j=0}^{N}c_j\left|x_{ij}\right|\geq Y_i+(1-h)e_i \tag{1.18}$$

$$\sum_{j=0}^{N}\alpha_j x_{ij}-(1-h)\sum_{j=0}^{N}c_j\left|x_{ij}\right|\leq Y_i-(1-h)e_i \tag{1.19}$$

$(i = 1,2,\ldots N, \quad c_j \geq 0, \;\; 0 \leq h \leq 1)$

Yukarıdaki doğrusal programlama modelinde verilen (1.17) eşitliği regresyon modelinin toplam bulanıklığının minimize edilmesini göstermektedir. Eşitlik (1.18) ve Eşitlik (1.19) gözlemlenen bulanık veri değeri $\tilde{Y}_i=(y_i,\ e_i)$ ile ilgilidir. y_i bulanık merkez ve e_i bulanık yayılım ölçüsüdür. Eğer gözlemlenen bir veri kesin ise, verinin *e* değeri sıfırdır. Dolayısıyla, belirli bir kesin sayı bulanık bir sayının özel bir durumu olur (Chang ve Ayyub, 2001:189).

Bulanık regresyon analizi klasik regresyon analizinin yetersiz kaldığı gerçek hayat problemlerine uygulanabildiği için, bulanık regresyon analizini çalışan birçok araştırmacı olmuştur. Fakat Tanaka'nın modeline ilişkin aşağıda belirtilenler gibi bir takım eleştiriler bulunmaktadır:

- Girdi ve çıktı verisi her zaman kesin ve doğrusal olmayabilir.
- Tanaka'nın orijinal modeli aykırı değerlere son derece duyarlıdır.
- Bulanık regresyon aralıklarına ilişkin uygun bir yorumlama yoktur.
- Bulanık doğrusal regresyon, bağımsız değişkenler fazlalaştıkça çoklu bağlantı sorunuyla karşı karşıya kalmaya yatkındır (Wang ve Tsaur, 2000a:357).

Bu yaklaşım daha sonraları Tanaka'nın kendi çalışmaları ve Tanaka ile Watada'nın birlikte çalışmaları tarafından geliştirildiyse de Redden ve Woodall'ın çalışmalarında belirttiği gibi modelin aykırı değerlere aşırı duyarlı olması sorunu devam etmektedir. Bununla birlikte, modele daha fazla veri dahil edildikçe sonsuz çözüm ortaya çıkmakta ve tahmini değerlerin yayılımı daha da genişlemektedir (Kao ve Chyu, 2002:402).

1998 yılında Diamond, bulanık parametreleri belirlemek için, klasik normal denklemleri belli bir ölçüye göre türetme yoluyla "*Bulanık En Küçük Kareler Yöntemi*" ni ortaya koymuştur. Tanaka'nın modelinin avantajı, programlama ve hesaplamadaki kolaylığıdır. Hesaplama yönünden Tanaka'nın modeli daha etkili iken, tahminleme yönünden, bulanık en küçük kareler yöntemi Tanaka'nın modelinden daha iyi sonuçlar vermektedir (Nasrabadi ve Nasrabadi, 2004:874).

Çalışmanın bu kısmında bahsedilen Tanaka modelinden sonra, bu model üzerine daha sonraları geliştirilen bulanık regresyon yöntemlerine yer verilecektir. Öncelikle birçok çalışmada karşılaşılan bulanık en küçük kareler regresyonu kısaca anlatılacaktır. Daha sonra bulanık regresyonda bir başka yöntem olan aralık regresyonu anlatılacaktır.

1.4.2. Bulanık En Küçük Kareler Regresyonu

Bulanık regresyon modelleri iki kısımda toplanabilmektedir. İlk kısım Tanaka'nın yöntemi ve onun genişletilmişini kapsamaktadır. Bu kısımda, bağımlı değişkenler için tahmin edilen değerlerin toplam belirsizliği minimize edilmektedir. İkinci kısımda ise bulanık en küçük kareler yöntemi ile tahmin edilen değerlerdeki toplam hataların karesi minimize edilmektedir (Modarres, Nasrabadi ve Nasrabadi, 2005:978). Bu konuda çalışma yapan araştırmacılar, çıktı verisinin toplam yayılımını minimize etmek için Tanaka'nın modelini kullanmışlardır ya da çıktı verisinin toplam hatasını minimize etmek için bulanık en küçük kareler yöntemini modele uyarlamışlardır. Tanaka'nın yöntemi uygulamalarda fazla yardımcı olamayan geniş bir tahmin aralığı sunarken, bulanık en küçük kareler yöntemi daha dar bir aralıkta tahmin yapmayı sağlamasına rağmen hesaplama yaparken fazla zaman almaktadır (Wang ve Tsaur, 2000b:637).

Tek değişkenli bir bulanık doğrusal regresyon modelinden, gözlenen ve beklenen değerler arasında minimum bulanıklık ölçütünün uyarlanmasıyla, minimum regresyon aralıkları elde edilebilir. Bulanık en küçük kareler yöntemi, klasik en küçük kareler yöntemine benzerdir. Buradan elde edilen regresyon aralığı, Tanaka'nın modeline göre daha dardır (Modarres ve diğer, 2005:980).

Bulanık en küçük kareler regresyonuna Celmins, Diamond, Savic ve Pedrycz ve Chang ve Ayyub tarafından farklı bakış açıları geliştirilmiştir. Celmins, bulanık veri ve bir model arasında uygun ölçüm tanımlamış ve bu ölçümü model uygunluk kriteri olarak kullanmıştır. Diamond, bir en küçük kareler metodu geliştirmiştir. Savic ve Pedrycz, klasik regresyona minimum bulanıklık kriterini katarak bulanık en küçük kareler regresyonu için bütünleşmiş bir yaklaşım geliştirmiştir. Savic ve Pedrycz, en küçük kareler prensibini ve minimum bulanıklık kriterini birleştirerek bulanık regresyon yöntemi formüle etmiştir. Yöntem iki ardışık adımla yapılır. İlk adım bulanık regresyon katsayılarının bulanık merkez değerlerini bulmak için klasik regresyonu kullanır. İkinci adım bulanık regresyon katsayılarının bulanık aralıklarını bulmak için minimum bulanıklık kriterini kullanır (Chang ve Ayyub, 2001:189).

İlk adımda, bulanık gözlemlerin merkezi değerleri ile ilgili mevcut bilgiler kullanılarak bir regresyon doğrusu oluşturulur. Bulanık veriler basitleştirilmiş kesin veriler gibi davranır ve regresyon analizi klasik regresyon olarak uygulanır. İlk adımın sonuçları, bulanık regresyon katsayılarının merkezi değeri olarak kullanılır. İkinci adımda, bulanık katsayılar minimum bulanıklık kriteri kullanılarak hesaplanır. Bulanık katsayıların aralıkları daha önce verilen eşitlikler (1.18) ve (1.19) ile birinci adımda bulunan bulanık merkezler kullanılarak hesaplanır (Chang ve Ayyub, 2001:190).

Wang ve Tsaur, bulanık katsayıları hesaplamak için bir bulanık en küçük kareler yöntemi geliştirmiştir. Yöntem, tahmin edilen $\tilde{Y}_i$ bulanık bağımlı değişken değeri ile gözlenen Y_i değerleri arasında minimum bulanıklığı sağlayarak tek değişkenli bir regresyon modelinden minimum regresyon aralıkları elde etmektedir.

$\tilde{Y}_i = \tilde{A}_0 + \tilde{A}_1 X_{i1}$ olarak tanımlı bulanık doğrusal regresyon modelinde, $\tilde{A}_0=(\alpha_0,c_0)$ ve $\tilde{A}_1=(\alpha_1,c_1)$ olmak üzere, α_0 ve α_1 merkez değerleri, c_0 ve c_1 de yayılım değerlerini göstermektedir. Buna göre $\tilde{Y}_i$ ile Y_i arasındaki minimize edilmesi gereken bulanık fark aşağıdaki eşitlik ile tanımlanır: (Wang ve Tsaur, 2000b:640)

$$r(\tilde{A}_0 + \tilde{A}_1 X_{i1}, Y_i) = \Sigma\, d\,(\tilde{A}_0 + \tilde{A}_1 X_{i1}, Y_i)^2 \qquad (1.20)$$

Eşitlik genişletildiğinde aşağıda (1.21) eşitliğindeki gibi olur:

$$r(\tilde{A}_0 + \tilde{A}_1 X_{i1}, Y_i) = (\alpha_0 + c_0 + \alpha_1 X_{i1} + c_1 X_{i1} - y_i - e_i)^2 + (\alpha_0 + \alpha_1 X_{i1} - y_i - e_i)^2$$
$$+ (\alpha_0 - c_0 + \alpha_1 X_{i1} - c_1 X_{i1} - y_i - e_i)^2 \qquad (1.21)$$

Bulanık en küçük kareler regresyonunda, (1.21) eşitliğinin, α_0, α_1 ve c_0, c_1 değerlerine göre türevleri alınıp sıfıra eşitlenmesiyle denklemler çoğaltılarak çözüm yapılmaktadır. Böylece, modelde yer alan merkez ve yayılım değerleri elde edilmektedir. Çözüm için bu parametrelerin pozitif olup olmamalarına göre, (1.21) denkleminin türevi alındığında dört olası durum ortaya çıkmaktadır.

<u>1. Durum:</u> $\alpha_0 > 0$, $\alpha_1 > 0$ ve c_0, c_1 kısıtsız ise

Bulanık mesafenin α_0, α_1, c_0 ve c_1'e göre 1. dereceden türevi alınarak sıfıra eşitlenirse, α_0, α_1, c_0 ve c_1 değerleri elde edilir:

$$\partial\, r(\tilde{A}_0 + \tilde{A}_1 X_{i1}, Y_i) / \partial \alpha_0 = 2\Sigma_i \{ \alpha_0 - c_0 + (\alpha_1 - c_1) X_{i1} - (y_i - e_i) + \alpha_0 + c_0 + (\alpha_1 + c_1) X_{i1} - (y_i + e_i) + \alpha_0 + \alpha_1 X_{i1} - y_i \} \quad (1.22)$$

$$\partial\, r(\tilde{A}_0 + \tilde{A}_1 X_{i1}, Y_i) / \partial c_0 = 2\Sigma_i \{ \alpha_0 - c_0 + (\alpha_1 - c_1) X_{i1} - (y_i - e_i)(-1) + \alpha_0 + c_0 + (\alpha_1 + c_1) X_{i1} - (y_i + e_i) + \alpha_0 + \alpha_1 X_{i1} - y_i \} \quad (1.23)$$

$$\partial\, r(\tilde{A}_0 + \tilde{A}_1 X_{i1}, Y_i) / \partial \alpha_1 = 2\Sigma_i X_{i1} \{ \alpha_0 - c_0 + (\alpha_1 - c_1) X_{i1} - (y_i - e_i) + \alpha_0 + c_0 + (\alpha_1 + c_1) X_{i1} - (y_i + e_i) + \alpha_0 + \alpha_1 X_{i1} - y_i \} \quad (1.24)$$

$$\partial\, r(\tilde{A}_0 + \tilde{A}_1 X_{i1}, Y_i) / \partial c_1 = 2\Sigma_i X_{i1} \{ - [\alpha_0 - c_0 + (\alpha_1 - c_1) X_{i1} - (y_i - e_i)] + [\alpha_0 + c_0 + (\alpha_1 + c_1) X_{i1} - (y_i + e_i)] \} \quad (1.25)$$

Yukarıdaki denklemler sıfıra eşitlendiğinde α_0, α_1, c_0 ve c_1 değerleri hesaplanmış olur. Sonuçlar aşağıdaki gibidir:

$$\alpha_1 = (N\Sigma_i X_{i1} y_i - \Sigma_i X_{i1} \Sigma_i y_i) / (N\Sigma_i X_{i1}^2 - (\Sigma_i X_{i1})^2 \quad (1.26)$$

$$\alpha_0 = \bar{y} - \alpha_1 \bar{X} \quad (1.27)$$

$$c_1 = (N\Sigma_i X_{i1} e_i - \Sigma_i X_{i1} \Sigma_i e_i) / (N\Sigma_{i=1} X_{i1}^{\,2} - (\Sigma_i X_i)^2 \quad (1.28)$$

$$c_0 = \bar{e} - c_1 \bar{X} \quad (1.29)$$

2. Durum: $\alpha_0 < 0$, $\alpha_1 > 0$, $c_0 < 0$ ve c_1 kısıtsız ise

$$\alpha_1 = (N\Sigma_i X_{i1}\, y_i - \Sigma_i X_{i1}\, \Sigma_i y_i) / (N\Sigma_i X_{i1}^2 - (\Sigma_i X_{i1})^2 \qquad (1.30)$$

$$\alpha_0 = \bar{y} - \alpha_1 \bar{X} \qquad (1.31)$$

$$c_1 = (N\Sigma_i X_{i1}\, e_i - \Sigma_i X_{i1}\Sigma_i e_i) / (N\Sigma_{i=1} X_{i1}^{\,2} - (\Sigma_i X_i)^2 \qquad (1.32)$$

$$c_0 = -(\bar{e} - c_1 \bar{X}) \qquad (1.33)$$

3. Durum: $\alpha_0 > 0$, $\alpha_1 < 0$, c_0 kısıtsız ve $c_1 < 0$ ise

$$\alpha_1 = (N\Sigma_i X_{i1}\, y_i - \Sigma_i X_{i1}\, \Sigma_i y_i) / (N\Sigma_i X_{i1}^2 - (\Sigma_i X_{i1})^2 \qquad (1.34)$$

$$\alpha_0 = \bar{y} - \alpha_1 \bar{X} \qquad (1.35)$$

$$c_1 = (N\Sigma_i X_{i1}\, e_i - \Sigma_i X_{i1}\Sigma_i e_i) / (N\Sigma_{i=1} X_{i1}^{\,2} - (\Sigma_i X_i)^2 \qquad (1.36)$$

$$c_0 = \bar{e} - c_1 \bar{X} \qquad (1.37)$$

4. Durum: $\alpha_0 < 0$, $\alpha_1 < 0$, $c_0 < 0$ ve $c_1 < 0$ ise

$$\alpha_1 = (N\Sigma_i X_{i1}\, y_i - \Sigma_i X_{i1}\, \Sigma_i y_i) / (N\Sigma_i X_{i1}^2 - (\Sigma_i X_{i1})^2 \qquad (1.38)$$

$$\alpha_0 = \bar{y} - \alpha_1 \bar{X} \qquad (1.39)$$

$$c_1 = (N\Sigma_i X_{i1}\, e_i - \Sigma_i X_{i1}\Sigma_i e_i) / (N\Sigma_{i=1} X_{i1}^{\,2} - (\Sigma_i X_i)^2 \qquad (1.40)$$

$$c_0 = -(\bar{e} - c_1 \bar{X}) \qquad (1.41)$$

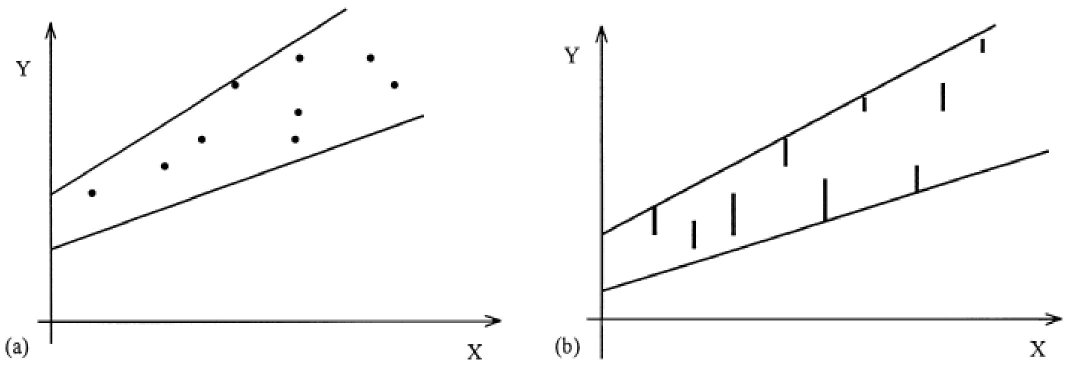
Tek değişkenli bir bulanık regresyon modeline ilişkin bulanık en küçük kareler yöntemi yukarıdaki gibi uygulanmıştır. Değişken sayısı arttığında bu yöntemi uygulamak da zorlaşacaktır. c_0, c_1 ve e_i değerleri sıfıra eşit olduğunda, bulanık en küçük kareler yöntemi, klasik en küçük kareler yöntemi ile aynı olacaktır. Bu yöntemde, minimize edilen regresyon aralıkları Tanaka'nın doğrusal programlama modeline göre daha dardır. Tahmin yönünden avantajlı olmasının yanında, bulanık en küçük kareler yönteminin dezavantajı oldukça uzun hesaplamalar gerektirmesidir (Wang ve Tsaur, 2000b:640-643).

Bulanık regresyon için geliştirilen en küçük kareler yöntemine ek olarak bulanık aralıkları tanımlamak için ortaya konulan yöntemlerden yaygın kullanılan bir yöntem de "*Aralık Regresyonu*" dur. Çalışmanın sonraki bölümünde aralık regresyonu anlatılacaktır.

1.4.3. Aralık Regresyonu

Bu yöntemde bulanık veriler ve bulanık regresyon katsayıları, aralık sayıları gibi davranırlar. Aralık işlemleri bulanık regresyon yöntemine uygulanır ve böylece yöntem "Aralık Regresyon Analizi" adını alır. Bulanık regresyon katsayıları, tüm bulanık çıktılar bulanık regresyon modelindeymiş gibi hesaplanır. Aşağıda Şekil 3'de Regresyon Aralıkları grafiksel olarak gösterilmektedir. Şekil 3(a)'da kesin *X* ve kesin *Y* verileri için regresyon aralığı, Şekil 3(b)'de ise kesin *X* ve bulanık *Y* için regresyon aralığı gösterilmiştir (Chang ve Ayyub, 2001: 192).

Şekil 3: Regresyon Aralıkları

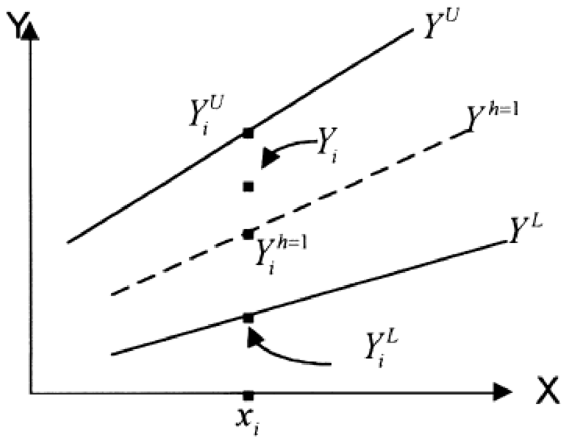


Kaynak: Chang ve Ayyub, 2001: 192

Klasik regresyonda, araştırmacılar bağımlı değişkenlerin kesin değerini doğru bir şekilde tahmin etmek için bir regresyon modeli kullanırlar. Fakat bulanık doğrusal regresyonda, bir aralık tanımlamak tanımlanan bu aralığın tahminsel bir anlamı olduğunu garanti etmez. Tanaka kendi yaptığı çalışmasında, tahmin edilen fiyat(cevap değişkeni) bulanık bir küme olduğu için, karar vericinin kendi insiyatifi ile tahmin edilen bulanık kümenin dışında bir fiyat seçebileceğini söylemiştir. Bu durum, bağımlı değişkenin değerini tahmin etmek için bir $\tilde{Y}_i$ aralığının kullanılmasını getirmiştir. $\tilde{Y}_i$ aralığı, Y_i^L ve Y_i^U, alt ve üst sınırları arasındaki sonsuz noktalarda yer almaktadır (Wang ve Tsaur, 2000a:357).

Bulanık doğrusal regresyon modelinde, her x_i değişkenine karşılık gelen $\tilde{Y}_i$ bulanık tahminleri için, n_i kadar nokta içeren $\tilde{Y}_i$ aralığı bulunmaktadır. $\tilde{Y}_i$ aralığı da, Δ eşit mesafesine sahip en alt değer Y_i^L ve en üst değer Y_i^U sınırları arasındadır *(i=1,2,…M)*. $\tilde{Y}_i$ bulanık regresyon aralığının Y_i^L ve Y_i^U, alt ve üst sınırları arasındaki konumu aşağıdaki gibi Şekil 4'te grafiksel olarak gösterilmiştir (Wang ve Tsaur, 2000a:357).

Şekil 4: Bulanık Regresyon Aralığı



Kaynak: Wang ve Tsaur, 2000a:358

$\tilde{Y}_i$ aralığının toplam değeri = $\int_{Y_i^L}^{Y_i^U} 1dY$ olur. Eşit Δ aralıkları ile n_i kadar noktada hesaplanırsa, $\tilde{Y}_i$ aralığının toplam değeri aşağıdaki gibi olmaktadır:

$$= Y_i^L + (Y_i^L+\Delta) + (Y_i^L +2\Delta) +\ldots+ (Y_i^L+ n_i\Delta) = (n_i+1)\, Y_i^L + (\Delta+\ldots+n_i\Delta)$$
$$=(n_i+1)\, Y_i^L + [n_i(n_i+1)/ 2]\, \Delta. \qquad (1.42)$$

Buna göre $\tilde{Y}_i$'nin ortalama değeri aşağıdaki gibi hesaplanır: (Wang ve Tsaur, 2000a:357)

$$\bar{Y}_i = [1/(n_i + 1)]\, \{(n_i + 1)Y_i^L + [n_i(n_i + 1)/2]\, \Delta\} \qquad (1.43)$$
$$= Y_i^L + (n_i\, \Delta)/2 = Y_i^L/2 + (Y_i^L + n_i\Delta)/2$$
$$=(Y_i^L + Y_i^U)/ 2$$

Aralık regresyonunda çeşitli h değerleri ile yani çeşitli bulanıklık seviyelerinde $\tilde{Y}_i$ bulanık parametreleri belirlenerek regresyon Y^L ve Y^U alt ve üst sınırları elde edilmeye çalışılır. Tahmin edilen parametrelerin optimum bir $\tilde{Y}_i$ aralığı içersinde yer alması istenir. Bunun için h seviyesinin 1 olması gerektiğini ($Y_i^{h=1}$), yani bulanık regresyonun 1 olasılığıyla gerçekleştiğini savunan teoremler mevcuttur.

Bu doğrultuda bulanık doğrusal regresyon aralığının belirlenmesi gerekmektedir, bunun için aşağıdaki bölümde bahsedilen teorem ele alınmıştır.

1.4.3.1. Bulanık Doğrusal Regresyon Aralığının Belirlenmesi

Teorem 1.1: Bulanık doğrusal regresyon modelinde bulanık parametrelerin üyelik fonksiyonları simetrik ise, $Y_i^{h=1}$ değerleri $\bar{Y}_i$'ye eşittir (Wang ve Tsaur, 2000a:358).

İspat: $Y_i^L = Y_i^{h=1} - c^t \mid X_i \mid$ (1.44)

$Y_i^U = Y_i^{h=1} + c^t \mid X_i \mid$ (1.45)

$$\bar{Y}_i = (Y_i^L + Y_i^U)/ 2 = [(Y_i^{h=1} - c^t \mid X_i \mid) + (Y_i^{h=1} + c^t \mid X_i \mid)] / 2 = Y_i^{h=1} \quad (1.46)$$

Burada M kadar nokta içeren veri kümesi ele alınmaktadır. (X_1, y_1), (X_2, y_2),…, (X_M, y_M) ve $Y_i^L \le Y_i \le Y_i^U$ olmak üzere ;

$Y_i = Y_i^L + a_i \Delta$ ve $Y_i^{h=1} = Y_i^L + n_i / 2\Delta$'dir. (1.47)

Y_i^L ve Y_i^U arasındaki fark $= n_i \Delta = (Y_i - Y_i^L) + (Y_i^U - Y_i)$ (1.48)

$$= (Y_i^{h=1} - Y_i^L) + (Y_i^U - Y_i^{h=1})$$

$\Sigma_i [(Y_i - Y_i^L) + (Y_i^U - Y_i)]^2 = \Sigma_i [(Y_i^{h=1} - Y_i^L) + (Y_i^U - Y_i^{h=1})]^2$ olur. (1.49)

Böylece;

$$\Sigma_i (Y_i - Y_i^L)^2 + \Sigma_i (Y_i^U - Y_i)^2 + 2\, \Sigma_i [(a_i \Delta) (n_i \Delta - a_i \Delta)] \quad (1.50)$$
$$= \Sigma_i (Y_i^{h=1} - Y_i^L)^2 + \Sigma_i (Y_i^U - Y_i^{h=1})^2 + 2 [(n_i \Delta / 2) (n_i \Delta / 2)]$$

$\Sigma_i (Y_i - Y_i^L)^2 + \Sigma_i (Y_i^U - Y_i)^2$ (1.51)

$=\Sigma_i (Y_i^{h=1} - Y_i^L)^2 + \Sigma_i (Y_i^U - Y_i^{h=1})^2 + 2\Sigma_i (Y_i^{h=1} - Y_i)^2$ elde edilir.

SST = SSR + SSE

Total Sum of Squares = Regression Sum of Squares + Error Sum of Squares

Tüm Hata Kareleri Toplamı = Regresyon Kareleri Top. + Hata Kareleri Top.

Buna göre, bulanık regresyon aralıkları için hata kareleri toplamları ayrı ayrı aşağıdaki gibi belirlenir:

$SST = \Sigma_i (Y_i - Y_i^L)^2 + \Sigma_i (Y_i^U - Y_i)^2$ (1.52)

$SSR = \Sigma_i (Y_i^{h=1} - Y_i^L)^2 + \Sigma_i (Y_i^U - Y_i^{h=1})^2$ (1.53)

$SSE = 2\Sigma_i (Y_i^{h=1} - Y_i)^2$ (1.54)

SST, $\tilde{Y}_i$'nin alt ve üst sınırları arasındaki değişimi ölçer. *SSE*, $\tilde{Y}_i$'nin tahmin edilmesinde $Y_i^{h=1}$ kullanıldığı zaman ortaya çıkan değişimi, *SSR* ise alt ve üst sınırlara göre $Y^{h=1}$'deki değişimi göstermektedir.

Ancak, aralık regresyonunda Y_i değerini tahmin etmek için en iyi seçeneğin $Y_i^{h=1}$ doğrusu olup olmadığı belirlenmelidir. Bunu araştırmak için aşağıdaki teorem incelenmiştir.

Teorem 1.2: Bulanık doğrusal regresyonda, bulanık parametre $\tilde{A}_j$'nin üyeliği simetrik olduğunda, Y_i'yi tahmin eden en iyi regresyon doğrusu $Y^{h=1}$'dir. (Wang ve Tsaur, 2000a:359)

İspat: $h = \alpha$ ve $\alpha \neq 1$ iken burada $Y^{h=1}$ doğrusuna simetrik $Y^{U,h=\alpha}$ ve $Y^{L,h=\alpha}$ alt ve üst regresyon doğruları oluşur. Veri kümesinin *M* kadar nokta içerdiği düşünülerek, (X_1, y_1), (X_2, y_2),…, (X_M, y_M) ve $Y_i^L \leq Y_i \leq Y_i^U$ olmak üzere ;

$Y_i = Y_i^L + a_i \Delta$ ve $Y_i^{h=1} = Y_i^L + n_i / 2\Delta$'dir. (1.55)

Buna göre, $Y^{h=1}$ doğrusunun tahmin yeteneğini ölçmek için $Y^{U,h=\alpha}$ ele alındığında aşağıdaki eşitlikler elde edilir:

$$Y^{U,h=\alpha} = Y_i^{h=1} + (1-\alpha)(Y_i^U - Y_i^{h=1}) \tag{1.56}$$
$$= Y_i^L + n_i\Delta/2 + [(1-\alpha)\, n_i\Delta/2] = Y_i^L + 2\, n_i\Delta - (n_i\Delta/2)\alpha$$
$$= Y_i^L + n_i\Delta/2 - (2-\alpha)$$

$$(Y_i - Y_i^L) + (Y_i^U - Y_i) = n_i\Delta = (Y_i^{h=\alpha} - Y_i^L) + (Y_i^U - Y_i^{h=\alpha}) \tag{1.57}$$

$$\Sigma_i\, [(Y_i - Y_i^L) + (Y_i^U - Y_i)]^2 = \Sigma_i\, [(Y_i^{h=\alpha} - Y_i^L) + (Y_i^U - Y_i^{h=\alpha})]^2 \tag{1.58}$$

$$\Sigma_i\, (Y_i - Y_i^L)^2 + \Sigma_i\, (Y_i^U - Y_i)^2 + 2\, \Sigma_i[(a_i\Delta)\, (n_i\Delta - a_i\Delta)] \tag{1.59}$$
$$= \Sigma_i\, [(Y_i^{h=\alpha} - Y_i^L)^2 + \Sigma_i\, (Y_i^U - Y_i^{h=\alpha})^2 + 2\, \Sigma_i\, [(n_i\Delta/2)(2-\alpha)\, (n_i\Delta/2)\alpha] \text{ olur.}$$

Böylece;

$$2\Sigma_i\, [(n_i\Delta/2)(2-\alpha)\, (n_i\Delta/2)\alpha] + 2\Sigma_i\, [(a_i\Delta)\, (n_i\Delta - a_i\Delta)] \tag{1.60}$$
$$= 2\Sigma_i\, [(2\alpha - \alpha^2)\, (n_i\Delta/2)^2 - a_i\, n_i\Delta^2 + a_i^2\Delta^2]$$

$$= 2\Sigma_i\, [(n_i\Delta/2 - a_i\Delta)^2 + (1-\alpha)^2 (n_i\Delta/2)^2] = 2\Sigma_i\, [(Y_i^{h=1} - Y_i)^2 + (1-\alpha)^2 (Y_i^{h=1} - Y_i^L)^2$$

$$\Sigma_i\, (Y_i - Y_i^L)^2 + \Sigma_i\, (Y_i^U - Y_i)^2 \tag{1.61}$$
$$= \Sigma_i\, [(Y_i^{U,h=\alpha} - Y_i^L)^2 + \Sigma_i\, (Y_i^U - Y_i^{U,h=\alpha})^2 + 2\Sigma_i\, [(Y_i^{h=1} - Y_i)^2 + (1-\alpha)^2 (Y_i^{h=1} - Y_i^L)^2$$

$$SST = SSR^{\wedge} + SSE^{\wedge}$$

Buna göre yeni oluşan regresyon kareler toplamı $SSR^{\wedge}$ ve hata kareler toplamı $SSE^{\wedge}$ aşağıdaki gibi belirlenir. Bu durumda $SSE^{\wedge} \geq SSE$ ve $SSR^{\wedge} \geq SSR$ olduğu görülmektedir.

$$SST = \Sigma_i (Y_i - Y_i^L)^2 + \Sigma_i (Y_i^U - Y_i)^2 \quad (1.62)$$

$$SSR^{\wedge} = \Sigma_i [(Y_i^{U,h=\alpha} - Y_i^L)^2 + \Sigma_i (Y_i^U - Y_i^{U,h=\alpha})^2 \quad (1.63)$$

$$SSE^{\wedge} = 2\Sigma_i [(Y_i^{h=1} - Y_i)^2 + (1-\alpha)^2 (Y_i^{h=1} - Y_i^L)^2 \quad (1.64)$$

Yapılan ispatın sonucu olarak, $h = 1$ seviyesinin yani $Y_i^{h=1}$ doğrusunun tüm α seviye regresyon doğruları $Y_i^{h=\alpha}$ arasında en iyi tahmini verdiği anlaşılmaktadır (Wang ve Tsaur, 2000a:360-361).

Aralık regresyonunda Y_i^L ve Y_i^U, alt ve üst regresyon doğruları $h=\alpha$ seviyesinde hesaplanan $Y_i^{h=\alpha}$ doğrusunu kapsamaktadırlar. Bu nedenle $Y_i^{h=\alpha}$ doğrusuna ulaşabilmek ve $\tilde{Y}_i$ aralığını belirleyebilmek için aralık regresyonunda alt ve üst regresyon doğrularının belirlenmesi gerekmektedir.

1.4.3.2. Alt ve Üst Regresyon Doğrularının Belirlenmesi

Aralık regresyonunda, $\tilde{Y}_i$ bulanık bağımlı değişkeninin tahmin edilmesi için belirlenen $\tilde{Y}_i$ aralığında Y_i^L ve Y_i^U, alt ve üst regresyon doğruları bu aralığın sınırlarını oluşturmaktadır. $h=\alpha$ seviyesinde hesaplanan $Y_i^{h=\alpha}$ tahmin doğrusunu kapsayan bu sınırları belirlemek için bir doğrusal programlama modeli çözümünün yapılması gerekmektedir (Chang ve Ayyub, 2001: 191).

$\hat{Y} = \tilde{A}_0 + \tilde{A}_1 X$ bulanık denklemi için bulanık regresyon katsayıları $\tilde{A}_0 = (\alpha_0, c_0)$ ve $\tilde{A}_1 = (\alpha_1, c_1)$ değerlerini belirlemek için aşağıdaki doğrusal programlama modeli kullanılır:

Minimum $nc_0 + c_1 \sum_{i=1}^{n} X_i$

Kısıtları altında $c_0 \geq 0$ ve $c_1 \geq 0$

$$(\alpha_0 - c_0) + (\alpha_1 - c_1)X_i \leq Y_i^L, \quad i=1,2,\ldots,n$$

$$(\alpha_0 + c_0) + (\alpha_1 + c_1)X_i \geq Y_i^U$$

Yukarıdaki doğrusal programlama modeline göre, Y_i^L ve Y_i^U her bulanık veri için alt ve üst sınırı göstermektedir. Amaç toplam bulanık yayılımın belirlenen kısıtlar altında minimize edilmesidir. Bu kısıtlar bulanık regresyon modelindeki tüm gözlenen değerlerin Y_i^L ve Y_i^U sınırları arasında kalmasını sağlamak için belirlenmiştir (Chang ve Ayyub, 2001:192).

Aralık regresyonunda böylece bu alt ve üst sınırlar arasında kalan bir $\tilde{Y}_i$ aralığına ulaşılması mümkün olmaktadır. Belirlenen $\tilde{Y}_i$ aralığı da, bulanık bağımlı değişken $\tilde{Y}_i$ için bir tahmin yapılmasını sağlamaktadır.

Bulanık regresyon analizi üzerine başka birçok çalışma bulunmaktadır. Çoğu çalışma sadece bağımlı değişkenin bulanık olması üzerinde durmuştur. Yalnızca birkaç tanesi hem bağımlı hem de bağımsız değişkenin bulanık olduğu durum üzerinde tartışmıştır (Kao ve Chyu, 2002:402). 1992 yılında, Sakawa ve Yano bulanık doğrusal regresyon modellerini bulanık girdi ve bulanık çıktı için incelemişlerdir. Modeldeki bulanık parametreleri belirlemek için, doğrusal programlama yaklaşımını kullanarak üç çeşit çok amaçlı programlama yöntemi geliştirmişlerdir (Nasrabadi ve Nasrabadi, 2004:874). Sakawa ve Yano, Tanaka modelindeki tahmin edilen aralığın tamamen gözlenen aralığı kapsaması gerekliliğini esneterek, bu aralığın sadece bir kısmını kapsamasının yeterli olacağını düşünmüşlerdir (Hojati, Bector ve Smimou, 2004:175). Bu tez çalışmasında girdisi kesin, çıktısı bulanık veriler üzerinde durulduğu için çalışmada bu modele yer verilmemiştir.

İKİNCİ BÖLÜM

2. LOJİSTİK REGRESYON ANALİZİ

İstatistiksel uygulamalarda, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi tanımlayabilmek amacıyla geliştirilen, yaygın olarak kullanılan alternatif tahminleme yöntemlerden birisi de lojistik regresyon analizidir. Son zamanlarda lojistik regresyon analizi kullanım kolaylığının yanında, sayısal olarak rahat yorumlanabilmesiyle ön plana çıkmış ve birçok uygulamada sıklıkla kullanılan bir yöntem haline gelmiştir.

Sonuç değişkeninin kesikli olması ve iki ya da daha fazla olası değer alması sık karşılaşılan bir durumdur. Son yıllarda çoğu alanda, böyle bir durum söz konusu olduğunda lojistik regresyon modelinin kullanılması, standart bir analiz yöntemi haline gelmiştir. Öncelikle anlaşılmalıdır ki; lojistik regresyon analizinin kullanım amacı istatistikte kullanılan diğer model yapılandırma teknikleriyle aynıdır. Amaç, en az değişkeni kullanarak en iyi uyuma sahip olacak şekilde sonuç değişkeni (bağımlı ya da cevap değişkeni) ile bağımsız değişkenler kümesi (açıklayıcı değişkenler) arasındaki ilişkiyi tanımlayabilen kabul edilebilir bir model kurmaktır (Lemeshow ve Hosmer, 2000:1).

Neden sonuç ilişkilerinin ortaya konulması amacıyla yapılan çoğu sosyo-ekonomik araştırmada, incelenen değişkenlerden bazıları olumlu-olumsuz, başarılı-başarısız, evet-hayır, memnun-memnun değil şeklinde iki düzeyli verilerden oluşmaktadır. Bu türde bağımlı değişkenin iki düzeyli ya da çok düzeyli kategorik verilerden oluşması durumunda; bağımlı değişken ile bağımsız değişken(ler) arasındaki neden-sonuç ilişkisinin incelenmesinde, lojistik regresyon analizi önemli bir yere sahiptir (Girginer ve Cankuş, 2008:185).

Lojistik regresyon modeli kategorik verilerin analizi için en önemli modeldir. Giderek artan, çok geniş bir uygulama alanı vardır. Önceleri sadece biyomedikal çalışmalarda kullanılsa da son 20 yıldır sosyal bilimler araştırmaları ve pazarlama alanlarında çoğunlukla kullanıldığı görülmektedir. Son dönemlerde lojistik regresyon işletme uygulamalarında popüler bir araç haline gelmiştir. Bazı kredi değerlendirme kuruluşları kendilerine gelen başvuruların kredilendirmeye değer olma olasılığını lojistik regresyon modelini kullanarak belirlemektedir (Agresti, 2002:165).

Son yıllarda yoğun bir şekilde kullanılan lojistik regresyon analizi, gözlemlerin gruplara atanmasında sık kullanılan üç yöntemden (diğerleri kümeleme analizi ve diskriminant analizi) birisidir. Kümeleme analizinde gözlemlerin atanacağı küme sayısı tam bilinmezken, diskriminant ve lojistik regresyon analizinde grup sayısı bilinmekte, mevcut veriler kullanılarak bir ayrımsama modeli elde edilmekte ve kurulan bu model yardımıyla veri kümesine eklenen yeni gözlemlerin gruplara atanması mümkün olabilmektedir (Bircan, A.Coşkun, S.Coşkun ve Kartal, 2004:42). Diskriminant analizi, bağımlı değişkenin nonmetrik olduğu durumlarda uygulanabilmektedir. Ancak bağımlı değişken sadece iki cevaplı olduğu zaman lojistik regresyon analizi şu nedenlerden dolayı tercih edilebilir. Birincisi diskriminant analizi, çoklu normal dağılım ve gruplar arası eşit varyans-kovaryans matrisi varsayımlarına dayalıdır. Lojistik regresyon ise bu katı varsayımları gerektirmez ve bu varsayımların gerçekleşmediği durumlarda çok daha sağlıklı sonuçlar verir. İkincisi bu varsayımlar gerçekleşmiş olsa bile bir çok araştırmacı regresyona benzerliğinden dolayı lojistik regresyonu tercih eder (Ulupınar, 2007:39).

Lojistik regresyon analizi; normallik, ortak kovaryansa sahip olma gibi çeşitli varsayımların sağlanamaması durumunda diskriminant analizi ve çapraz tablolara alternatif bir yöntemdir. Bağımlı değişkenin 0 ve 1 gibi ikili (binary) ya da ikiden çok düzey içeren kesikli değişken olması durumunda da, normallik varsayımının sağlanamaması nedeniyle doğrusal regresyon analizine alternatif olmaktadır. Ayrıca elde edilen modelin matematiksel olarak çok esnek olması ve kolay yorumlanabilir olması bu yönteme olan ilgiyi artırmaktadır (Bircan, 2004:187).

Lojistik regresyon analizi değişkenler arasında çoklu bağlantı olmadığını varsayar. Yani herhangi bir değişken diğer değişkenlerin doğrusal kombinasyonu seklinde yazılmamalıdır. Gözlem sayısının az olması çoklu bağlantı olasılığını arttırır. Bu yüzden çoklu bağlantı probleminin araştırılması ve düzeltilmesi gerekmektedir. Ancak araştırmacının amacı birim veya bireyi uygun sınıfa yerleştirmekse çoklu bağlantı problemi ihmal edilebilir (Ulupınar, 2007:40).

Lojistik regresyon yöntemi, sınıflama ve atama işlemi yapmaya yardımcı olan bir regresyon yöntemidir. Normal dağılım varsayımı, süreklilik varsayımı önkoşulu yoktur. Bağımlı değişken üzerinde açıklayıcı değişkenlerin etkileri olasılık olarak elde edilerek, risk faktörlerinin olasılık olarak belirlenmesi sağlanır (Özdamar, 2002:475).

Lojistik regresyon modelleri bağımlı değişkenin yapısına göre üçe ayrılmaktadır. Bu modellerden İkili (Binary) Lojistik Regresyon Modeli, kategorik bağımlı değişkenin iki durumlu (örn: var-yok) olduğu durumda kullanılmaktadır. Çoklu (Multinomial) Lojistik Regresyon modeli kategorik bağımlı değişkenin çok kategorili ve nominal (örn: medeni durum: evli- bekar- boşanmış) olduğu durumlarda kullanılırken; çok kategorili ve sıralı bir yapı söz konusu ise (örn: likert tipi ölçekler, az-orta-çok) Sıralı (Ordinal) Lojistik Regresyon modelleri kullanılmaktadır (Kaşko, 2007:18). Bu tez çalışmasında uygulamalarda en çok yer bulan İkili Lojistik Regresyon Analizi üzerinde durulmuş ve çalışmanın bu bölümünde ikili modelin teorik anlatımına yer verilmiştir.

Lojistik regresyon analizi ile doğrusal regresyon analizleri yaklaşım olarak birbirlerine benzemekle beraber aralarında önemli farklar da bulunmaktadır. Bağımlı değişkenin ikili veya çoklu olduğu durumlarda kullanılan lojistik regresyon, doğrusal regresyondan model ve varsayımları yönüyle farklılaşmaktadır. Lojistik regresyonu doğrusal regresyondan ayıran temel farklılıklar alt bölümde verilmiştir.

2.1. Lojistik Regresyon Analizinin Doğrusal Regresyonla İlişkisi

Lojistik regresyonu doğrusal regresyondan ayıran en belirgin özellik lojistik regresyonda sonuç değişkenin ikili veya çoklu olmasıdır. Lojistik regresyon ve doğrusal regresyon arasındaki bu fark hem parametrik model seçimine, hem de varsayımlara yansımaktadır (Lemeshow ve Hosmer, 2000:1).

Bir değişkenin bağımsız değişkenlere bağlı olarak hangi değeri alacağını tahminlemede doğrusal regresyon analizi kullanılırken, iki sonuçlu bir değişkenin bağımsız değişkenlere bağlı olarak hangi değerleri alacağının olasılığının tahminlenmesinde lojistik regresyon analizi kullanılmaktadır (Özdemir, 2008:278).

Lojistik regresyonda da, doğrusal regresyon analizinde olduğu gibi bazı değişken değerlerine dayanarak tahmin yapılmaya çalışılır. Ancak bu iki yöntem arasında üç önemli fark vardır;

1. Doğrusal regresyon analizinde tahmin edilecek olan bağımlı değişken sürekli iken, Lojistik regresyon analizinde bağımlı değişken kesikli bir değer almaktadır.
2. Doğrusal regresyon analizinde bağımlı değişkenin değeri, Lojistik regresyon analizinde ise bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı tahmin edilir.
3. Doğrusal regresyon analizinde bağımsız değişkenin çoklu normal dağılım göstermesi şartı aranırken, Lojistik regresyon analizinde böyle bir şart yoktur (Bircan, 2004:187)

Doğrusal regresyon yönteminden hareketle lojistik regresyon modelinin oluşturulması modelin anlaşılmasına yardımcı olacaktır. Lojistik regresyon analizine başlarken öncelikle tahminleme sorununda kullanılacak lojistik modelin oluşturulması gerekmektedir.

2.2. Lojistik Regresyon Modeli

Herhangi bir regresyon probleminde en önemli nicelik, verilen bir bağımsız değişken değeri için sonucun ortalama değeridir. Bu nicelik "koşullu ortalama" diye adlandırılır ve "$E(y \mid x)$" şeklinde gösterilir. "y" sonuç değişkenini, "x" ise bağımsız değişkenin değerini simgeler. $E(y \mid x)$ niceliği, "verilen x değeri için, y'nin beklenen değeri" olarak ifade edilir. Doğrusal regresyonda, koşullu ortalama x'e göre doğrusal bir eşitlikle gösterilebilir. Şöyle ki;

$$E(y \mid x) = \beta_0 + \beta_1 x \qquad (2.1)$$

(2.1) ifadesine göre x, $-\infty$ ve $+\infty$ arasında değişen değerler alırken, $E(y \mid x)$ mümkün olan herhangi bir değeri alabilmektedir.

Bağımlı değişken ikili olduğu zaman koşullu ortalama, 0 ile 1 arasında değişmek durumundadır [$0 \leq E(y \mid x) \leq 1$]. x 'deki her birim değişme sonucunda $E(y \mid x)$'de gerçekleşen değişim, koşullu ortalama 0'a ya da 1'e yaklaştıkça gittikçe azalmaktadır.

İkili bir bağımlı değişkenin analizinde kullanılmak üzere birçok dağılım fonksiyonu önerilmiştir. Lojistik dağılımın seçilmesi için iki önemli neden bulunmaktadır. Bu nedenlerden birisi matematiksel açıdan bakıldığında, lojistik dağılımın çok esnek ve kolay kullanılan bir fonksiyon olması, ikincisi ise dağılımın kendisinin biyolojik olarak mantıklı bir ifade olmasıdır.

Gösterimi basitleştirmek için, $\pi(x) = E(y \mid x)$ ifadesi, yani verilen x'e göre y'nin koşullu ortalaması yerine kullanılsın. Bu durumda lojistik regresyon modelinin formu şöyledir;

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \qquad (2.2)$$

Lojistik regresyon çalışmasına merkez olacak "logit dönüşüm" $\pi(x)$'deki dönüşümdür. (Koşullu ortalamanın değerinin 0 ile 1 arasında olması kısıtının sağlanması için bu dönüşüm yapılmalıdır.) $\pi(x)$'in bu dönüşümü şöyle tanımlanır;

$$g(x) = ln\left[\frac{\pi(x)}{1-\pi(x)}\right] \tag{2.3}$$

$$= \beta_0 + \beta_1 x$$

Bu dönüşümün önemi $g(x)$'in, doğrusal regresyon modeli için uygun özelliklerin birçoğuna sahip olmasıdır. Kendi parametrelerinde doğrusal olan $g(x)$ logiti, sürekli olabilir ve x'in tanım aralığına göre – ∞'dan + ∞'a kadar değer alabilir.

Doğrusal regresyon modelinde sonuç değişkeni için yapılan gözlem $y = E(y \mid x) + \varepsilon$ şeklinde gösterilebilir. ε niceliği "hata" diye adlandırılır ve gözlemin koşullu ortalamadan sapma miktarını ifade eder. ε değeri için en genel varsayım, ε değerinin 0 ortalamasına ve bağımsız değişkenlerin seviyeleri arasında sabit bir varyansa sahip normal dağılımlı bir değer olduğudur. Verilen x için sonuç değişkeninin koşullu dağılımı, $E(y \mid x)$ ortalamasına ve sabit bir varyansa sahip bir normal dağılımdır. Fakat iki sonuçlu bağımlı değişken için durum farklıdır. Bu durumda verilen x için sonuç değişkeni $y = \pi(x) + \varepsilon$ diye ifade edilir. Burada ε 'nin bir veya iki olası değeri vardır:

Eğer $y = 1$ ise, $\pi(x)$ olasılığıyla $\varepsilon = 1- \pi(x)$ olur.
Eğer $y = 0$ ise, $1 - \pi(x)$ olasılığıyla $\varepsilon = - \pi(x)$ olur.

Dolayısıyla ε , 0 ortalamalı ve $\pi(x)[1- \pi(x)]$ varyansına sahip bir dağılım olur. Yani sonuç değişkeninin koşullu dağılımı, $\pi(x)$ koşullu ortalaması tarafından olasılığı verilen bir binom dağılımı olur (Lemeshow ve Hosmer, 2000:5-7).

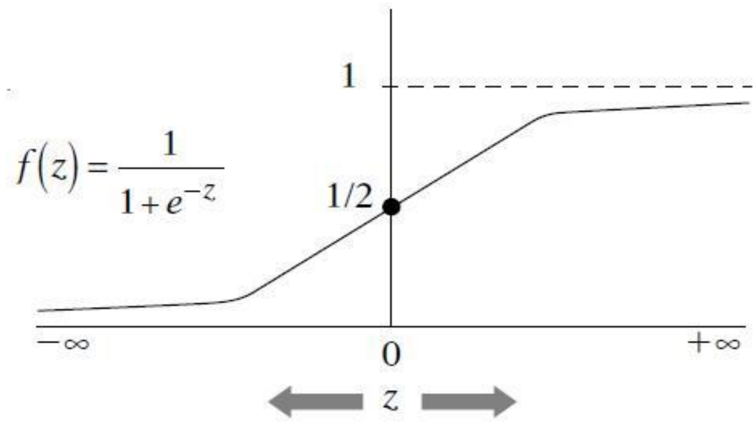
Lojistik regresyon modelinin popüler olmasını sağlayan önemli bir neden dayandığı lojistik fonksiyondur (Kleinbaum, 2002:5). Tüm tahmin modelleri gibi lojistik regresyon modeli de bir tahmin modeli olarak ele alındığında modele bağlı lojistik fonksiyonun tanımlanması gerekmektedir.

2.2.1. Lojistik Fonksiyon

Lojistik fonksiyon, z değerine bağlı olarak f(z) biçiminde tanımlandığında, lojistik model için bir fonksiyon belirlenir. f(z) fonksiyonu lojistik modele bağlı olan lojistik fonksiyonu ifade etmektedir. Bu çerçevede lojistik fonksiyon ve lojistik fonksiyon grafiği aşağıdaki gibi tanımlanır:

$$f(z) = \frac{1}{1+e^{-z}} \qquad (2.4)$$

Şekil 5: Lojistik Fonksiyon Grafiği



Kaynak: Kleinbaum, 2002:5

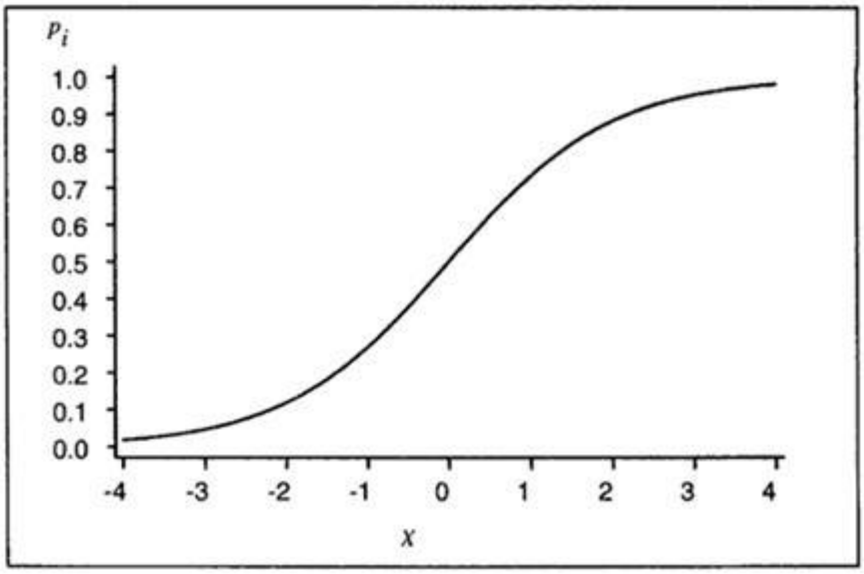
Lojistik fonksiyon grafiği Şekil 5'de gösterilmektedir. Şekil 5'e göre grafiğin sol tarafında *z* değeri, -∞ olduğunda; lojistik fonksiyon f(*z*) = 0 olmaktadır. Grafiğin sağ tarafında ise *z* değeri, +∞ olduğunda; f(z) = 1 olmaktadır.

Yukarıda verilen Şekil 5'deki grafikte görüldüğü gibi, *z*'nin tanım kümesi (-∞, +∞) olmaktadır. Buna karşılık *z*'nin değer kümesi [0,l]'dir. Lojistik fonksiyon f(*z*), *z*'nin değerinden bağımsız olarak, 0 ile 1 aralığında kalmaktadır. Böyle bir model, daima 0 ile 1 aralığında kalan bir olasılığı tanımlamak için tasarlanmıştır. Lojistik modeller, tahminlerin 0 ile 1 aralığında olması gerektiği durumlarda bu kısıtı sağlamaktadır. Bu durum da lojistik modelin kullanılmasını yaygınlaştırmaktadır (Kleinbaum, 2002:5-6).

Lojistik modelin popüler olmasının diğer bir nedeni ise, lojistik fonksiyonun şeklidir. Şekil 5'teki grafikte görüldüğü gibi; *z* değeri, -∞'dan başlayarak artarken f(*z*)'nin değeri 0'a yaklaşmaktadır, daha sonra 1'e doğru artmakta ve *z*, +∞'a geldiğinde f(*z*), 1'de dengelenmektedir. Bu da "S" şeklinde bir fonksiyon oluşturmaktadır. S şeklindeki lojistik fonksiyon; eğer *z* değişkeninin, risk faktörlerinin katkılarını temsil eden bir değer olduğu düşünülürse, f(*z*)'nin de verilen *z* değeri için, riski temsil ettiğini göstermektedir. f(*z*)'nin S şekli, *z* değeri düşük seviyelerde olduğunda riskin de belli bir seviyeye kadar minimum olduğunu göstermektedir. Fakat bu seviye aşıldığında (kıvrım noktası), bu riskin hızla artacağını ve *z* değeri arttıkça da oldukça yükselerek 1'e yaklaşacağını ifade etmektedir (Kleinbaum, 2002:6-7).

Lojistik regresyon fonksiyonu düzden ziyade eğri bir sekil aldığı için, fonksiyon *z* değişkenindeki birimsel değişimle etkilenen f(*z*)'in değişim oranına işaret eder. Eğri üzerindeki belirli bir z değerinin tanjantı olan düz doğru, o noktadaki değişim oranını belirler (Agresti, 2002:166). Bunun gibi S şeklindeki fonksiyon gösterimi, *y* bağımlı değişkeninin ikili olduğu durumlarda tahmin modelini açıklamak için uygun bulunmuştur (Ryan, 1997:256). Lojistik regresyonun S şeklindeki grafik gösterimi aşağıda Şekil 6'de verilmektedir.

Şekil 6: Lojistik Regresyon Grafiği



Kaynak: Allison, 2001:15

Lojistik fonksiyon tanımlandıktan sonra lojistik regresyon analizinde bahsedilmesi gereken iki önemli kavram da "odds oranı" ve "logit model" dir. Lojistik modelin anlaşılması için bu iki kavramın ele alınması faydalı olacaktır.

2.2.2. Odds Oranı ve Logit Model

Lojistik model oluşturulurken izlenen yolda "odds oranı" nın elde edilmesi önemli bir yer tutmaktadır. Odds oranı olasılık kavramı ile yakından ilişkilidir. Bir olayın olasılığı olduğu gibi, o olayın bir "odds" u da bulunmaktadır.

Logit modelin anlaşılması için odds oranının açıklanması gerekmektedir. Olasılık, çoğu insanın bir olayın gerçekleşme şansını değerlendirmek için tanımladığı bir ifadedir. Bu şansın değerinin 0 ile 1 aralığında değiştiği otomatik olarak düşünülebilir; buradan 0 olmasının olayın kesinlikle gerçekleşmeyeceği, 1 olmasının ise kesinlikle gerçekleşeceği anlamına geldiği anlaşılmaktadır. Ancak bir olayın gerçekleşme şansını tanımlamanın farklı yolları da bulunmaktadır ve “odds” bunlardan biridir (Allison, 2001:12).

Olasılık ve odds arasındaki basit ilişki aşağıda (2.5) ve (2.6) eşitliğindeki gibi gösterilmektedir. Bir olayın gerçekleşme olasılığı “*p*” ise, olayın gerçekleşmeme olasılığı *1–p*’dir. Olayın odds’u yani bir olayın gerçekleşme ve gerçekleşmeme olasılıklarının oranı “*O*” olarak tanımlanır:

$$O = \frac{p}{1-p} = \frac{olayın\ gerçekleşme\ olasılığı}{olayın\ gerçekleşmeme\ olasılığı} \tag{2.5}$$

$$p = \frac{O}{1+O} \tag{2.6}$$

Tablo 2: Olasılık ve Odds Arasındaki İlişki

Olasılık	Odds
0.1	0.11
0.2	0.25
0.3	0.43
0.4	0.67
0.5	1.00
0.6	1.50
0.7	2.33
0.8	4.00
0.9	9.00

Kaynak: Allison, 2001:12

Olasılık ve Odds arasındaki basit ilişki Tablo 2’de gösterilmektedir. Odds’un 1’den küçük olması, olasılığın 0,5’ten küçük olmasına; odds’un 1’den büyük olması ise olasılığın 0,5'ten büyük olmasına işaret eder. Odds'un en alt sınırı olasılıkta olduğu gibi sıfırdır, fakat odds'ta olasılıktaki gibi bir üst sınır bulunmamaktadır (Allison, 2001:12).

Lojistik modelin oluşturulmasında daha önce bahsedilen logit dönüşüm, odds oranının logaritması alınarak yapılan dönüşümdür. (2.3) eşitliğinde gösterilen logit dönüşümle elde edilen model, "lojistik regresyon model" ini ortaya koymaktadır.

$E(y \mid x) = \beta_0 + \beta_1 x$ şeklinde ifade edilen doğrusal olasılık modelinde olasılıklar 0 ile 1 arasında sınırlanmaktadır fakat doğrusal fonksiyonlar aslında sınırsızdır. Bu sorunun çözümü için olasılığın dönüşüme uğraması ve sınırlarından kurtulması gerekmektedir. Olasılığın odds'a dönüşümü üst sınırı ortadan kaldırır. Odds'un logaritmasını almak ise alt sınırı ortadan kaldırır. Bunu da, açıklayıcı değişkenlerin doğrusal fonksiyonuna eşitleyerek logit model elde edilir; (Allison, 2001:13)

$$\log\left[\frac{p_i}{1-p_i}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \qquad (2.7)$$

(2.7) eşitliğindeki logit modelde yer alan p_i değeri, y =1 olma olasılığıdır. Eşitliğin sol tarafı "logit" ya da "log-odds" olarak ifade edilir (Allison, 2001:13). (2.7) eşitliğinin sağ tarafında hata terimi artık bulunmamaktadır. Çünkü sol tarafı, hata terimini ortadan kaldırmayı sağlayan $E(y \mid x)$'in bir fonksiyonudur. β_1 ise, x'te birim başına meydana gelen değişimin log-odds değerinde yarattığı değişimi göstermektedir. x'te meydana gelen bir birimlik artış, odds değerini " e^{β} " çarpımsal faktör etkisiyle artırmaktadır (Ryan, 1997:257).

Odds değeri - ∞ ile + ∞ arasında değişen bir değişken yaratır. Odds'un doğal logaritması alındığında, diğer bir deyişle logiti elde edildiğinde; odds, 1'den 0'a yaklaştıkça mutlak değer olarak giderek artan ve pozitif yönde 1'den sonsuza yaklaştıkça da yine giderek artan bir değer olur. Eğer y bağımsız değişkeni için odds değerinin doğal logaritması kullanılırsa, tahminlenen olasılığın, minimum ve maksimum değerleri aşması gibi bir problem artık kalmayacaktır (Menard, 2002:13).

Şunu anlamak gerekir ki; olasılık, odds ve logit ifadeleri aslında aynı şeyi açıklamak için oluşan üç farklı yoldur. Olasılık ve odds, genellikle anlaşılması daha kolay yollardır. Olasılığın logitini elde etmek, matematiksel olarak, ikili bağımlı değişkenleri analiz etmek için kullanılan en iyi yoldur (Menard, 2002:13).

Lojistik regresyon analizinde model oluşturulduğunda modelde bilinmeyen parametreler bulunmaktadır, bu parametreler lojistik modelin katsayılarıdır. Bu nedenle, analize devam edebilmek için, oluşturulan lojistik modeldeki parametrelerin tahmini yapılmalıdır.

2.3. Lojistik Modellerde Parametrelerin Tahmini

$(x_i\ ,\ y_i)$, i=1,2,3,…,n ikilisi ile gösterilen, n bağımsız gözlemin olduğu varsayılsın. Burada y_i, iki sonuçlu sonuç değişkenini, x_i, i'nci gözlem için bağımsız değişkenin değerini göstersin. Ayrıca sonuç değişkeni, belli bir özelliğe sahip olup olmamasına göre 0 ve 1 değerleri ile kodlansın. Buna göre (2.2) eşitliğindeki lojistik regresyon modelini veri kümesine uydurmak için, bilinmeyen β_0 ve β_1 parametrelerinin tahmin edilmesi gerekir (Lemeshow ve Hosmer, 2000:7).

Bu çalışmada parametrelerin tahmin edilmesinde kullanılan başlıca üç yöntemden bahsedilecektir. Bunlar; "en çok olabilirlik", "yeniden ağırlıklandırılmış iteratif en küçük kareler" ve "minimum lojit ki-kare" yöntemleridir.

2.3.1. En Çok Olabilirlik Yöntemi

Lojistik regresyon modelinde parametrelerin tahmin edilmesinde en çok kullanılan yöntem "En Çok Olabilirlik Yöntemi" dir. En çok olabilirlik yöntemi, doğrusal regresyon analizindeki en küçük kareler yöntemine benzerlik göstermektedir (Ryan,1997:258).

Doğrusal regresyonda, bilinmeyen parametreleri tahmin etmek için en sık kullanılan yöntem "En Küçük Kareler" yöntemidir. Bu yöntemde modele göre y'nin gözlenen değerleri ile beklenen değerleri arasındaki sapmaların karesinin toplamını en çok küçükleyecek β_0 ve β_1 değerleri seçilir. Doğrusal regresyondaki genel varsayımlar altında, en küçük kareler yöntemi, istenen istatistiksel özelliklere sahip kestiriciler üretir. Ancak iki sonuçlu bir modele en küçük kareler yöntemini uyguladıktan sonra kestiriciler aynı özelliklere sahip olamazlar.

Lojistik regresyon modelde, doğrusal regresyon modelindeki en küçük kareler yönteminin yerine gelen genel tahmin yöntemi "en çok olabilirlik yöntemi" diye adlandırılır. Genel olarak düşünürsek, en çok olabilirlik yöntemi, gözlenen veri kümesine ulaşma olasılığını maksimize edecek şekilde bilinmeyen parametrelerin değerlerinin tahmin edilmesini sağlar. Bu yöntemin uygulanması için ilk olarak "olabilirlik fonksiyonu" diye bir fonksiyon tanımlanması gerekir. Bu fonksiyon, gözlenen verinin olasılığını, bilinmeyen parametrelerin bir fonksiyonu olarak gösterir. Bu parametrelerin "en çok olabilirlik kestiricileri" bu fonksiyonun değerini maksimum yapan değerlerden seçilir. Dolayısıyla, sonuçta oluşan kestiriciler gözlenen veri ile en yakın kestiriciler olurlar (Lemeshow ve Hosmer, 2000:8).

En çok olabilirlik yöntemi, istatistiksel modellerin tüm çeşitleri için yaygın olarak kullanılan genel bir tahmin yöntemidir. Popüler olmasının iki nedeni vardır: Birincisi, en çok olabilirlik kestiricilerinin, büyük örneklemlerde istenen bazı iyi özelliklere sahip olduğu bilinir. Genel olarak bu kestiriciler tutarlı, asimptotik olarak etkin ve normal dağılımlıdır. İkinci neden ise; çözüm için başka olasılıklar bulunmadığında, en çok olabilirlik kestiricilerinin türetilmesi gerekmektedir. En çok olabilirlik yöntemi, kategorik bağımlı değişkenler söz konusu olduğunda iyi sonuçlar vermektedir (Allison, 2001:16-17).

Eğer *y*, 0 ya da 1 olarak gösterilirse, denklem (2.2)'de verilen $\pi(x)$ ifadesi, verilen *x* değeri için *y*'nin 1'e eşit olması koşullu olasılığını verir. Bu da $P(y=1 \mid x)$ olarak gösterilir. ($\beta' = [\beta_0, \beta_1]$ rastgele değeri, parametrelerin vektörüdür.) $1-\pi(x)$ niceliği, verilen *x* için *y*'nin 0'a eşit olması koşullu olasılığını verir ve bu da $P(y=0 \mid x)$ olarak gösterilir. Dolayısıyla (x_i , y_i) ikilileri için $\pi(x_i)$, hesaplanmış x_i için $\pi(x)$ değerini göstermek üzere;

$y_i = 1$ olan ikililerin olabilirlik fonksiyonuna katkısı $\pi(x_i)$ 'dir.

$y_i = 0$ olan ikililerin olabilirlik fonksiyonuna katkısı $1-\pi(x_i)$ 'dir.

Buna göre (x_i , y_i) ikililerinin olabilirlik fonksiyonuna katkılarını göstermek için şöyle bir yol izlenir:

$$\zeta(x_i) = \pi(x_i)^{y_i}\left[1-\pi(x_i)\right]^{1-y_i} \tag{2.8}$$

Gözlemlerin bağımsız olduğu varsayımından dolayı, yukarıdaki (2.8) ifadesinde verilen terimlerin bir çarpımı olarak "olabilirlik fonksiyonu" şu şekilde elde edilir:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} \zeta(x_i) \tag{2.9}$$

En çok olabilirlik ilkesi, (2.9) eşitliğini maksimum yapan **ß** tahmininin kullanılmasını öngörmektedir. Ancak matematiksel olarak, bu eşitliğin logaritması ile çalışmak daha kolaydır. Buna göre "log olabilirlik" kavramı şöyle tanımlanır:

$$L(\boldsymbol{\beta}) = \ln[l(\boldsymbol{\beta})] = \sum_{i=1}^{n} \{y_i ln[\pi(x_i)] + (1-y_i) ln[1-\pi(x_i)]\} \tag{2.10}$$

L(**ß**)'yi maksimum yapan **ß** değerini bulmak için, L(**ß**) değeri, β_0 ve β_1'e göre türevi alınıp 0'a eşitlenir. Böylece "olabilirlik eşitlikleri" diye adlandırılan aşağıdaki eşitlikler elde edilir;

$$\sum_{i=1}^{n}\left[y_i - \pi(x_i)\right] = 0 \qquad (2.11)$$

$$\sum_{i=1}^{n} x_i\left[y_i - \pi(x_i)\right] = 0 \qquad (2.12)$$

Doğrusal regresyonda, karelerin sapmaları toplamının **ß**'ya göre türevi alınarak elde edilen olabilirlik eşitlikleri bilinmeyen parametrelerde doğrusaldırlar ve çözümleri kolaydır. Ancak lojistik regresyon için (2.11) ve (2.12) eşitliklerindeki olabilirlik eşitlikleri, β_0 ve β_1 'de doğrusal değildirler ve dolayısıyla çözümleri için özel yöntemlere ihtiyaç duyulur. Bu yöntemler doğaları gereği iteratif yani tekrarlıdırlar (Lemeshow ve Hosmer, 2000:8-9). Tekrarlanan bu tahmin süreci, test edilmesi ve yeniden tahmin edilmesi iterasyon olarak adlandırılır. Olabilirlik fonksiyonundaki değişim sonraki aşamalarda ihmal edilebilir duruma gelinceye kadar çözüme devam edilir. Olabilirlik fonksiyonunu maksimize edecek en iyi parametre kümesini oluşturmak için yapılan tüm bu işlemler için bilgisayar uygulamalı algoritmalar tasarlanmıştır (Menard, 2002:14).

(2.11) ve (2.12) eşitliklerinden elde edilen **ß** değeri "en çok olabilirlik tahmini" diye adlandırılır ve $\hat{\boldsymbol{\beta}}$ ile gösterilir. Genel olarak "^" sembolü bu niceliğin en çok olabilirlik tahminini gösterir. Örneğin, $\hat{\pi}(x_i)$, $\pi(x_i)$'nin en çok olabilirlik tahminidir. Bu nicelik, $x = x_i$ olarak verildiği zaman, y'nin 1 olma koşullu olasılığı için bir tahmin verir. Bu da lojistik regresyon modelinin uyarlanan ve tahmin edilen değerini gösterir. Bu durumda (2.11) eşitliği aşağıdaki gibi olur;

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n}\hat{\pi}(x_i) \qquad (2.13)$$

Verilen (2.13) eşitliği, (2.11) ifadesinden elde edilen eşitliktir ve y'nin gözlenen değerlerinin toplamının, beklenen (tahmin edilen) değerlerinin toplamına eşit olduğunu gösterir (Hosmer ve Lemeshow, 2000:9-10).

2.3.2. Yeniden Ağırlıklandırılmış İteratif En Küçük Kareler Yöntemi

Doğrusal regresyonda bilinmeyen parametreleri bulmak için sıklıkla kullanılan yöntem en küçük kareler yöntemidir. 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 (Menard, 2002:14).

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

Bu durumda lojit (r_j / n_j)'nin açıklayıcı değişkenler üzerinde $w_j = n_j / P_j\ (1- P_j)$ ağırlığı ile ağırlıklandırılmış regresyonu 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 (Ürük, 2007:11).

2.3.3. Minimum Lojit Ki-Kare Yöntemi

Ağırlıklı en küçük kareler tahmin yönteminin özel bir biçimi olan ve Berkson tarafından geliştirilen bu yöntemde, 2*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.

Tahminde kullanılan ağırlık değerleri $n_j \cdot P_j(1-P_j)$ olarak elde edilmektedir. Bu bilgiler ışığında yöntem, lojit değeri olarak tanımlanan sonuç değişkeninin, açıklayıcı değişkenler ile (tanımlanan ağırlık değerleri 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 (Ürük, 2007:11) .

Kısaca anlatılan bu üç tahmin yöntemi dışında kullanılan başka yöntemler de bulunmaktadır. Ancak çok özel durumlarda kullanılmaları nedeniyle, bu çalışmada değinilmemiştir. Lojistik modellerde parametrelerin tahmin edilmesinden sonra modeldeki katsayıların anlamlılığının ölçülmesi işlemine geçilir.

2.4. Modeldeki Katsayıların Anlamlılık Testi ve Yorumlanması

Parametrelerin tahmininden sonra uyarlanan modelde atılması gereken ilk adım, modeldeki değişkenlerin anlamlılığının ölçülmesidir. Bu işlem; modeldeki bağımsız değişkenlerin sonuç değişkenleri ile ilişkisinin ne kadar anlamlı olup olmadığını belirlemek için formül üretimi ve istatistiksel bir hipotezin test edilmesini kapsamaktadır.

Herhangi bir modeldeki değişkenin katsayısının anlamlılığını test etmek için gereken yaklaşım şu soru ile ilgilidir: “Söz konusu değişkeni içeren model, bu değişkeni içermeyen modele göre sonuç (cevap) değişkeni hakkında daha çok bilgi verebiliyor mu?” Bu soru sonuç değişkeninin gözlenen değerlerinin, söz konusu değişkeni içeren ve içermeyen iki modelden elde edilen tahminler ile karşılaştırılmasıyla cevaplanır. Eğer değişkenin modelde olduğu tahmini değerler, o değişkene sahip olmayan modele göre daha iyi ve açık ise, söz konusu değişken “anlamlı” dır.

Doğrusal regresyonda eğim katsayısının anlamlılığının ölçülmesi, ilk olarak “Varyans Tablosu Analizi”nin açıklanmasıyla yapılır. Bu tablo, gözlemlerin ortalamalarından sapmalarının karelerinin toplamını (total sum of squares, *SST*) iki parçaya ayırır:

1) *SSE* (error sum of squares): gözlemlerin regresyon çizgisinden yani tahmin edilen değerlerden sapmalarının kareleri toplamı

2) *SSR* (regression sum of squares): regresyon modelindeki tahmin edilen değerlerin, bağımlı değişkenin ortalamasından sapmalarının kareleri toplamı

Bu yöntem iki modeldeki gözlenen ve tahmin edilen değerlerin karşılaştırılmasını göstermektedir. Doğrusal regresyonda gözlenen ve tahmin edilen değerlerin karşılaştırılması, ikisi arasındaki uzaklığın karesini baz alır. Modele göre *i*’nci birey için, y_i gözlenen değeri ve $\hat{y}_i$ tahmin edilen değeri göstermek üzere, bu karşılaştırma için kullanılan istatistik şudur:

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2.14}$$

Söz konusu bağımsız değişkene sahip olmayan model için tek parametre β_0’dır ve $\hat{\beta}_0 = \bar{y}$, yani cevap değişkeninin ortalamasıdır. Bu durumda $\hat{y}_i = \bar{y}$ olur ve *SSE* toplam varyansa eşit olur. Modele bağımsız değişken eklendiğinde, *SSE*’deki her düşüş bağımsız değişken için eğim katsayısının, 0 (sıfır) olmamasından kaynaklanacaktır. *SSE* değerindeki değişim ise, değişkenliğin regresyonundan kaynaklanmaktadır ve bu da *SSR* ile gösterilmektedir. (*SSR* = *SST* – *SSE*). Şöyle ki: (Lemeshow ve Hosmer, 2000:11-12)

$$SSR = \left[\sum_{i=1}^{n} (y_i - \bar{y})^2\right] - \left[\sum_{i=1}^{n} (y_i - \hat{y}_i)^2\right] \tag{2.15}$$

Değişkenlerin anlamlılıklarının ölçülmesi için uygulanan testlerden birisi olabilirlik oranı testidir. Bu test *SSR* değerini, değişkenin yer aldığı ve almadığı iki ayrı modelde kıyaslayarak modeldeki bağımsız değişkenlerin anlamlı olup olmadığı üzerinde durur.

2.4.1. Olabilirlik Oranı Testi

Doğrusal regresyonda anlamlılık ölçülürken *SSR* değerinin büyüklüğü üzerinde durulur. Yani; büyük değerler, bağımsız değişkenin önemli olduğunu, küçük değerler ise bu bağımsız değişkenin cevabı tahmin etmek için önemli olmadığını gösterir. Lojistik regresyonda da temel prensip aynıdır. Bağımsız değişkenin dahil olduğu ve olmadığı modellerdeki gözlenen değerler, tahmin edilen değerlerle karşılaştırılır. Bu karşılaştırmanın yapılması log-olabilirlik fonksiyonuna dayanmaktadır ve bunun için aşağıda (2.16) eşitliğinde verilen olabilirlik fonksiyonu tanımlanır:

$$D = -2ln\left[\frac{(Geçerli\ modelin\ olabilirliği)}{(Doymuş\ modelin\ olabilirliği)}\right] \tag{2.16}$$

Doymuş bir model, değişken sayısı kadar parametre içeren modeldir. Yukarıdaki (2.16) eşitliğinde parantez içinde yer alan ifade, olabilirlik oranı (likelihood ratio) olarak adlandırılır. Eşitlikteki “-2*ln*” değeri matematikseldir ve hipotez testinde kullanılmak üzere dağılımı bilinen bir nicelik elde edilmesi için gereklidir. *D* (deviance) istatistiği, doğrusal regresyondaki hata kareleri toplamı ile aynı rolü oynar, bu test olabilirlik oranı testi olarak adlandırılır ve test için kullanılan istatistik (2.17) eşitliğindeki gibidir:

$$D = -2\sum_{i=1}^{n}\left[y_i ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1-y_i)ln\left(\frac{1-\hat{\pi}_i}{1-y_i}\right)\right] \tag{2.17}$$

Bağımsız değişkenin anlamlılığını değerlendirmek için; bağımsız değişkenin dahil olduğu ve olmadığı eşitliklerin *D* değerleri birbirleriyle karşılaştırılır. Bağımsız değişkenin modelde yer almasından kaynaklanan, *D*'deki değişim aşağıdaki şekilde elde edilir. (Burada, $\beta_1=0$ hipotezi altında *G* istatistiğinin 1 serbestlik derecesi ile Ki-kare (χ^2) dağılımına uyduğu ve yeterli bir *n* örneklem büyüklüğüne sahip olduğu varsayılmaktadır).

$$G = D\ (\text{Değişkeni içermeyen model için}) - D\ (\text{Değişkeni içeren model için}) \qquad (2.18)$$

$$G = -2ln\left[\frac{(\text{Değişkeni içermeyen modelin olabilirliği})}{(\text{Değişkeni içeren modelin olabilirliği})}\right] \qquad (2.19)$$

Değişkenlerin anlamlılık testi için birbirine istatistiksel olarak denk olan iki test daha önerilmektedir. Bunlar Wald ve Score testlerdir. Olabilirlik oranı testi için geçerli varsayımlar bu testler için de geçerli sayılmaktadır (Lemeshow ve Hosmer, 2000:12-16).

2.4.2. Wald Testi

Wald testi, $\hat{\beta}_1$ parametresinin en çok olabilirlik tahminlerinin standart hata tahminlerine oranlanmasıyla uygulanır. $\beta_1 = 0$ hipotezi altında ortaya çıkan oran, standart normal dağılıma sahiptir (Lemeshow ve Hosmer, 2000:16).

$$W = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \qquad (2.20)$$

Doğrusal regresyonda, diğer parametreler modelde bulunurken tek parametrenin test edildiği durumlarda *t* istatistiği kullanılmaktadır. Lojistik regresyonda ise Wald testi buna karşılık gelmektedir. Ancak, Wald testi *t* istatistiği ile aynı formda olmasına rağmen *t* dağılımına uymamakta ve normal dağılım göstermektedir (Ryan, 1997:269).

Normal dağılıma uyan yani Z dağılımı gösteren Wald testinin karesi alınarak da test uygulanabilir. Z istatistiğinin karesi, 1 serbestlik derecesi ile ki-kare dağılımına uymaktadır. Wald testi, her değişken için en çok olabilirlik katsayılarını ve bunun standart hatalarını hesaplar. Çoğu bilgisayar yazılımı ayrıca ki-kare istatistiğini ve p değerini hesaplayarak sonuç çıktısında gösterir. Büyük örneklemlerde Wald testi ve Olabilirlik oranı testinin çok yakın sonuçlar verdiği fakat örneklem küçük olduğunda bu iki istatistiğin farklı sonuçlar verdiği görülmüştür. Böyle bir durumda olabilirlik oranı testinin daha iyi sonuçlar verdiği istatistikçiler tarafından ortaya konulmuş, bu nedenle de, bunun gibi bir şüphe olduğunda olabilirlik oranı testinin kullanılması önerilmiştir (Kleinbaum, 2002:135).

Olabilirlik oranı testi iki modelin tahmin edilmesini gerektirirken, testin hesaplanması sadece çıkarma işlemlerinden oluşmaktadır. Wald testi ise tek bir modelin tahmin edilmesini gerektirir, fakat test matris çarpımlarının hesaplanmasını kapsamaktadır (Long, 1997:97).

Olabilirlik oranı testi (G) ve Wald (W) testinin her ikisi de β_1 parametresinin en çok olabilirlik tahminlerinin hesaplanmasına ihtiyaç duyar. Score testi ise bu hesaplamayı gerektirmeyen bir yöntemdir. Ancak testin bazı bilgisayar programlarında bulunmayışı kullanılmasını kısıtlamaktadır (Lemeshow ve Hosmer, 2000:16).

2.4.3. Score Testi

Score testi, log-olabilirliklerin türevlerinin dağılım teorisine dayanmaktadır. Genel olarak, matris hesaplamalarını gerektiren çok değişkenli bir testtir. Score testi için test istatistiği ST aşağıdaki gibidir:

$$ST = \sum_{i=1}^{n} \frac{x_i(y_i - \bar{y})}{\sqrt{\bar{y}(1-\bar{y})\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{2.21}$$

Özet olarak lojistik regresyonda bir değişkenin katsayısının önemini test etmek için kullanılan yöntem, doğrusal regresyondaki kullanılan benzer yaklaşımlardan ortaya çıkar. Ancak ikili sonuç değişkeni için önem testi uygulanmak istediğinde lojistik regresyonda olabilirlik fonksiyonu kullanılır (Lemeshow ve Hosmer, 2000:16-17).

Lojistik regresyon analizinde model oluşturulduktan sonra yapılan parametre tahminleri ve değişkenlerin anlamlılık testleri adımlarını son olarak modelin uyum iyiliğinin ölçülmesi izler. Yani oluşturulan modelin, sonuç değişkenini açıklamaya ne kadar yardımcı olduğu değerlendirilmelidir.

2.5. Modelin Uyum İyiliğinin Değerlendirilmesi

Modelde bulunması gereken değişkenler modele doğru bir şekilde alındıktan sonra, modelin sonuç değişkenini açıklamakta ne kadar etkin olduğunun bilinmesi gerekmektedir. Yapılan bu araştırmaya "modelin uyum iyiliği" denir.

Sonuç değişkeninin gözlenen değerlerinin y ile, modelden tahmin edilen değerlerinin ise $\hat{y}$ ile gösterildiği varsayılsın. Bu durumda y ve $\hat{y}$ arasındaki uzaklığın özet istatistiklerinin küçük olması, her $(y_i , \hat{y}_i)$ ikilisinin bu özet istatistiklere katkısının sistematik olmaması ve modelin hata yapısı ile düşük ilişkili olması halinde modelin uyumlu olduğu sonucuna varılır. Yani uyarlanan modelin değerlendirilmesi, hem y ve $\hat{y}$ arasındaki uzaklığın özet istatistiklerinin hesaplanmasını hem de bu istatistiklerin bileşenlerinin eksiksiz şekilde incelenmesini içermektedir. Uyum iyiliğinin değerlendirilmesi için gereken yaklaşımda izlenen adımlar şöyledir:

(1) Uyumluluğun genel ölçütlerinin hesaplanması ve değerlendirilmesi,

(2) Bu özet istatistikleri oluşturan öğelerin incelenmesi,

(3) y ve $\hat{y}$ arasındaki farkın ya da uzaklığın diğer ölçütlerinin incelenmesidir (Lemeshow ve Hosmer, 2000:143).

Lojistik regresyon analizinde, logit modeldeki katsayılar tahmin edildikten sonra, genel olarak modelin güvenilirliğinin test edilmesi gerekir. Bu amaçla bağımsız değişkenler ile kategorik bağımlı değişken arasında önemli derecede ilişki olup olmadığının hipotezleri, yani teorik modelin verileri iyi temsil edip etmediğinin hipotezleri oluşturularak test edilmelidir (Ege ve Bayraktaroğlu, 2009:147). Bunu değerlendirmek için alt bölümde bahsedilen bazı istatistiklerden faydalanılmaktadır.

2.5.1. Pearson Ki-kare (χ^2) ve Deviance İstatistiği

Doğrusal regresyonda model uyumunun özet istatistikleri, gözlenen ve tahmin edilen değerler arasındaki sapmaların (y - $\hat{y}$) bir fonksiyonudur. Lojistik regresyonda da gözlenen ve tahmin edilen değerler arasındaki farkı ölçmek için birçok olası istatistik bulunmaktadır. Bu istatistiklerden bahsedilmesi gereken ikisi Pearson ve Deviance istatistikleridir. (y - $\hat{y}$) sapmalarına dayanarak elde edilen Pearson Ki-kare istatistiği (χ^2) ve Deviance (D) aşağıdaki gibi ifade edilir: (Lemeshow ve Hosmer, 2000:145-146)

$$r(y_i, \pi_i) = \frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i(1-\hat{\pi}_i)} \tag{2.22}$$

$$\chi^2 = \sum_{i=1}^{n} r(y_i, \hat{\pi}_i)^2 \tag{2.23}$$

$$D = \sum_{i=1}^{n} d(y_i, \hat{\pi}_i)^2 \tag{2.24}$$

D istatistiği daha önce gösterilen (2.17) eşitliğindeki olabilirlik oranı test istatistiğinin doymuş modelde uygulanması ile elde edilmektedir ve (2.24) eşitliğinde gösterilmektedir. (2.22) ve (2.23) eşitliğinde gösterilen χ^2 istatistiği için de aynı yaklaşım geçerlidir.

Deviance ve χ^2 istatistiklerinin her ikisi de ki-kare dağılımı göstermektedir. Bu iki istatistiğin sayısal değeri genellikle birbirinden farklılaşmaktadır fakat bu farkın pratikte pek bir önemi olmayacaktır. İki istatistik arasında büyük farklar olması, D ya da χ^2 istatistiğinin ki-kare dağılımına uygun olmadığı anlamına gelmektedir (Collett, 2002:87).

Doğrusal regresyon modelindeki F ve R^2 istatistiklerine paralel olarak lojistik regresyon modelinde, uyum iyiliğini ölçmek için G_M (Model Ki-kare) ve R_L^2 istatistikleri kullanılır. Doğrusal regresyon modelinde parametre seçme kriteri olan hata kareleri toplamına karşılık olarak lojistik regresyon modelinde log-olabilirlik parametre seçme kriteri kullanılır. İstatistiksel programlarda genelde log-olabilirlik'in -2 ile çarpılmış hali verilir ve -2LL (-2Log-Likelihood) olarak gösterilir. Log-olabilirlik negatif olduğunda, -2LL pozitiftir ve büyük değerler bağımlı değişkenin kötü tahmin edildiğini göstermektedir. Bu haliyle, lojistik regresyonda kullanılan -2LL değeri ki-kare dağılımına uygun olarak test etmeyi sağlamaktadır. (Menard, 2002:20-21).

2.5.2. Model Ki-Kare İstatistiği

Lojistik regresyon analizinde model uyumunu test ederken D_0 olarak gösterilen bağımsız değişkenleri içermeyen modelin -2LL istatistiği, doğrusal regresyon analizindeki bağımlı değişkenlerin ortalamadan sapma kareleri toplamına (total sum of squares, *SST*) karşılık gelmektedir. 0 ve 1 olarak kodlanan ikili bağımlı değişken için eğer $n_{y=1}$ ise, y=1 olduğunda n durum sayısını, N toplam durum sayısını ve $P(y=1) = n_{y=1} / N$ ifadesi de y'nin 1'e eşit olma olasılığını göstermek üzere;

$$D_0 = -2\{n_{y=1}\, ln[P(y=1)] + (N - n_{y=1}) ln[1 - P(y=1)]\} \qquad (2.25)$$

$$= -2\{n_{y=1}\, ln[P(y=1)] + (n_{y=0}) ln[P(y=0)]\}$$

Lojistik regresyon modelinde bağımsız değişkeni içeren modelin -2LL istatistiği D_M ile gösterilir. D_M, doğrusal regresyon analizindeki tahmin edilen ve gözlenen değerler arasındaki hata kareleri toplamına (error sum of squares, *SSE*) karşılık gelmektedir. Doğrusal regresyon analizinde bu sapmaların kareleri toplamı arasındaki farka (residual sum of squares, *SSR*) bakıldığı gibi lojistik regresyon analizinde D_M ve D_0 arasındaki farka bakılmaktadır. Aradaki bu fark G_M ya da *Model* χ^2 olarak ifade edilmektedir.

$$SST - SSE = SSR$$

$$D_M - D_0 = G_M \text{ (Model } \chi^2\text{)}$$

Lojistik regresyon modeli için G_M, boş hipotezi yani sabit katsayı β_0 hariç diğer katsayıların sıfıra eşit olup olmadığını test etmektedir. Boş hipotez (H_0) ve boş hipotezin tersini, yani katsayıların en az birinin sıfıra eşit olmadığını savunan alternatif hipotez (H_1) aşağıdaki gibi gösterilmektedir:

$$H_0 = \beta_1,\ \beta_2, \ldots, \beta_k = 0 \tag{2.26}$$

$$H_1 = \beta_1,\ \beta_2, \ldots, \beta_k \neq 0 \tag{2.27}$$

Eğer G_M istatistiksel olarak anlamlı ise yani ($p \leq 0{,}05$) ise H_0 reddedilir. İdeal olarak G_M' nin istatistiksel olarak anlamlı olduğu ancak D_M' nin anlamlı olmadığı bir modele ulaşmak istenir. (Menard, 2002:21-22).

2.5.3. R_L^2 (Pseudo-R^2) İstatistiği, Cox-Snell R^2 ve Nagelkerke R^2 İstatistikleri

Doğrusal regresyon analizinde determinasyon katsayısı olarak bilinen R^2 değeri modelin açıklayıcı gücünü yani bağımsız değişkenlerin bağımlı değişkenler üzerindeki etkisini ölçmek için kullanılmaktadır. Bu istatistik, model tarafından açıklanan cevap değişkenlerinin toplam varyanslarının oranlanmasıyla ölçülmektedir. R^2 istatistiği, ayrıca bağımlı değişkenin gözlenen ve tahmin edilen değerleri arasındaki korelasyonun karesinin alınması ile de hesaplanabilir. 0 ile 1 arasında bir değer alan ve değer ne kadar büyükse modelin o kadar uyumlu olduğunu ifade eden R^2 şu şekilde hesaplanmaktadır: (Collett, 2002:89-90).

$$R^2 = 1 - \frac{SSR}{SST} \tag{2.28}$$

Doğrusal regresyon analizindeki hata karelerinin oranlamasıyla elde edilen değerden hareketle lojistik regresyon analizinde -2LL istatistiğine paralel bir dönüşümle R_L^2 istatistiği elde edilmektedir. R_L^2, modeldeki bağımsız değişkenlerin D_0 tarafından ölçülen varyasyonun azalmasında ne kadar etkili olduğunu ölçmektedir. Bu varyasyon 0 ile 1 arasında değer almaktadır. Değerin 0 olması; $G_M=0$, $D_M=D_0$ ve modeldeki bağımsız değişkenlerin bağımlı değişkenleri tahmin etmekte elverişsiz olduğu anlamına gelmektedir. Değerin 1 olması; $G_M=D_0$, $D_M=0$ ve modelin bağımlı değişkenleri tahmin etmekte mükemmel olduğunu göstermektedir. R_L^2 değeri bilgisayar programlarının sonuç çıktılarında *McFadden* R^2 olarak *Pseudo* R^2 tablolarından da elde edilebilir. L_0, geçerli modelin olabilirlik fonksiyonunu L_M ise tüm kestiricileri içeren doymuş modelin olabilirlik fonksiyonunu ifade etmek üzere R_L^2 istatistiği (2.29) eşitliğindeki gibi hesaplanabilir: (Menard, 2002:24-25)

$$R_L^2 = 1 - \left(\frac{L_0}{L_M} \right)^{2/n} \tag{2.29}$$

R_L^2 istatistiğine alternatif olarak bilgisayar programlarında model özet tablosunda ya da Pseudo R^2 tablosunda, *Cox-Snell* R^2 ve *Nagelkerke* R^2 istatistikleri de yer almaktadır. R_N^2 olarak ifade edilen bu istatistik ise, R_L^2 değerinin bu değeri maksimum yapan değere bölünmesiyle elde edilir. Şöyle ki: (Menard, 2002:25)

$$R_N^2 = \frac{1 - \left(\frac{L_0}{L_M} \right)^{2/n}}{1 - (L_0)^{2/n}} = \frac{R_L^2}{R_{\max}^2} \tag{2.30}$$

Cox-Snell R^2 ve Nagelkerke R^2 istatistikleri SPSS programında model özet tablolarında -2LL istatistiği ile beraber bulunur ve modelin uyum iyiliğini gösterirler. Diğer bir deyişle, bağımsız değişkenlerin, bağımlı değişkeni açıklamakta ne kadar iyi olduğunu gösterirler. Değer ne kadar büyükse değişkenler modeli açıklamakta o kadar başarılıdırlar.

Lojistik regresyon analizinde yer alan başlıca istatistikler yukarıda bahsedildiği gibidir. Çalışmanın bu bölümünde anlatılan lojistik regresyon analizinden sonra çalışmaya konu olan tahmin yöntemlerinin uygulanmasına yönelik örneğin verildiği üçüncü bölüme geçilecektir. Uygulama kısmında çalışmanın birinci bölümünde anlatılan bulanık regresyon analizi ve ikinci bölümde yer alan lojistik regresyonla ilgili literatürde yer alan çalışmalar verilecek, daha sonra bu çalışmanın uygulamasına konu olan tahmin modeli çözülecektir.

ÜÇÜNCÜ BÖLÜM

3. SEKTÖR PAYLARININ TAHMİNLENMESİNDE BULANIK DOĞRUSAL REGRESYON ANALİZİ VE LOJİSTİK REGRESYON ANALİZİNİN UYGULANMASI

Bu çalışmaya konu olan istatistiksel tahmin yöntemleri de işletmelerin tahminleme sürecinde tüm tahmin yöntemleri gibi uygulamalardaki kullanımları ile önem kazanmaktadır. Çalışmanın bu bölümünde, önceki bölümlerde anlatılan iki istatistiksel tahmin yöntemi Bulanık Doğrusal Regresyon Analizi ve Lojistik Regresyon Analizinin uygulanmasına ilişkin bankacılık sektör paylarının tahminlenmesine yönelik örnek bir uygulama yapılmıştır ve elde edilen sonuçlar karşılaştırılmıştır. Analizde, Türk Bankacılık sektöründe faaliyet gösteren 45 bankaya ait veriler kullanılmıştır. Literatürde her iki yöntemi birlikte ele alan bir çalışmaya rastlanamamıştır, ancak yöntemleri ayrı ayrı ele alan ve farklı birçok alanda uygulama yapılan çalışmalar mevcuttur. Uygulama örneğine geçmeden önce çalışmanın birinci bölümünde anlatılan Bulanık doğrusal Regresyon Analizi, ikinci bölümünde anlatılan Lojistik Regresyon Analizi ile ilgili literatürde yer alan çalışmalara yer verilmiştir.

3.1. Bulanık Doğrusal Regresyon Analizi ve Lojistik Regresyon Analizine İlişkin Literatür Taraması

Çalışmaya konu alan iki tahmin yöntemine ilişkin literatür taramasında, ilk olarak birinci bölümde anlatılan Bulanık Doğrusal Regresyon Analizi ile ilgili tahminlemeye yönelik yapılan daha önceki çalışmalar özetlenmiştir.

Tanaka, Uejima ve Asai, klasik regresyon prensiplerini esneterek bulanık modele sahip doğrusal regresyon analizini ilk kez ortaya koyanlar olmuşlardır. Girdi ve çıktı değişkenlerinin bulanık olmadığı, fakat sistem bilgisinin bulanık olduğu varsayılmakta ve amaç fonksiyonu bağımlı değişkenin tahmin değerinin yayılımının minimizasyonuna dayanmaktadır. Konut fiyatlarının tahminine yönelik yapılan uygulama doğrusal programlama tekniği kullanılarak çözümlenmekte ve sonuçta bulanık regresyon modelinin sistemin belirsizliğini değerlemede ve tahmin etmede kullanışlı olduğu kanısına varılmaktadır (Tanaka ve diğer., 1982).

Moskowitz ve Kim bulanık doğrusal regresyonda bulanık parametrelerin yayılmaları, üyelik fonksiyonları şekilleri ve bulanıklık seviyesini ifade eden *h* değeri arasındaki ilişkiyi belirlemişlerdir (Moskowitz ve Kim, 1993).

Ming, Friedman ve Kandel, çalışmalarında bulanık verilerle oluşturulan bir en küçük kareler yöntemi tanımlamışlardır ve oluşturdukları yeni model daha öncekilerle karşılaştırılarak uygulanabilirliği gösterilmiştir (Ming, Friedman ve Kandel, 1997).

Diamond ve Körner, bulanık doğrusal modelin en küçük kareler yöntemini, negatif yayılımları engellemek ve yorumlamak için genişleterek uygulamışlardır. Uyarlanan modeller klasik regresyondaki determinasyon katsayısı kullanılarak karşılaştırılmıştır (Diamond ve Körner, 1997).

Kim ve Chen, klasik regresyon analizine alternatif olarak sunulan bulanık doğrusal regresyon ve nonparametrik doğrusal regresyon analizlerini ele alarak, kesin olmayan verilerle çalışmada hangi yöntemin daha iyi bir seçim olacağı konusunda karşılaştırma yapmışlardır (Kim ve Chen, 1997).

Kim ve Bishu, gözlenen ve tahmin edilen bulanık değerler arasındaki farkları minimize etme kısıtını karşılamaya dayalı olarak bulanık regresyon analizinde bazı düzenlemeler yaparak karar vericinin tahmin sürecinde alacağı riskleri de azaltmayı amaçlamışlardır ve bunun için iki sayısal örnekle bulanık regresyon modellerini değerlendirmişlerdir (Kim ve Bishu, 1998).

Wang ve Tsaur, Bulanık regresyon modelinde bulanık aralıklar üzerinde durarak regresyon aralık analizinden bahsetmişlerdir, bulanık regresyon aralıklarının analizine ilişkin teoremler ortaya koymuş ve bunların açıklamasını yaparak bulanıklık seviyesi h ve değişken seçimini açıklamaya yönelik sayısal örnekler vermişlerdir (Wang ve Tsaur, 2000a).

Wang ve Tsaur, Tanaka tarafından tanımlanmış kesin (bulanık olmayan) girdi ve bulanık çıktıya sahip problemin çözümü için geliştirilmiş bulanık en küçük kareler yöntemini önermişlerdir, bu yöntemi Tanaka'nın yöntemi ile karşılaştırmışlardır (Wang ve Tsaur, 2000b).

D'Urso ve Gastaldi, bulanık regresyon analizine yeni bir yaklaşım getirerek iki doğrusal modele dayanan 'çift doğrusal uyarlamalı model'i önermişlerdir. Bu modeller, çekirdek regresyon model ve yayılım regresyon modelidir. Birincisi bulanık gözlemlerin merkez değerlerini ikincisi ise yayılımlarını ifade etmektedir. Çoğu gerçek uygulamada karşılaşılan bu merkezler ve yayılımlar arasındaki bağımlılık, bu çalışmada aralarında doğrusal ilişkiler kurularak incelenmiş ve buna yönelik örneklerin sonuçları verilmiştir (D'Urso ve Gastaldi, 2000).

Chang ve Ayyub, bulanık regresyon ve klasik regresyon arasındaki farklılıkları tanımlamışlardır. Çalışmada bulanık regresyonun üç yaklaşımı özetlenmiştir. İlk yaklaşım, en uygun ölçüt ile bulanıklığın minimizasyonu esasına dayanmaktadır. İkinci yaklaşımda uygun ölçüt olarak hataların en küçük kareleri kullanılmaktadır ve çalışmada buna dair iki yöntem özetlenmiştir. Üçüncü yaklaşım aralık regresyon analizi olarak tanımlanmaktadır. Her bir bulanık regresyon yöntemi ile klasik regresyon yöntemi arasındaki farklılığı değerlendirmek için sayısal örnekler ve grafiksel sunumlar kullanılmıştır. Çalışmada klasik regresyon model verilerindeki belirsizliğin rastgelelik tipi ile geleneksel bulanık regresyon model verilerindeki belirsizliğin bulanıklık tipi arasındaki temel farklılıklar karşılaştırmalı olarak değerlendirilmiştir. Bu bulanıklık ve rastgeleliği tek bir regresyon modelinde bütünleştirmek amacıyla, bir melez bulanık regresyon analizi ortaya konulmuştur (Chang ve Ayyub, 2001).

Kao ve Chyu, bulanık regresyon analizinde bağımsız değişkenlerin büyüklüğü arttıkça tahmin edilen değişkenlerin yayılımlarının da artması sorunundan yola çıkarak iki aşamalı bir model geliştirmişlerdir. Birinci aşama, verilerin genel eğilimini gösteren kesin bir regresyon doğrusu elde etmek amacıyla klasik en küçük kareler yönteminin uygulanabilmesi için bulanık gözlemlerin bulanıklığının giderilmesini içermektedir. İkinci aşamada, bulanık regresyon modelindeki verilerin bulanıklığını gösteren hata terimi, modele en iyi açıklayıcı gücü kazandırmak için belirlenmektedir. Çalışmada gösterilen iki sayısal örnekte, geliştirilen iki aşamalı yöntemin daha önceki çalışmalara göre daha iyi performans sergilediği sonucuna varılmıştır (Kao ve Chyu, 2002).

Kao ve Chyu, bulanık regresyon analizindeki en küçük kareler yöntemi ile ilgili bir çalışmada bulunmuşlardır. Hata kereleri toplamının üyelik fonksiyonları oluşturulmuş ve regresyon katsayılarının fonksiyonu olarak belirlenmişlerdir. Modelin uygulanışı kesin girdi-bulanık çıktı, bulanık girdi-bulanık çıktı ve üçgensel olmayan bulanık gözlemler olarak üç durumda incelenmiştir. Elde edilen sonuçlar en küçük kareler yönteminin regresyon katsayılarını açıklamakta, üçgensel olmayan durum da dahil olmak üzere çoğu durum için güçlü olduğunu ortaya koymuştur (Kao ve Chyu, 2003).

M. Nasrabadi ve E. Nasrabadi, bulanık-kesin çıktı ve bulanık-kesin girdi ile modellenen bulanık regresyon üzerinde durmuşlar ve modelin çözümünde matematiksel programlamaya dayalı bir yaklaşım öne sürmüşlerdir. Bu yaklaşımın programlama ve hesaplamadaki avantajlarını değerlendirerek, gözlenen ile beklenen değerler arasındaki toplam yayılımın en aza indirgendiğini savunmuşlardır (M. Nasrabadi ve E. Nasrabadi, 2004)

Sanchez ve Gomez, faiz oranlarındaki belirsizliğin bulanık sayılar ile tanımlanmasından yola çıkarak bu oranların bulanık sayılarla nasıl tahmin edileceğine ilişkin sorunu ele almışlardır ve faiz oranlarındaki zamansal yapının ayarlanması için bulanık regresyon tekniklerine dayalı bir yöntemle çözüm sunmuşlardır. Bu yöntem gelecek için bulanık sayılarla öngörülen oranların hesaplanmasının sağlamaktadır (Sanchez ve Gomez, 2004).

Hojati, Bector ve Smimou sadece bağımsız değişkenlerin bulanık olduğu ve hem bağımlı hem bağımsız değişkenlerin bulanık olduğu iki durum için bulanık regresyon hesaplamasına yönelik basit ve iyi çözüm sunan, doğrusal programlama temeline dayanan yeni bir yöntem önermişlerdir. Önerilen yaklaşım ile literatürdeki diğer yaklaşımları karşılaştırmışlardır (Hojati, Bector ve Smimou, 2005).

Modarres, E. Nasrabadi ve M. Nasrabadi, kesin girdi ve bulanık çıktıya sahip bulanık regresyon modellerindeki parametreleri tahmin etmek için bir matematiksel programlama yöntemi geliştirmişlerdir. Yöntem, gözlenen ve tahmin edilen yayılım değerleri arasındaki toplam farkın karesinin minimize edilmesine diğer bir deyişle hataların karelerinin minimize edilmesine dayanmaktadır. Çalışmada iki örnek sunulmuş ve önerilen yöntem diğer yöntemlerle karşılaştırılarak hesaplamada kolaylık sağladığı sonucuna varılmıştır (Modarres ve diğer,2005).

Yücel, yaptığı tez çalışmasında bulanık regresyon yöntemini kullanarak, Türkiye’de 1980-2004 döneminde kayıt dışı ekonominin tahminine yönelik bir uygulamada bulunmuştur. Kayıt dışı ekonomi modeli kurulurken Tanzi’nin nakit para denklemi esas alınmıştır ve bağımlı değişken olarak nakit para oranının seyrettiği aralık bulanık olarak tahmin edilmeye çalışılmıştır, tahmin edilen aralıklar her yıl için nakit paranın dar anlamlı paraya oranı için, alt ve üst sınır belirlenmesini sağlamıştır. Elde edilen bulgular kayıt dışı ekonominin büyüklüğünü ölçebilecek nitelikte olmasa da nakit para oranının artış ve azalışlarına bakılarak, oranın büyüdüğü dönemlerde kayıt dışı ekonominin de büyüdüğü düşünülmektedir (Yücel,2005).

Stahl, çalışmasında bulanık bağımsız değişken ve bulanık parametreler kullanarak doğrusal bulanık regresyon modeli kurmuştur. Bulanık parametreleri oluşturmak için en küçük kareler metodu kullanılmıştır. En küçük kareler metodunun güçlü bir tahmin edici olduğunu açıklamıştır (Stahl, 2006).

Sanchez, bulanık regresyonu kullanarak sigorta şirketleri için hak korumaya yönelik bir uygulama yapmışlardır. Analizde Isubichi ve Nii tarafından genişletilen Tanaka'nın bulanık regresyon yöntemi ile Sherman'ın hak koruma planı birleştirilerek bir yöntem sunulmuştur (Sanchez, 2006).

He, Chan ve Wu, üretkenlik, müşteri memnuniyeti ve karlılık arasındaki ilişkiyi klasik regresyon ve yeni bulanık regresyon yaklaşımı kullanılarak belirlenmeye çalışmışlardır. Hong Kong'dan 22 adet örnek firma seçilmiş ve değişkenlere ait değerler elde edilmiştir. Klasik en küçük kareler yöntemi kullanılarak model tahmin edilmiştir, fakat bu yöntemle üretkenlik, karlılık ve müşteri memnuniyeti gibi kesin olmayan veriler için regresyon katsayılarının tahmin edilmesi bazı kısıtlar doğurmuştur. Bu nedenle mevcut bulanık doğrusal regresyon yönteminde bu kısıtlar dikkate alınarak 'geliştirilmiş bulanık doğrusal regresyon' yöntemi oluşturulmuştur. Ortaya çıkan kısıtlar ve bulanıklık, geliştirilen model ile çözülmüş ve sonuçlar karşılaştırılmıştır (He, Chan ve Wu, 2007).

Bozdağ ve Türe, bulanık doğrusal programlama modelini portföy seçimi problemine yönelik bir uygulamada kullanarak, belirsiz durumların ve portföy yöneticilerinin deneyimlerinin de karar sürecine katılmasını amaçlamışlardır. Çalışmada İstanbul menkul kıymetler borsasında işlem gören 26 hisse senedine ilişkin Ocak 2003 - Eylül 2005 dönemine ait veriler kullanılmıştır. Optimum portföy yatırımcının farklı risk davranışlarına göre tanımlanmış 6 farklı senaryoya göre belirlenmiştir ve elde edilen sonuçlar, riskten kaçınma ile getiri düzeyi arasında açık bir ilişki olduğuna işaret etmiştir (Bozdağ ve Türe, 2008).

Lu ve Wang, tahmin edilen bağımlı değişkenin yayılımlarının gözlenen bağımlı değişkenin yayılımlarına uyduğu bir geliştirilmiş bulanık doğrusal regresyon modeli önermişlerdir ve modelin etkinliğini göstermek için dört adet sayısal örnek kullanmışlardır (Lu ve Wang, 2009).

Stölzle, Koissi ve Shapiro, regresyon katsayılarının bulanıklığını test etmek için Tanaka (1982) ve He (2007)'nin bulanık regresyon modeline dayanan bir test geliştirmişlerdir. Regresyon katsayılarındaki yayılımı, bağımlı ve bağımsız değişkenler arasındaki ilişkinin bulanıklığının istatistiksel ölçüsü olarak yorumlamışlardır. Geliştirdikleri testin bulanık regresyon katsayılarını nasıl belirlediği, Alman emlak ve mali sigorta şirketlerinin ödeme gücünü tahminlemeye yönelik örnek bir modelle gösterilmiştir (Stölzle, Koissi ve Shapiro, 2010).

Çalışmanın ikinci bölümünde yer alan, önceleri biyolojik alanlarda kullanılırken son yıllarda sosyal bilimlerdeki uygulamalarına da sıkça rastlanılan Lojistik Regresyon Analizi ile ilgili yapılan daha önceki çalışmalar ise aşağıdaki gibi özetlenmiştir.

Lojistik modelin biyolojik deneylerin analizi için kullanımı ilk olarak Berkson (1944) tarafından önerilmiş, Cox (1970) bu modeli gözden geçirerek çeşitli uygulamalarını yapmış, özet gelişmeler ise ilk Andersson (1979, 1983) tarafından verilmiştir. Ayrıca verilerin lojistik modele uyumu ile ilgili birçok çalışmalar da yapılmıştır. Bunlar arasında Aranda-Ordaz (1981) ve Johnson(1985) tarafından yapılan çalışmalar en önemlileridir. Pregibon (1981) iki grup lojistik modelde etkin (influential), aykırı (outlier) gözlemleri ve belirleme ölçütlerini (diagnostic), Lesaffre (1986), Lesaffre ve Albert (1989) ise çoklu grup lojistik modellerde etkin ve aykırı gözlemlerle belirleme ölçütlerini incelemişlerdir. Lojistik regresyon modellerinin yaygın bir şekilde kullanılır hale gelmesi, katsayı tahmin yöntemlerinin geliştirilmesi ve lojistik regresyon modellerinin daha ayrıntılı incelenmesine sebep olmuştur. Cornfield (1962), lojistik regresyondaki katsayı tahmin işlemlerinde diskriminant fonksiyonu yaklaşımını ilk kez kullanarak popüler hale getirmiştir. Lee (1984) basit dönüşümlü (cross-over) deneme planları için lojistik modeller üzerinde durmuştur. Bonney (1987) lojistik regresyon modelinin kullanımı ve geliştirilmesi üzerinde çalışmıştır. Robert ve diğerleri (1987) lojistik regresyonda standart ki-kare, olabilirlik oranı, “pseudo” en çok olabilirlik tahminleri, uyum mükemmelliği ve hipotez testleri üzerine çalışmalar yapmışlardır. Duffy (1990) lojistik regresyonda hata terimlerinin dağılışı ve parametre değerlerinin gerçek değerlere yaklaşımını incelemiştir. Başarır (1990) klinik verilerde çok değişkenli lojistik regresyon analizi ve ayrımsama sorunu üzerinde çalışmıştır. Hsu ve Leonard (1995) lojistik regresyon fonksiyonlarında

Bayes tahminlerinin elde edilmesi işlemleri üzerine çalışmışlar ve lojistik regresyonda Monte Carlo dönüşümünün kullanılabileceğini göstermişlerdir. Akkaya ve Pazarlıoğlu (1998) lojistik regresyon modellerinin ekonomi alanında kullanımını örneklerle incelemişlerdir (Bircan, 2004:186).

Altaş ve Giray, çalışmalarında mali başarısızlığı tahminlemek için model geliştirmek amacıyla lojistik regresyon yöntemini kullanmışlardır. Tekstil sektöründe faaliyet gösteren İMKB'ye kayıtlı işletmelere ait veriler araştırma çerçevesini oluşturmuştur. Çalışmada öncelikle bu işletmelerin 2001 yılına ait bilançoları yardımıyla mali oranları (rasyolar) hesaplanmış, dönem sonu kar-zarar durumuna bakılarak, işletmeler o dönem için mali başarısız ya da başarılı olarak değerlendirilmiştir. Modelde anlamlı olan değişkenlerin belirlenmesi için öncelikle elde edilen mali oranlara faktör analizi uygulanmış, elde edilen faktör skorları bağımsız değişken olarak alınarak uygulanan lojistik regresyon analizi sonuçlarına göre mali başarısızlığı etkileyen faktörler belirlenmiştir (Altaş ve Giray, 2005).

Keskin Benli, çalışmasında bankaların mali başarısızlıklarının öngörülmesine yönelik lojistik regresyon ve yapay sinir ağı modeline dayanan mali başarısızlık öngörü modelleri geliştirmiştir. Çalışma sonucunda yapay sinir ağı modelinin mali başarısızlığı öngörme gücünün lojistik regresyon modelinden daha üstün olduğu tespit edilmiştir (Keskin Benli, 2005).

Tezcan, yaptığı tez çalışmasında sigorta sektörüne yönelik lojistik regresyon analizini kullanarak bir uygulama yapmıştır. Çalışmada, 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 sınıflandırılmıştır. 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 (Tezcan, 2006).

Özer, Türk bankacılık sektöründe lojistik regresyon yöntemini kullanarak kriz olasılığının incelenmesine ilişkin bir araştırmayı tez olarak sunmuştur. Çalışmada, Latin Amerika, Avrupa ve Asya'da yaşanan bankacılık krizleri incelenmiş ve bu krizlerin öngörüsü için bankacılık verileri ve bazı temel makro ekonomik veriler baz alınarak bir lojistik regresyon modeli oluşturulmuştur. Analizde, bankaların sermaye yeterliliği, aktif kalitesi, likidite ve karlılık ve gelir-gider yapısının temel göstergeleri olana oranlarla bazı temel makro ekonomik göstergeler; cari açık/gayri safi yurtiçi hasıla, ithalat/ihracat, reel kur, kısa vadeli mevduat faizi, kapasite kullanım oranı değişken olarak kullanılmıştır. Araştırmada anlamlı bulunan değişkenler kriz göstergesi olarak modele alınmış ve Türk ekonomisinin son 25 yıl içinde yaşadığı krizlerin etkisi modelde görülmüştür, geleceğe yönelik yapılan tahminlerde ise 2005 yılı son dönemi ve 2006 yılı Mart dönemi için kriz olasılığının olmadığı sonucuna varılmıştır (Özer, 2006).

Hout, Heijden ve Gilchrist, lojistik regresyon modelini cevap değişkenlerinin rastgele cevaplar olması durumunda değerlendirmiştir. Rastgele cevaplar genellikle karşılıklı görüşme ya da anketlerde cevap vericinin isteksizliği durumunda ortaya çıkmaktadır ve yanlış sınıflandırılan kategorik veriler arasında yer almaktadır. Çalışmada kullanılan tek değişkenli model genel doğrusal model olarak alınmış olup çok değişkenli model Fisher algoritması yardımıyla rastgele hale getirilerek oluşturulmuştur. Analizde lojistik regresyon kullanılarak işsizlik tazminatına uyulmamasının düzenlenmesine ilişkin bir çalışma yapılmıştır (Hout, Heijden ve Gilchrist, 2007).

Taç ve Budak, işletmelerin TS-ISO 14000 standartlarını uygulama kararlarına etki eden faktörlerin belirlenmesinde lojistik regresyon analizini kullanmışlardır. Bu kapsamda işletmeler, ISO 14000 Çevre Yönetim Sistemleri Standartlarını uygulayan işletmeler ve uygulamayan işletmeler olarak ele alınmıştır. Araştırma anket çalışmasına dayanmaktadır. Elde edilen veriler ki-kare testi ile analiz edilerek işletmelerin standartlar uygulama kararlarına etki eden istatistiksel olarak anlamlı faktörler belirlenmiş ve bu faktörler için lojistik regresyon modeli uygulanmıştır (Taç ve Budak, 2007).

Ulupınar, 2001 krizi döneminde Türk bankalarının karlılıklarını lojistik regresyon analizi ile inceleyerek bir tez çalışması yapmıştır. Çalışmada, lojistik regresyon analizi teorik olarak incelenmiş, bankaların kriz dönemi, öncesi ve sonrasında karlılıklarını etkileyen faktörler belirlenmeye çalışılmıştır. Türkiye Bankalar Birliği'nden elde edilen her bir rasyo grubundaki değişkenler kendi içinde Faktör Analizi'ne tabi tutularak değişken sayısı azaltılmaya çalışılmıştır. Daha sonra lojistik regresyon analizi ile kriz öncesi olarak seçilen 1996 yılı için, kriz dönemi olarak seçilen 2001 yılı için ve kriz sonrası dönem olarak seçilen 2005 yılı için modeller elde edilerek üç dönem karşılaştırılmıştır. Bağımlı değişken olarak karlılık rasyoları; aktif karlılığı(ROA), özkaynak karlılığı(ROE) ve net faiz marjı(NIM), bağımsız değişken olarak da Türkiye Bankalar Birliği tarafından yayınlanan dört grup finansal rasyo grubu; sermaye yeterliliği, aktif kalitesi, likidite ve gelir-gider rasyo grupları alınmıştır. Bu verilerle üç dönem ayrı ayrı ele alınarak her bağımlı değişken için faktör analizi sonucu elde edilen bağımsız değişkenler ile İkili Lojistik Regresyon Analizi uygulanarak elde edilen sonuçların genel bir değerlendirmesi yapılmıştır (Ulupınar, 2007).

Roy ve Guria, lojistik regresyon modelini gözlemlerin silinmesi tekniğini kullanarak tanımlamaya yönelik çalışmışlardır. Model, en çok olabilirlik yöntemi ile uyumlaştırılmış ve bir gözlem silindikten sonra uyumlu hale getirilen modeldeki tahminlerde oluşan değişim ve sapmalar gözlemlenmiştir. Her gözlem silinmesi sonrası regresyonun yeniden yapılmasına gerek duyulmamış aksine hesaplamada zamandan tasarruf edilmiştir (Roy ve Guria, 2008).

Girginer ve Cankuş, toplu taşıma araçlarından biri olan tramvaya yönelik yolcu memnuniyeti, Eskişehir tramvay sistemi (Estram) örneğini, İkili Lojistik Regresyon Analizi ile incelenmişlerdir. Araştırma Eskişehir'in sahip olduğu her iki üniversiteden basit tesadüfî örnekleme yoluyla seçilen 300 öğrenci üzerinde gerçekleştirilmiştir. Uygulanan lojistik regresyon analizi sonucunda; öğrencilerin Estram'dan memnuniyetleri üzerinde modele alınan tüm bağımsız değişkenlerin negatif etkileri olduğu belirlenmiştir (Girginer ve Cankuş, 2008).

Cengiz, bireylerin mevcut kredi kartını değiştirme tercihi olasılıklarını belirlemeye yönelik demografik değişkenleri ve etki faktörlerini gösteren bir model oluşturmuş ve lojistik regresyon yöntemi ile analiz etmiştir. Araştırma Trabzon, Ordu ve Giresun'da 581 kişi üzerinde uygulanmıştır. Yapılan binary (ikili) lojistik regresyon analizi sonucunda cinsiyet, yaş, kredi kartının kullanabileceği yerlerin çokluğu, kredi kartının limitinin yüksek olması, internetten alışveriş imkanının iyi durumda olması ve kredi kartını sağlayan kurumun imajının tesirli olması değişkenlerinin, kredi kartını değiştirmeme isteği olasılığını değişik seviyelerde etkiledikleri tespit edilmiştir (Cengiz, 2009).

Kim ve Gu, konaklama sektöründe kar payı ödemesi yapan firmaların finansal özelliklerinin değerlendirilmesine yönelik araştırmada lojistik regresyon analizini kullanmışlardır. ABD'de konaklama sektöründe faaliyet gösteren firmaların 2005 yılında ulaşılan 69 firmanın verileri kullanılarak kar payı ödemeleri ve ödememeleri bakımından ayırıcı finansal özelliklerin tanımlanması amaçlanmış ve uygulanan lojistik modelde firma büyüklüğü ile karlılığın anlamlı değişkenler olduğu, bunun yanında yatırım harcamalarının ise ödemeyi zorlaştırdığı görülmüştür (Kim ve Gu, 2009).

Dong, Lai ve Yen, bankacılık sektöründe kredi notlarının değerlendirmesine ilişkin sorunu lojistik regresyon modeli kullanarak değerlendirmişlerdir. Lojistik regresyonun tahmin doğruluğunu geliştirmek için rastgele katsayılar ile analiz edilmesini önermişlerdir. Önerilen kredi notu değerlendirme yönteminin, kredi risk yönetimine katkıda bulunması amaçlanmıştır ve elde edilen sonuçlar değerlendirilmiştir (Dong, Lia ve Yen, 2010).

Bu tez çalışmasının uygulama kısmına, çalışmada anlatılan iki tahmin yönteminin gösterilmesi amacıyla bankaların sektör paylarının tahminlenmesi konu olmuştur. Bu nedenle literatür taramasında Türk Bankacılık sektöründe tahminlemeye yönelik yapılan bazı çalışmalar da aşağıda özetlenmiştir.

Kaya, banka karlılığına ilişkin yaptığı çalışmada 1997-2000 dönemi için panel veri kullanarak karlılık göstergelerinin (net faiz marjı, aktife göre getiri, özkaynağa göre getiri) mikro ve makro belirleyicilerini, Ho ve Saunders (1981) tarafından geliştirilen iki aşamalı yaklaşım kullanarak tespit edilmeye çalışmıştır. Seçilen üç karlılık göstergesinden net faiz marjının mikro belirleyicileri olarak özkaynaklar, likidite, personel harcamaları, mevduatlar ve piyasa payı; makro belirleyicileri olarak enflasyon ve konsolide bütçe açığı anlamlı bulunmuştur. İkinci gösterge olan aktife göre getirinin mikro belirleyicileri olarak kaynak, likidite, personel harcamaları, krediler, kötü aktifler ve mevduatlar; makro belirleyicileri olarak enflasyon ve konsolide bütçe açığı anlamlı bulunmuştur. Üçüncü gösterge olan özkaynağa göre getirinin mikro belirleyicileri olarak ise, özkaynaklar, menkul değerler cüzdanı, likidite, personel harcamaları, krediler, mevduatlar, yabancı para pozisyonu ve piyasa payı; makro belirleyicileri olarak ise enflasyon, konsolide bütçe açığı ve reel faizin anlamlı belirlenmiştir (Kaya, 2002).

Ünsal ve Güler, yaptıkları çalışmada diskriminant ve lojistik regresyon analizlerinin teorik yapısı özetleyerek Türk Bankacılık Sektöründe bu iki tekniğin nasıl sonuçlar verdiğini incelemiş ve 1997-2003 dönemine ait veri için bankaları mali başarılı ve başarısız olarak sınıflandırmada hangi tekniğin daha iyi sonuçlar verdiğini belirlemeyi amaçlamışlardır. Değişken seçiminde hem önsel bilgiden hem de adımsal yöntemlerden yararlanılabilir. Sonuçlar dikkate alındığında, Türk Bankacılık Sektöründe bankaları sınıflandırmak ve bankaların mali durumlarını öngörmek için önsel bilgi ile seçilen değişkenlerle lojistik regresyon analizinin uygun olduğuna karar verilmiştir (Ünsal ve Güler, 2005).

Aktaş ve Kargın, çalışmalarında, Türk Bankacılık Sektöründeki yabancı bankalar ile ulusal bankaları bazı finansal oranlar açısından karşılaştırmıştır. Araştırmanın sonuçlarına göre, yabancı bankalar daha yüksek “Sermaye Yeterliliği” ve “Likidite” oranlarına sahiptir. Ayrıca bazı “Gelir Gider Yapısı”na ilişkin oranlarda da farklılık bulunmaktadır. Söz konusu farklılıklar istatistiksel olarak anlam bulunmuştur (Aktaş ve Kargın, 2007).

Bumin, yaptığı çalışmada, Türk bankacılık sektörünün karlılık performansını, özkaynak karlılık oranının ayrıştırılması yoluyla 2002-2008 dönemi verilerini esas alarak, bankaları fonksiyonlarına ve sahiplik yapılarına göre gruplandırarak oran analizi yöntemiyle incelemiştir. Söz konusu dönemde genel olarak bankacılık sektörünün karlılığı artış göstermiş, ancak küresel piyasalarda yaşanan gelişmeler nedeniyle 2008 yılında karlılık oranlarında düşüş yaşanmıştır (Bumin, 2009).

3.2. Araştırmanın Amacı

Araştırmanın amacı işletmelerin tahminleme sürecinde kantitatif karar verme tekniklerinin kullanımına yönelik bulanık doğrusal regresyon analizi ve lojistik regresyon analizi ile model oluşturmak, işletmelerin karar vermesinde önemli bir araç olan iki tahmin yönteminin kullanılmasını ve uygulanışını göstermek, elde edilen sonuçları karşılaştırmaktır. Bu kapsamda uygulama için bankacılık sektör paylarını her iki analiz yöntemiyle tahmin etmeye yönelik bir örnek uygulama yapılmış olup elde edilen sonuçlar değerlendirilecek ve buna göre ileriye yönelik tahminlerde bulunularak yöntemlerin etkinliği kıyaslanacaktır.

3.3. Araştırmanın Modeli ve Veri Seti

Araştırmada model oluşturulmasında kullanılan veri seti bir bağımlı değişken ve dört adet bağımsız değişkenden oluşmaktadır. Bağımlı değişken bankaların "sektör payları" olmak üzere bağımsız değişkenler "sermaye yeterliliği", "likidite", "özkaynak karlılığı (ROE)" ve "vergi öncesi karlılık" tır. Analizde kullanılan veriler "Türkiye Bankalar Birliği"nin internet sitesinde yayınladığı istatistiki raporlarda yer alan 2008 ve 2009 yılına ait "Seçilmiş Rasyolar" raporundan elde edilmiş ve Türk Bankacılık sektöründe faaliyet gösteren 45 bankaya ait veriler kullanılmıştır. (http://www.tbb.org.tr/tr/Banka_ve_Sektor_Bilgileri/Tum_Raporlar.aspx, Erişim: 30.05.2010). Araştırmada model oluşturmak ve geleceğe yönelik verilerle tahmin yapabilmek için 2008 yılı verileri, oluşturulan bu modelle tahmin yapmak içinse

2009 yılı verilerinden faydalanılmıştır. Sektör paylarının tahmin edilmesinde kullanılan dört bağımsız değişkenin belirlenmesinde, "Bankacılık Düzenleme ve Denetleme Kurulu"nun yayınladığı "Finansal Piyasalar Raporları"nda yer alan bankacılık sektörüne ilişkin finansal sağlamlık göstergeleri arasında bulunmaları ve bankacılık sektörü ile ilgili yapılan çalışmalarda çoğunlukla değişken olarak analizlerde kullanılmaları etkili olmuştur. (http://www.bddk.org.tr/websitesi/turkce/Raporlar/Finansal_Piyasalar_Raporlari/Finansal_Piyasalar_Raporlari.aspx, 30.05.2010). Modelin kurulmasına uygun bir yapı oluşturması açısından finansal sağlamlık göstergeleri arasında yer alan rasyolar denenmiş, bu dört değişkenin en anlamlı yapıyı oluşturması dolayısıyla uygulamada, belirtilen dört değişkenle analiz yapılmıştır.

Yapılan uygulamada bankaların sektör paylarının tahminlenmesine yönelik model için aşağıda, dört adet bağımsız değişkeni olan, bulanık doğrusal regresyon ve lojistik regresyon olmak üzere iki model gösterilmektedir ve bu modellere ait parametreler yer almaktadır. Analizde yer alan dört açıklayıcı değişken için bulanık regresyon modelinde sabit terim de dahil olmak üzere merkezi (a) ve yayılım (c) değerlerini gösteren 10 adet parametre, lojistik modelde ise sabit katsayı (β_0) da dahil olmak üzere 5 adet parametre bulunmaktadır.

<u>Bulanık doğrusal regresyon modeli:</u>

$$\tilde{Y}_i = \tilde{A}_0 + \tilde{A}_1 x_1 + \tilde{A}_2 x_2 + \tilde{A}_3 x_3 + \tilde{A}_4 x_4$$

Y = Bankaların sektör payları

x_1 = Sermaye yeterliliği (Özkaynaklar/(Kredi+Piyasa+Operasyonel Riske Esas Tutar)

x_2 = Likidite (Likit Aktifler/Toplam Aktifler)

x_3 = Özkaynak karlılığı (ROE = Net Dönem Karı(Zararı)/Özkaynaklar)

x_4 = Vergi öncesi karlılık (Vergi Öncesi Kar/Toplam Aktifler)

Bulanık parametreler:

$\tilde{A}_0\ (a_0, c_0),\quad \tilde{A}_1\ (a_1, c_1),\quad \tilde{A}_2\ (a_2, c_2),\quad \tilde{A}_3\ (a_3, c_3),\quad \tilde{A}_4\ (a_4, c_4)$

Bulanık Model için Doğrusal Programlama Amaç Denklemi:

$$MIN\sum_{j=0}^{N}\left(c_j\sum_{i=1}^{M}\left|x_{ij}\right|\right) = Nc_0 + \left(c_j\sum_{i=1}^{M}\left|x_{ij}\right|\right)$$

Kısıtlar:

$$\sum_{j=0}^{N} a_j x_{ij} + (1-h)\sum_{j=0}^{N} c_j\left|x_{ij}\right| \geq Y_i \qquad (j = 0,1,3,4)$$

$$\sum_{j=0}^{N} a_j x_{ij} - (1-h)\sum_{j=0}^{N} c_j\left|x_{ij}\right| \leq Y_i$$

$c_j \geq 0$

Lojistik regresyon modeli:

$$ln\left(\frac{p}{1-p}\right) = ln(odds) = e^{\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4}$$

$$p = \pi(x) = \frac{odds}{1+odds} = \frac{e^{\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4}}$$

Yukarıda Bulanık doğrusal regresyon ve Lojistik regresyon modellerinin uygulamada kullanılan dört bağımsız değişkene sahip genel gösterimleri verilmiştir. Model oluşturmak için kullanılan 2008 yılının bağımlı ve bağımsız değişkenlere ait verileri aşağıda Tablo 3'te gösterilmektedir.

Tablo 3: Modelde Kullanılan Değişkenlere Ait Veriler

	Y	x_1	x_2	x_3	x_4
1	14,8	20,1	21,9	29	2,6
2	7,2	14,5	14,9	23,7	2,5
3	7,4	14,3	30,4	13,3	1,8
4	0,01	186,4	93,2	6,6	5,6
5	12,1	18,2	21,4	15,2	2,4
6	0,5	14,1	28,9	14,1	1,8
7	0,5	18,5	17,8	17,1	3,2
8	1,1	14,7	18,2	14,8	2,3
9	0,4	17,9	26,9	2,8	0,5
10	0,1	34,5	70,3	6,9	1,4
11	2,1	17,7	31,6	11,5	1,3
12	12,6	16,1	31,9	18,5	2,4
13	13,8	15,2	41,3	16	1,8
14	9	15,7	13,1	15,2	2,1
15	1	65,1	85,1	12,2	12,4
16	0,1	34,1	31	1,1	-0,2
17	0,8	17,9	51,7	10,2	1,9
18	2,7	17,2	24,5	13,7	1,7
19	0,1	40,7	57	10,1	6,3
20	0,5	17,9	35,1	4,5	0,4
21	3,8	16	17,1	12,8	1,9
22	1,7	14,8	32,2	8	1,6
23	2,1	15,4	29,5	11	2,1
24	2,3	13,8	27,5	8,7	1,1
25	0,2	22,3	22,7	1,9	0,1
26	0,1	21,9	31,6	0,4	0,1
27	0,2	34,4	81,2	16,9	3,3
28	0,04	51,6	41	25,4	5,4
29	0,01	86,8	72,1	12,3	6,9
30	0,03	121,8	97,2	14,2	10,9
31	0,1	23,2	75	-47,4	-3,3
32	0,1	71,8	94,3	16,6	3,2
33	1	70,8	20	5,7	6
34	0,7	104,9	13,6	12,6	7,5
35	0,1	79,2	32,6	7,6	4,5
36	0,03	35,6	42,6	1	0,5
37	0,02	64,8	62,9	9,6	6,4
38	0,01	49,2	33,1	15,8	14,5
39	0,2	60,1	91,5	16,6	3,9
40	0,01	22,2	33,1	3,3	1,6
41	0,9	21,1	31,2	15,9	2,4
42	0,2	32,4	23,1	7,2	2,6
43	0,01	13,9	70,7	-1,1	-0,4
44	0,01	108,9	79,6	-3,2	-1,5
45	0,1	78,5	24,9	-15,9	-14,3

3.4. Bankaların Sektör Paylarının Tahminlenmesinde Bulanık Doğrusal Regresyon ve Lojistik Regresyon Modelinin Uygulanması

Bankaların sektör paylarını tahminlenmesinde bulanık doğrusal regresyon ve lojistik regresyon modelinin değişkenlere ilişkin belirtilen verilerle uygulanması iki alt bölümde incelenmiş olup, ilk olarak bulanık regresyon modeli ile ardından lojistik regresyon modeli ile uygulanışı gösterilmiştir.

3.4.1. Bulanık Doğrusal Regresyon Modelinin Uygulanması

Bulanık doğrusal regresyon modelinin uygulanmasında klasik regresyondaki gibi normallik varsayımı, rastgelelik incelemeleri, durağanlık testleri vb. gerekmediği daha önce de bahsedilmiştir. Bu doğrultuda modelin uygulanması için gereken tek kriter değişkenlere ait verilerin doğrusal olmalarıdır. Bu nedenle, bulanık modelde yer alan değişkenlere ait verilerin doğrusallaştırılmaları için 45 bankaya ait verilerin doğal logaritmaları (ln) alınarak doğrusal bir form oluşturulmuş ve değişkenler analizde bu şekilde yer almışlardır (Bakınız: EK 1). Bankaların sektör paylarının tahmininde bulanık regresyon analizi için doğrusal programlama tekniği kullanılmıştır ve bağımlı değişkeni etkilediği düşünülen açıklayıcı değişkenler analize katılmıştır. Analizde yer alan 45 bankaya ait veriler, oluşturulan bulanık doğrusal programlama modelinde her bir gözlem için alt ve üst sınır elde etmek amacıyla 90 adet kısıt oluşturmakta ve bu kısıtlar tahminleme yapmak için bir bulanık regresyon aralığı sunmaktadır. Kısıtlar, her bir bankaya ait sektör payı için bir alt ve bir üst sınır hesaplanabilmesi adına bağımlı değişkene ait değerler büyük eşit ($\geq$) ve küçük eşit ($\leq$) şeklinde eklenerek oluşturulmuştur. Bulanık regresyon modelinde bulunan yayılımların minimizasyonunu sağlayan doğrusal programlama amaç denklemi aşağıda gösterildiği gibidir.

$$\text{Min } (45c_0 + 153{,}81c_1 + 161{,}67c_2 + 97{,}98c_3 + 34{,}62c_4)$$

Bulanık doğrusal regresyon modelinde kısıtlar oluşturulurken gözlemlerin bulanıklığını ifade eden uygun bir *h* seviyesinin belirlenmesi gerekmektedir, araştırmacıların genellikle bu seviyeyi 0,5 olarak aldıklarından daha önce söz edilmiştir. Yapılan bu çalışmada da *h* seviyesi 0,5 olarak belirlenmiştir ve tahmin aralığını oluşturan kısıtlar buna göre oluşturulmuştur. Aşağıda 1. gözlem (banka) değerleri kullanılarak (doğal logaritmaları alınarak doğrusallaştırılmış değerler) oluşturulan ilk iki kısıt gösterilmektedir. Bulanık regresyon analizinde, kısıtlar 45 banka için hesaplandığında daha önce de belirtildiği gibi 90 kısıt meydana gelmektedir.

$a_0 + 3{,}001a_1 + 3{,}086a_2 + 3{,}367a_3 + 0{,}956a_4 + 0{,}5c_0 + 2{,}308c_1 + 2{,}393c_2 + 2{,}674c_3 + 0{,}262c_4 \geq 2{,}695$

$a_0 + 3.001a_1 + 3{,}086a_2 + 3{,}367a_3 + 0{,}956a_4 - 0{,}5c_0 - 2{,}308c_1 - 2{,}393c_2$ $2{,}674c_3 - 2{,}695c_4 \leq 2{,}695$

Bulanık modelin doğrusal programlama ile çözümü "POM-QM for Windows" programı kullanılarak yapılmış ve bulanık parametrelere ait merkezi değerler ve yayılım değerleri aşağıdaki gibi bulunmuştur.

$a_0 = 0$	$c_0 = 0$
$a_1 = 0$	$c_1 = 0{,}0161$
$a_2 = 0{,}1562$	$c_2 = 0{,}2707$
$a_3 = 0{,}2632$	$c_3 = 0{,}2711$
$a_4 = 0$	$c_4 = 0{,}9743$

Buna göre her bağımlı değişken için elde edilen bulanık parametre, bulanık doğrusal regresyon modelinde yerine konulduğunda aşağıda gösterilen eşitlik oluşturulmuştur. Sabit terimler olan a_0 ve c_0 değerleri sıfır olarak bulunduğu için etkisiz oldukları nedeniyle modelde yer almaları gerekmemektedir.

$\tilde{Y}$= (0, 0,0161)x_1 + (0,1562, 0,2707)x_2 + (0,2632, 0,2711)x_3 + (0, 0,9743)x_4

Elde edilen merkezi ve yayılım değerleri ışığında modelde bulunan bulanık parametreler için alt ve üst sınır değerleri hesaplanmıştır. Üst sınır değeri için, a_i ve c_i değerlerinin toplamı, alt sınır değeri için bu değerlerin farkları alınmıştır ve doğrusal programlama ile tahmin edilen bulanık parametrelerin alt ve üst sınır değerleri aşağıda Tablo 4'te verilmiştir.

Tablo 4: Bulanık Parametrelerin Alt ve Üst Sınır Değerleri

	A_0	**A_1**	**A_2**	**A_3**	**A_4**
ÜST (a_i+c_i)	0	0,0161	0,4269	0,5343	0,9743
ALT (a_i-c_i)	0	-0,0161	0,1145	-0,0075	-0,9743

Buna göre bulanık parametrelere ilişkin alt ve üst sınır değerler kullanılarak bulanık doğrusal regresyon analizi için oluşan alt ve üst regresyon doğruları aşağıdaki gibi elde edilir.

$\tilde{Y}_{üst}$ = 0,0161x_1 + 0,4269x_2 + 0,5343x_3 + 0,9743x_4

$\tilde{Y}_{alt}$ = -0,0161x_1 + 0,1145x_2 – 0,0075x_3 – 0,9743x_4

Oluşturulan alt ve üst bulanık regresyon modellerinin tahmin yeteneğini ölçmek için geleceğe yönelik tahminde bulunmak amacıyla 2009 yılının aynı değişkenlere ait verileri (ln değerleri) kullanılarak sektör paylarına ilişkin bulanık regresyon aralığı tahminlenmiştir. 2009 yılına ait kullanılan verilerin gerçek değerleri ve doğal logaritmaları alınarak doğrusallaştırılmış değerleri EK 2'de verilmiştir. Bu verilerle yapılan tahmine ilişkin alt ve üst sınırlar ile gerçekleşen Y_i değerleri aşağıda yer alan Tablo 5'te gösterilmektedir.

Tablo 5: Tahmin Edilen Bulanık Aralık ve Ln(Y_i) Değerleri

	$\tilde{Y}_{üst}$	$\tilde{Y}_{alt}$	Ln(Y_i)	
1	4,66	0,34	2,75	1
2	4,17	0,33	2,03	1
3	3,95	0,26	2,09	1
4	2,80	0,00	-5,06	0
5	4,40	0,27	2,48	1
6	3,11	0,26	-0,79	0
7	4,15	0,28	-0,73	0
8	3,54	0,23	0,11	0
9	1,61	0,07	-1,32	0
10	-0,44	-0,08	-2,05	0
11	3,37	0,23	0,63	1
12	4,55	0,29	2,58	1
13	4,07	0,26	2,65	1
14	3,63	0,27	2,09	1
15	5,23	0,16	-2,29	0
16	3,85	0,19	-2,11	0
17	3,58	0,18	-0,49	0
18	4,13	0,28	0,98	1
19	5,61	0,28	-1,43	0
20	2,10	0,13	-0,73	0
21	3,80	0,27	1,30	1
22	2,61	0,14	0,34	1
23	3,50	0,20	0,55	1
24	3,17	0,20	0,65	1
25	3,57	0,25	-2,01	0
26	0,28	-0,04	-1,94	0
27	5,46	0,25	-1,90	0
28	5,07	0,30	-3,01	0
29	4,82	0,16	-4,69	0
30	5,99	0,20	-3,69	0
31	5,47	0,42	-3,10	0
32	5,42	0,25	-2,66	0
33	4,00	0,13	0,00	0
34	4,41	0,19	-0,21	0
35	3,19	0,13	-1,83	0
36	3,22	0,13	-2,63	0
37	3,63	0,22	-4,44	0
38	2,94	0,11	-4,35	0
39	4,48	0,22	-1,68	0
40	2,91	0,11	-3,82	0
41	4,15	0,26	-0,15	0
42	2,82	0,16	-1,61	0
43	6,27	0,24	-4,82	0
44	4,93	0,24	-3,65	0
45	4,94	0,29	-5,76	0

Bulanık doğrusal regresyon analizinde belirlenen alt ve üst sınır regresyon modelleri kullanılarak ileriye dönük tahminde bulunabilmek için 2009 yılına ait veriler kullanılmıştır. Bu veriler ile tahmin edilen bulanık regresyon aralığının alt ve üst sınırları ile gerçekte olan Y_i değerleri yani 45 bankaya ait sektör payları ln değerleri elde edilerek yukarıda Tablo 5'te gösterilmiştir. Sektör payları oran olarak gerçekleştiği için bu değerin aralık olarak tahminlenmesi uygun bulunmuştur. Buna göre tahmin edilen bulanık aralığın gerçekleşen değerlerle karşılaştırması yapılmış ve ne derece doğru bir aralık verdiği incelenmiştir. Tahmin edilen aralığın içinde gerçekleşen bağımlı değişken değerlerine tablodaki son sütunda 1 değeri, bu aralığın içinde gerçekleşmeyenlere ise 0 değeri verilmiştir. Bu durumda, 45 banka içersinden 13 tanesi bu aralığa denk düşmüş kalanı ise aralığın dışında kalmıştır, dolayısıyla bulanık regresyonun doğru tahmin etme oranı yaklaşık %29 olarak belirlenmiş ve tahmin gücü bu uygulama için düşük bulunmuştur.

3.4.2. Lojistik Regresyon Modelinin Uygulanması

Bankaların sektör paylarının tahminlenmesine yönelik yapılan bu uygulamada oluşturulan lojistik regresyon modelinde bağımlı değişkenin ikili olduğu durum ele alınmıştır, dolayısıyla bağımlı değişken istenilen ve istenilmeyen iki durum olarak sınıflandırılmıştır. Bu uygulamada yapılan lojistik regresyon analizinde, bağımlı değişkeni temsil eden sektör paylarının 45 banka içinden ilk on banka arasına girip girmemesi durumu incelenmiştir. Buna göre ilk on banka arasına girenler 1 olarak, ilk on banka arasına giremeyenler ise 0 olarak kodlanmış ve lojistik modelin çözümü yapılmıştır. Modelin çözümü için "SPPS" programı kullanılmış ve "enter" metodu ile çözümde aşağıdaki sonuçlar elde edilmiştir.

Tablo 6: Sınıflandırma Tablosu

Gözlenen		Tahminlenen		
		sektör payları ilk 10		Doğru Sınıflama Oranı
		0	1	
sektör payları ilk 10	0	33	2	94,3
	1	1	9	90,0
Toplam Oran				93,3

Lojistik regresyon analizinden elde edilen "Block 1" çıktılarında görülen sınıflandırma tablosu Tablo 6'da gösterilmiştir. Toplam 45 bankaya ait sektör payları 2008 yılı için ilk on banka arasına girip girmeme durumuna göre sınıflandırılmıştır. Buna göre, ilk 10 banka arasına giremeyen ve 0 olarak atanan 35 bankadan 2'sinin aslında ilk 10 banka arasında yer alması gerektiği ve 1 olarak atanması gerektiği anlaşılmaktadır. Sonuçta 0 olarak atananların doğru sınıflama oranı %94,3 çıkmıştır. İlk 10 banka arasına girdiği belirlenen ve 1 olarak atanan 10 bankadan 1'inin ise 0 olarak atanması gerektiği görülmektedir. Bunların doğru sınıflandırma oranı ise %90 çıkmıştır. Modelin tamamına bakıldığında toplam doğru sınıflandırma oranı %93,3 olarak bulunmuş ve modelin sınıflandırmada başarılı olduğu anlaşılmıştır.

Tablo 7: Modelin Katsayılarının Genel Testi

		Ki-kare	df	Sig.
Step 1	Step	32,067	4	0,00
	Block	32,067	4	0,00
	Model	32,067	4	0,00

Tablo 7'de Modelin katsayılarının genel testine ilişkin veriler görülmektedir. Buna göre enter metoduyla 1. adımda (Step 1) değişkenlerin modele dahil edilmesiyle çıkan sonuçlarda 11 iterasyon sayısına ulaşıldığı, bu aşamadan sonra –2LL(-2Loglikelihood) değerinin çok küçük bir değişim gösterdiği ve iterasyona son verildiği anlaşılmaktadır. İterasyon geçmişine bakıldığında sadece sabit katsayının modelde olduğu ilk –2LL değeri 47,674 olarak, modele değişkenler girdikten sonraki yapılan hesaplamalarda ise bu değerin 15,606 olduğu görülmektedir (Bakınız: EK 3). Bu iki değer arasındaki fark ise Model Ki-kare istatistiğini vermekte ve modeldeki katsayılara ilişkin genel test yapılmasını sağlamaktadır. Tablo 7'de gösterildiği gibi model ki-kare değeri 32,067 ve bunun sig. (anlamlılık) değeri ise 0,00 bulunmuştur. Anlamlık değeri 0,05'ten küçük olduğu için modelin %95 güven seviyesinde istatistiksel olarak anlamlı olduğuna karar verilmiştir.

Tablo 8: Modeldeki Değişkenlere Ait Veriler

	β	S.E.	Wald	df	Sig.	Exp(β)
x_1	-0,486	0,257	3,573	1	0,059	0,615
x_2	-0,091	0,098	0,857	1	0,355	0,913
x_3	0,709	0,351	4,080	1	0,043	2,032
x_4	-2,902	1,465	3,921	1	0,048	0,055
Sabit	6,406	6,314	1,029	1	0,310	605,271

Lojistik regresyon analizinin SPSS programındaki çözümünde elde edilen çıktılara bakıldığında, yukarıda değişkenlerin modele alındığındaki değerlerini gösteren Tablo 8'de anlamlı bulunan değişkenler ile lojistik regresyon modeli oluşturulmuştur. Tablo 8'de değişkenlere ait "β" değerleri (katsayılar), S.E (standart hata), Wald istatistiği, df (serbestlik dereceleri), sig. (anlamlılık) ve exp(β) yani e^{β} değerleri görülmektedir. Exp(β) değeri, modeldeki diğer değişkenler sabit tutularak ilgili değişkenin 1 birim artması durumunda bağımlı değişkenin ne oranda değişeceğini ifade etmektedir. Wald istatistiği ise değişkenlere ait β değerlerinin standart hatalarına oranının karesi alınarak elde edilmektedir. Bu bağlamda, x_1 değişkeninin lojistik regresyon katsayısı -0,486 ve standart hatası 0,257 olduğundan Wald istatistiği $(-0,486/0,257)^2 = 3,573$ olarak hesaplanmaktadır. Wald istatistiğinin anlamlılığı için katsayıların anlamlılığı test edilirken yapıldığı gibi sig. değerine bakılmaktadır. Buna göre x_1'in anlamlılık değeri 0,059 çıkmıştır. Anlamlık değerinin 0,05'ten küçük olması bu değişkenin modelde bulunmasının %95 güven düzeyinde anlamlı olduğunu ifade etmektedir. 0,059 değerinin 0,05'e çok yakın olması, modelde bağımlı değişkenin değerini etkilemesi ve bu değerin %90 güven düzeyinde kabul edilebilir olması nedeniyle modele dahil edilmesi uygun bulunmuştur.

Aşağıda değişkenlerin katsayılarına ait anlamlılık hipotezleri verilmiştir. H_0 hipotezi, değişkenlerin modelde bulunmalarının anlamsız olduklarını ve katsayılarının 0'a eşit olduklarını, alternatif hipotez H_1 ise değişkenlerin katsayılarının 0'dan farklı olduğunu savunmaktadır. Bu durumda x_1 değişkeni için H_0 reddedilerek H_1 kabul edilmiştir. Aynı şekilde diğer açıklayıcı değişkenlerin anlamlık değerlerine bakıldığında x_2'nin 0,355 değerle anlamsız olduğu görülmekte ve H_0 kabul edilmektedir, x_3'ün 0,043 değerle, x_4'ün 0,048 değerle anlamlı olduğu anlaşılmakta ve H_0 reddedilmektedir.

$$H_0 = \beta_1, \beta_2, \beta_3, \beta_4 = 0$$
$$H_1 = \beta_1, \beta_2, \beta_3, \beta_4 \neq 0$$

Bu doğrultuda anlamlı bulunan bağımsız değişkenler x_1, x_3 ve x_4'ün katsayıları ve sabit katsayı modele alındığında aşağıdaki lojistik regresyon modeli elde edilmiştir.

$$\ln(\text{odds}) = e^{6,406 - 0,486x_1 + 0,709x_3 - 0,2,902x_4}$$

Tablo 9: Modelin Uyum İyiliği

-2 Log Likelihood	Cox & Snell R^2	Nagelkerke R^2
15,606	0,510	0,780

Yukarıda yer alan Tablo 9'da modelin uyum iyiliğinin belirlenmesini sağlayan -2 Loglikelihood, Cox & Snell R^2 ve Nagelkerke R^2 istatistiklerine ilişkin değerler görülmektedir. Bu istatistikler bağımlı değişkenin bağımsız değişkenler tarafından ne ölçüde açıklanabildiğine, aralarındaki ilişkinin ne oranda olduğuna işaret etmektedir. -2LL değeri 15,606 bulunmuş ve Tablo 7'deki değerler ile ele alındığında modelin uyumlu olduğundan bahsedilmiştir. Cox & Snell R^2 istatistiği modelin %51 oranında, Nagelkerke R^2 istatistiği ise %78 oranında uyumlu olduğunu göstermektedir. Bu durumda modelin açıklayıcı gücünün iyi seviyede olduğu söylenebilir.

Lojistik regresyon analizinde sektör paylarının tahminlenmesine yönelik oluşturulan model ile geleceğe dönük tahminde bulunabilmek amacıyla 2009 yılına ilişkin aynı bağımlı ve bağımsız değişkenlere ait gerçekleşen veriler kullanılarak bir tahminde bulunulmuştur. Bu tahminle lojistik regresyon için oluşturulan modelin tahminlemede ne kadar etkin olduğu gözlenmiştir. Buna göre oluşturulan modelle kullanılarak lojistik regresyon analizi uygulanmış ve 2009 yılına ilişkin hesaplamalar yapılarak Tablo 10'da gösterildiği gibi sonuçlar elde edilmiştir.

Tablo 10: Lojistik Regresyon Analizi Tahmin Sonuçları

	ln(odds)	odds	p	Y_i	
1	8,87	7119,79	1,00	**15,59**	1
2	9,04	8463,66	1,00	**7,60**	1
3	4,03	56,00	0,98	**8,11**	1
4	-87,21	0,00	0,00	0,01	1
5	-0,94	0,39	0,28	**11,94**	**0**
6	4,08	59,09	0,98	0,45	**0**
7	-1,25	0,29	0,22	0,48	1
8	0,70	2,02	0,67	1,12	1
9	-3,91	0,02	0,02	0,27	1
10	-7,37	0,00	0,00	0,13	1
11	1,89	6,60	0,87	1,89	**0**
12	1,48	4,39	0,81	**13,21**	1
13	2,42	11,29	0,92	**14,18**	1
14	1,68	5,35	0,84	**8,08**	1
15	-34,82	0,00	0,00	0,10	1
16	-12,18	0,00	0,00	0,12	1
17	-2,68	0,07	0,06	0,61	1
18	2,39	10,93	0,92	**2,66**	1
19	-19,98	0,00	0,00	0,24	1
20	-4,81	0,01	0,01	0,48	1
21	4,52	92,01	0,99	**3,67**	1
22	-0,82	0,44	0,31	1,41	1
23	-1,33	0,27	0,21	1,73	1
24	0,41	1,50	0,60	**1,91**	1
25	-7,54	0,00	0,00	0,13	1
26	-2,99	0,05	0,05	0,14	1
27	-26,91	0,00	0,00	0,15	1
28	-16,68	0,00	0,00	0,05	1
29	-50,47	0,00	0,00	0,01	1
30	-82,44	0,00	0,00	0,02	1
31	-37,27	0,00	0,00	0,05	1
32	-40,07	0,00	0,00	0,07	1
33	-39,15	0,00	0,00	1,00	1
34	-63,37	0,00	0,00	0,81	1
35	-31,09	0,00	0,00	0,16	1
36	-5,95	0,00	0,00	0,07	1
37	-41,08	0,00	0,00	0,01	1
38	-19,94	0,00	0,00	0,01	1
39	-18,13	0,00	0,00	0,19	1
40	-7,55	0,00	0,00	0,02	1
41	-2,64	0,07	0,07	0,86	1
42	-9,64	0,00	0,00	0,20	1
43	-42,45	0,00	0,00	0,01	1
44	-31,01	0,00	0,00	0,03	1
45	-0,64	0,53	0,35	0,00	1

Tablo 10'da görülen lojistik regresyon analizinin tahmin sonuçlarında modele alınan değişkenlere ait veriler ile hesaplanan ln(odds) değerleri ve p değerleri verilmiştir. 2009 yılı için gerçekleşen Y_i bağımlı değişken değerleri yani sektör payları da tabloda yer almıştır ve ilk 10 banka arasına girenler kalın yazılarak vurgulanmıştır. Lojistik regresyon analizi için tahmine konu olan bankaların ilk 10 banka arasına girip giremediklerinin olasılığını hesaplanan p değerleri vermektedir. Olasılıkların yüksek olduğu durumlar analizdeki bankaların ilk on banka arasına girdiğini, olasılıkların düşük olduğu durumlar ilk on banka arasına giremediklerini ifade etmektedir. Bu durumun doğru tahminlendiği bankalara tablonun son sütununda 1 değeri atanmıştır, bankanın ilk on arasında yer aldığı fakat olasılık değerinin düşük kalarak bunu yanlış tahminlediği ya da tersi durumlara ise 0 değeri atanmıştır. Böylece lojistik regresyon analizinin tahmin sonuçlarının doğruluğu değerlendirilmiştir.

Buna göre %15,59 değerle en yüksek sektör payına sahip bankanın ilk on banka içinde olma olasılığı %100 sonucunu vermiş ve lojistik modelin bunu doğru sınıflandırdığı belirlenerek 1 değeri atanmıştır. Aynı şekilde %8,11 sektör payı ile ilk on arasına giren 3. bankanın da lojistik regresyon analizi sonucunda %98 olasılıkla ilk on banka arasında yer alacağı tahminlenmiştir ve sonuç doğru olduğu için 1 atanmıştır. %0,01 ile çok düşük bir sektör payına sahip olan 4. bankanın ilk on banka arasına girme olasılığı %0 olarak hesaplanmış ve lojistik regresyonun doğru sonucu verdiği görülmüştür. 5. bankaya bakıldığında lojistik regresyonda yanlış bir tahmin söz konusu olmuştur. %11,64'lük bir sektör payına sahip olan bankanın ilk on arasına girme olasılığı %28 çıkmış yani ilk on arasına giren bir bankaya düşük olasılık tahminlendiği için bunun yanlış sınıflandırıldığı belirlenmiştir. Yine %0,45 sektör payına sahip 6. bankanın da p değeri %98 hesaplandığından ilk on banka arasında yer aldığı sonucu çıkmış ve yanlış sınıflandırıldığı için bu değere 0 atanmıştır. Üçüncü ve son yanlış atama ise %1,89 sektör payına sahip 11. bankadır. Bu bankanın da %87 olasılıkla ilk on banka arasında yer aldığı sonucu alınmıştır. Ancak banka ilk ona giremese de sektör paylarına bakıldığında ilk on banka arasına giren bankaların hemen ardında 11. olarak bulunmaktadır, dolayısıyla burada lojistik regresyon sonucunun çok da yanıldığı söylenemez. %87 ihtimalle bu bankanın ilk on arasında yer alması aslında mümkündür.

Sonuçta lojistik regresyon tıpkı 2008 yılının sonuçlarında çıktığı gibi 2009 yılı için yapılan tahminde de sadece 3 banka için yanılmıştır. Tüm tahmin sonuçları ele alındığında, lojistik regresyon modelinin verdiği sonuçların sektör paylarını tahminlemede başarı gücünün yüksek olduğu açıkça görülmektedir. Yapılan analizde Tablo 10'daki sonuçlarda görüldüğü gibi uygulamada yer alan 45 bankadan sadece 3 tanesinde yanılma tespit edilmiştir ve modelin tahmin gücünün oldukça yüksek olduğu göze çarpmıştır.

SONUÇ VE ÖNERİLER

İşletmelerin tahminleme sürecinde alternatif analiz yöntemlerinin uygulanmasına yönelik olarak yapılan bu çalışmada bulanık doğrusal regresyon analizi ve lojistik regresyon analizinin uygulanması anlatılmıştır. İşletmeler için tahmin yapmak, geleceği öngörmek kuşkusuz günümüzde önemli bir araç haline gelmiştir. Tahminleme süreci tüm işletme kararlarının temelini oluşturmaktadır. İşletmeler yaşamlarını devam ettirebilmek ve rekabet edebilmek amacıyla tahminlerde bulunurlar ve bu tahminleri esas alarak planlar yapıp kararlar alırlar. Bu nedenle geleceğe yönelik tahminlerde bulunmak için şirketler, birçok uygulamada faydalı olan çeşitli istatistiksel analiz tekniklerini kullanmaya yönelmişlerdir.

Bu tahmin yöntemlerinden sebep sonuç ilişkisine dayanan regresyon tekniği sıklıkla kullanılmaktadır. Regresyon analizleri birbiri ile ilişkili olan bağımlı ve bağımsız değişkenler arasında modeller oluşturularak geleceğe yönelik tahminler yapmayı sağlamaktadırlar. Bu çalışma regresyon analizi yöntemlerinden Bulanık doğrusal regresyon analizi ve lojistik regresyon analizi ele alınmıştır. Bu iki yöntem klasik regresyon analizine alternatif oluşturarak son yıllarda sıklıkla uygulanan yöntemler olmuşlardır. Dolayısıyla bu çalışmada da her iki analiz yöntemi teoriksel olarak incelenmiş olup, uygulamadaki kullanımlarına ilişkin örnek teşkil etmesi amacıyla banka sektör paylarının tahminlenmesine yönelik örnek bir uygulama sunulmuştur.

Bu kapsamda çalışmanın birinci bölümünde bulanık doğrusal regresyon analizi ele alınmıştır. Birinci bölümde bulanık regresyon düşüncesine temel oluşturan bulanık mantıktan bahsedilmiştir. Klasik düşüncenin yani sadece iyi ve kötü gibi kesin verilerin dışında belirsizliklerin de hayatın her alanında bulunduğundan bahsedilmiştir. Bulanık regresyon analizi de, sistem yapısından kaynaklanan bu belirsizlikleri analize katabilme gücüyle bazı tahminleme sorunlarında karar vericiye bir alternatif oluşturmaktadır. Bu doğrultuda birinci bölüme bulanık regresyonun Tanaka tarafından geliştirilen temel modeli anlatılmış ve modelin doğrusal programlama tekniği ile çözümü için oluşturulan kısıtlar ve amaç denklemi gösterilmiştir. Birinci bölümün devamında bulanık regresyon analizinin halen gelişmekte ve temel model üzerinden geliştirilen yöntemlerinin önemli olan iki tanesine değinilmiştir. Bunlar bulanık en küçük kareler yöntemi ve aralık regresyonudur.

Çalışmanın ikinci bölümünde lojistik regresyon analizi yer almaktadır. Lojistik regresyon analizi, son yıllarda sosyal bilimler uygulamalarında oldukça öne çıkan bir yöntem olmuştur ve bağımlı değişkenin kesikli olduğu durumlarda sınıflama yapmaya yarayan bir tekniktir. Bu çalışmada lojistik regresyonda, bağımlı değişkenin ikili olduğu yani istenilen durum için 1 değerini istenilmeyen durum için 0 değerini aldığı durum üzerinde durulmuştur. Bu doğrultuda, ikinci bölümde lojistik regresyon modeli anlatılmış, lojistik regresyonun fonksiyonu, odds oranı ve logit modele ilişkin bilgilere yer verilmiştir. Bölümün devamında modelde bulunan değişkelere ait parametrelerin tahmini ve parametre tahmin yöntemlerine değinilmiştir. Daha sonra analizde yer alan değişkenlerin katsayılarının anlamlılığı ve yorumlanmasından bahsedilerek, modelin uyum iyiliğini belirlemekte kullanılan istatistiksel ölçütler anlatılmıştır.

Üç bölümden oluşan çalışmanın üçüncü bölümünde ise sektör paylarının tahminlenmesine yönelik bulanık doğrusal regresyon analizi ve lojistik regresyon analizinin uygulanması anlatılmıştır. Her iki tahmin yönteminin kullanılmasını ve uygulanmasını göstermek amacıyla yapılan uygulamada Türk bankacılık sektöründe yer alan 45 bankaya ait veriler kullanılarak sektör paylarının tahminlenmesi örneği verilmiştir. Analizde kullanılan veriler Türkiye Bankalar Birliğinin yayınladığı istatistiksel raporlar arasından yer alan 2008 ve 2009 yılına ait Seçilmiş Rasyolar Raporundan elde edilmiştir. Bu doğrultuda, analizde bağımlı değişken olarak sektör payları, bağımsız değişken olarak sermaye yeterliliği, likidite, vergi öncesi karlılık ve özkaynak karlılığı oranları kullanılmıştır.

Çalışmada her iki yöntemin uygulanmasının ve tahminlemede kullanımının gösterilmesi amaçlanmış olduğundan değişken seçiminde önsel bilgilerden yararlanılmıştır. Seçilen değişkenlerin Bankacılık Düzenleme ve Denetleme Kurulunun yayınladığı Finansal Piyasalar Raporunda bankacılık sektörüne ilişkin finansal sağlamlık göstergeleri arasında yer almaları nedeniyle bu dört bağımsız değişkenin kullanılması uygun bulunmuştur. Analize başlamadan önce finansal sağlamlık göstergeleri arasında bulunan belirli rasyolar denenmiş ve anlamlı bir modele ulaşılması seçilen değişkenlerle sağlanmıştır. Bu kapsamda, bankaların sektör paylarının tahmin edilmesi için çalışmanın son bölümünde öncelikle bulanık doğrusal regresyon analizi ile daha sonra lojistik regresyon analizi ile uygulama yapılmış ve elde edilen sonuçlar yorumlanmıştır.

Modellerin oluşturulmasında 45 banka için 2008 yılına ait gerçek veriler kullanılmıştır. Çalışmaya konu olan iki analiz yöntemi kullanılarak oluşturulan bu modeller ile ileriye dönük tahmin yapabilmek için 2009 yılına ait gerçekleşen veriler kullanılarak 2009 yılı için tahminde bulunulmuş ve gerçeğe ne kadar yakın sonuçlar elde edildiği gözlemlenmiştir. Burada amaç bankacılık sektörü için sektör paylarını tahminlemek adına bir model belirmesinden ziyade, çalışmada yer alan analiz yöntemlerinin tahminleme sürecinde kullanımına yönelik bir uygulama yapmaktır.

Buna göre, banka sektör paylarının tahminlenmesine ilişkin yapılan uygulamada öncelikle bulanık regresyon analizi uygulanarak model oluşturulmuştur. Bulanık regresyon analizinde bağımlı değişkeni etkilediği düşünülen değişkenlerin hepsi analize dahil edilebilmektedir ve değişkenlerde aranan tek kısıt doğrusal olmalarıdır, bu nedenle analizde yer alan veriler doğal logaritmaları alınarak doğrusallaştırılmıştır. Modelde yer alan değişkenlere ait değerler Tablo 3'te doğrusallaştırılmış değerler ise EK 1'de gösterilmiştir. Bulanık regresyon modelinde, her değişken için sabit değerle birlikte merkezi(a_i) ve yayılım(c_i) değerleri olarak toplam 10 adet parametrenin tahmin edilmesi gerekmektedir. Bunun için, 45 bankaya ait değerlerle doğrusal programlama tekniği için 90 adet kısıt oluşturulmuştur ve bulanıklık seviyesi *h* değeri 0,5 alınmıştır. Amaç denkleminde ise modelin toplam yayılımının yani bulanıklığının minimize edilmesi amaçlanmıştır. Oluşturulan doğrusal programlama modeli POM-QM for Windows programı yardımıyla çözülerek değişkenlere ait merkezi ve yayılım değerleri elde edilmiştir. Bulanık regresyon tahmin aralığının belirlenmesi için elde edilen merkezi ve yayılım değerleri yardımıyla alt ve üst sınırların regresyon modelleri oluşturulmuştur.

Bulanık regresyon analizinde tahmin yapabilmek için bir aralık sunan, alt ve üst regresyon modelleri ile 2009 yılına ilişkin tahminde bulunmak amacıyla, modelde 2009 yılına ait bağımsız değişkenlerin verileri kullanılarak bir tahmin aralığı elde edilmiştir. Burada kullanılan veriler daha önce belirtildiği gibi doğal logaritmaları alınmış değerlerdir. Bağımlı değişken sektör payları için tahminlenen alt ve üst sınırlara ait sonuçlar ve gerçekleşen Y_i sektör payı değerleri Tablo 5'te gösterilmiştir. Bu sonuçlara göre gerçekleşen bağımlı değişken değerlerinin tahmin edilen aralık içersinde yer alıp almadıkları incelenmiştir. Buna göre Tablo 5'te doğru aralıkta tahminlenenlere 1 değeri, bu aralık içersinde yer almayanlara ise 0 değeri atanmıştır. Bulanık regresyonun, bu aralıkları ne oranda doğru tahmin ettiği değerlendirildiğinde 45 banka arasından 13 bankanın sektör paylarının tahminlenen alt ve üst sınırlar arasında bulunduğu tespit edilmiştir. Dolayısıyla, bu uygulama için yapılan tahmin örneğinde bulanık regresyon analizinin tahmin yeteneğinin zayıf kaldığı sonucuna varılmıştır.

Uygulama kısmının devamında sektör paylarının tahminine yönelik lojistik regresyon analizinin uygulanması anlatılmıştır. Lojistik regresyonda da bulanık regresyonda kullanılan aynı değişkenler yer almıştır. Lojistik regresyon analizinde bağımlı değişkenin iki düzeyli olduğu durum incelendiği için bağımlı değişkeni temsil eden sektör payları iki grupta sınıflandırılmıştır. Sektör payları büyüklüğü olarak ilk on banka arasına girip girmeme durumuna göre bağımlı değişken 1 ve 0 değerlerini almıştır. Öncelikle 2008 yılına ait 45 bankanın verileri kullanılarak lojistik regresyon modeli oluşturulmuştur. Analizde SPPS programı kullanılarak elde edilen sonuçlarda Tablo 6'da gösterilen sınıflandırma tablosu değerlendirildiğinde, modelin ilk on banka arasına girenleri doğru sınıflandırma oranı %94,3 ilk on banka arasına giremeyenleri doğru sınıflandırma oranı ise %90 çıkmıştır. Toplam orana bakıldığında bunun %93,3 oranda doğru tahmin etme oranı ile sınıflandırma gücünün başarılı olduğu belirlenmiştir.

Modelin katsayılarının genel testine ilişkin değerlerin yer aldığı Tablo 7 değerlendirildiğinde anlamlılık değerinin 0,00 çıkmış olduğu görülmüş ve bu değer 0,05'ten küçük olduğu için modeldeki katsayılar genel olarak anlamlı bulunmuştur. Modelde yer alan değişkenlere ait verilerin gösterildiği Tablo 8'de yer alan sonuçlara göre, değişkenlerin anlamlıkları değerlendirilmiş ve lojistik modele alınacak değişkenlere karar verilmiştir. Tablo 8'de, bağımsız değişkenlere ait, katsayılar, standart hatalar, Wald istatistiği, serbestlik dereceleri, anlamlılık değerleri ve exp(β) değerleri gösterilmiştir. Bu sonuçlar değerlendirildiğinde; x_1, x_3 ve x_4 değişkenlerinin anlamlı olduklarına ve lojistik modelde yer almalarına karar verilmiştir. Buna göre, lojistik modelde değişime en çok etki eden -2,902 katsayı(β) değeriyle x_4 değişkeni olmuştur. Katsayının negatif olması bağımlı değişkenin logaritmik değerini negatif yönde etkilediğini ifade etmektedir. Oluşturulan modelin uyum iyiliğinin belirlenmesi için Tablo 9'da gösterilen istatistikler değerlendirilmiştir ve Cox&Snell R^2 istatistiği modelin %51 oranında, Nagelkerke R^2 istatistiği ise %78 oranında uyumlu olduğunu göstermiştir. Bu durumda modelin bağımlı değişkeni açıklayıcı gücünün iyi seviyede olduğu sonucuna varılmıştır.

Çalışmanın devamında, oluşturulan lojistik regresyon modeli ile tahmin yapabilmek amacıyla yine 2009 yılına ait bağımsız değişken verileri kullanılarak bağımlı değişken sektör payları için bir tahminde bulunulmuştur. Bu tahmine ilişkin elde edilen lojistik regresyon analizi sonuçları Tablo 10'da gösterilmiş ve sonuçlar yorumlanmıştır. Bunun için 45 bankaya ait ln(odds) ve p değerleri hesaplanmıştır. Elde edilen p olasılıkları bağımlı değişken için tahminlenen ilk on banka arasına girme olasılığını ifade etmektedir. Bu sonuçlara göre, p olasılıklarının doğru tahminler verdiği durumlara 1 değeri, yanlış tahminlediği durumlara ise 0 değeri verilmiştir. Bu durumda, lojistik regresyon tahminlerinin sadece 3 bankanın sınıflandırmasında yanıldığı gözlemlenmiştir. İlk on banka arasında yer alması gereken bir bankaya düşük olasılık sonucu çıkmış, ilk on arasında yer almayan iki bankaya da yüksek olasılık sonucu çıkmıştır. Ancak lojistik regresyondaki yanılma payının çok düşük bir seviyede olduğu anlaşılmaktadır. 45 gözlem içersinden sadece 3 tanesinde yanılma görülmesi, oluşturulan lojistik regresyon modeli ile yapılan tahminin oldukça iyi sonuçlar verdiğini kanıtlamıştır. Dolayısıyla bu uygulama için modelin tahminleme başarısı yüksek bulunmuştur.

Lojistik regresyon analizi, bulanık regresyon analizinden farklı olarak sınıflama yapmaya yaramakta ve bağımlı değişkene ait istenen durumun gerçekleşme olasılığını belirlemektedir, bulanık regresyon ise bağımlı değişken değerine ilişkin bir alt ve üst sınır sunarak, değerin bu aralıkta gerçekleşmesini beklemektedir. Ancak bu uygulamada sektör paylarının değerine ilişkin bulanık regresyonun sunduğu aralıkların gerçekleşen değerlerden çoğunlukla saptığı tespit edilmiştir. Bunun nedeni, kısıt sayısı arttıkça modeldeki belirsizliğin de artmasından dolayı bulanık regresyon tahminlerinin hata paylarının artması olarak düşünülmektedir. Bir başka neden de kullanılan bağımsız değişkenlerin, bağımlı değişkeni açıklamakta yetersiz kalmaları olabilir. Ancak, lojistik regresyon analizinde aynı değişkenler modelde kullanılarak bankaların ilk on arasında yer almaları durumu oldukça yüksek bir başarıyla belirlenebilmiş ve modelin tahmin gücünün yüksek olduğu gözlenmiştir. Bulanık regresyonda tüm değişkenler modele alınabilmişken, lojistik regresyonda yapılan anlamlılık testleri sonucu x_3 değişkeni modelde yer almamıştır. Her iki regresyon yönteminde de bağımlı değişken değerinde en büyük değişimi sağlayan bağımsız değişken katsayıları ele alındığında x_4 olarak tespit edilmiştir.

Sonuçta bu çalışmaya konu olan yöntemler ile yapılan örnek uygulama, oluşturulan modellerin tahminleme yeteneğini ortaya koymaya yöneliktir. Tahminler geleceğe ilişkin belirsizliklerle yapıldığından mükemmel olmaları çok zordur. Her zaman hata vardır ama yapılan uygulamalarda amaç bu hataları en aza indirgemek ve geleceğe katkı sağlamaktır. Bu çalışmada iki farklı analiz yöntemi kullanılarak yapılan tahminde elde edilen sonuçlar karşılaştırıldığında, lojistik regresyon analizinin tahminlemede daha başarılı olduğu tespit edilmiştir. Bulanık regresyon analizin uygulamaya konu olan sektör paylarını tahminlemede yetersiz kaldığı sonucuna varılmıştır.

İşletmelerin tahminleme sürecinde yaptıkları analizlerde, kullanılan veri grubunun genel yapısına göre farklı tahmin yöntemlerini denemelerinde fayda vardır.

Bu çalışmada da, farklı tahmin yöntemlerinden ikisi kullanılarak yapılan uygulama ile tahminde bulunmak ve gerçekleşen sonuçlarla karşılaştırarak uygun yöntemin belirlenmesi esas alınmıştır. Bu bağlamda çalışmanın, işletmelerin tahminleme sürecinde yapılacak analizlerin kullanımına ve uygulanmasına dair katkı sağlaması hedeflenmiştir. Dolayısıyla, işletmelere tahminde bulunurken veri grubunun yapısına göre farklı analiz yöntemlerini denemeleri, elde edilen sonuçları kontrol ederek ve mevcut veriler ile kıyaslayarak bu yöntemlerden işletme için uygun olanını seçmeleri önerilebilir. Uygun tahmin yönteminin kullanılması için mevcut veriler ile modeller oluşturulmalı ve anlamlı farklılıkların bulunup bulunmamasına göre firma kendi yapısına uygun tahmin yöntemini seçerek analize başlamalıdır. Özellikle kategorik veriler ile çalışıldığı durumlarda ve sonuçların oran olarak tahminlenmek istenildiği uygulamalarda lojistik regresyon analizinin kullanılması önerilmektedir. Klasik yöntemlerin sonuç vermediği, bağımsız değişken sayısının fazla olduğu ve bağımlı değişkenin değerini etkilediği bilinen değişkenlerin modele dahil edilmek istendiği durumlarda ise bulanık regresyon analizi kullanılabilir.

KAYNAKÇA

Agresti, A. (2002), *Categorical Data Analysis*, Second Edition, John Wiley&Sons Inc, USA

Aktaş, H. ve Kargın, M. (2007), Türk Bankacılık Sektöründeki Yabancı ve Ulusal Bankaların Finansal Oranlar Açısından Karşılaştırılması, *Celal Bayar Üniversitesi Yönetim ve Ekonomi Dergisi*, Cilt:14, Sayı:2, s.31-45

Allison, P. D. (2001), *Logistic Regression Using the SAS System:Theory and Application,* SAS Institute Inc, USA

Altaş, D. ve Giray, S. (2005), Mali Başarısızlığın Çok Değişkenli İstatistiksel yöntemlerle Belirlenmesi: Tekstil Sektörü Örneği, *Marmara Üniversitesi Sosyal Bilimler Dergisi*, Sayı 2, s.13-28

Altaş, İ.H. (1999), Bulanık Mantık: Bulanıklık Kavramı, *Enerji, Elektrik, Elektromekanik-3e*, Sayı 62, s.80-85, BileşimYayıncılık A.Ş., İstanbul

Bircan, H. (2004), Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama, *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, Sayı:2, s.185-208

Bircan, H., Coşkun, A., Coşkun, S., Kartal, M. (2004), Lojistik Regresyon Analizinin İncelenmesi ve Diş Hekimliğinde Bir Uygulaması, *Cumhuriyet Üniversitesi Diş Hekimliği Fakültesi Dergisi*, Cilt:7, Sayı:1, s.41-50

Bozdağ, N. ve Türe, H. (2008), Bulanık Doğrusal Programlama ve İMKB Üzerine bir Uygulama, *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi 10/1*, s.1-18

Bumin, M. (2009), Türk Bankacılık Sektörünün Karlılık Analizi: 2002-2008, *Maliye Finans Yazıları*, Bankacılık Düzenleme ve Denetleme Kurumu, Sayı:84, s.39-60

Cengiz, E. (2007), Bireylerin Kredi Kartını Değiştirme Tutumları, *Fırat Üniversitesi Sosyal Bilimler Dergisi*, Cilt 19, Sayı 2, s.179-196

Chang, Y.H.O. ve Ayyub, B.M. (2001), Fuzzy Regression Methods – A Comparative Assessment, *Fuzzy Sets and Systems 119*, s.187-203

Collett, D. (2002), *Modelling Binary Data*, Second Edition, Texts in Statistical Science, Chapman and Hall/CRC, USA

Çağman, N. (2006), Bulanık Mantık, *Bilim ve Teknik*, Sayı:463, s.50-51

Diamond, P. ve Körner, R. (1997), Extended Fuzzy Linear Models and Least Squares Estimates, *Computers Math Applications*, Vol.33, No.9, s.15-32

Dong, G., Lai, K. K. ve Yen, J. (2010), Credit Scorecard Based on Logistic Regression with Random Coefficients, *Procedia Computer Science 1*, s.2457-2462

D'Urso, P. ve Gastaldi, T. (2000), A Least Squares Approach to Fuzzy Linear Regression Analysis, *Computational Statistics&Data Analysis 34*, s.427,440

Ege, İ. ve Bayraktaroğlu, A. (2009), İMKB Şirketlerinin Hisse Senedi Getiri Başarılarının Lojistik Regresyon Tekniği ile Analizi, *ZKÜ Sosyal Bilimler Dergisi*, Cilt 5, Sayı 10, s.139-158

Girginer, N. ve Cankuş, B. (2008), Tramvay Yolcu Memnuniyetinin Lojistik Regresyon Analiziyle Ölçülmesi: Estram Örneği, *Celal Bayar Üniversitesi Yönetim ve Ekonomi Dergisi*, Cilt:15, Sayı:1, s.181193

He, Y. Q., Chan, L. K. ve Wu, M. L. (2007), Balancing Productivity and Costumer Satisfaction for Profitability: Statistical and Fuzzy Regression Analysis, *European Journal of Operations Research 176*, s.252-263

Hojati, M., Bector C.R. ve Smimou K. (2004), A Simple Method for Computation of Fuzzy Linear Regression, *European Journal of Operational Research 166*, s.172-184

Hosmer, D. W. ve Lemeshow, S. (2000), *Applied Logistic Regression*, Second Edition, John Wiley&Sons Inc, USA

Hout, A., Heijden, P.G.M. ve Gilchrist, R. (2007), The Logistic Regression Model with Response Variables Subject to Randomized Response, *Computational Statistics&Data Analysis 51*, s.6060-6069

Kao, C. ve Chyu, C. (2002), A Fuzzy Linear Regression Model with Better Explanatory Power, *Fuzzy Sets and Systems 126*, s. 401-409

Kao, C. ve Chyu C. L. (2003), Least Squares Estimates in Fuzzy Regression Analysis, *European Journal of Operations Research148*, s.426-435

Kaşko, Y. (2007), *Çoklu Bağlantı Durumunda İkili (Binary) Lojistik Regresyon Modelinde 1. Tip Hata ve Testin Gücü*, Yüksek Lisans Tezi, Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara

Kaya Türker, Y. (2002), Türk Bankacılık Sektöründe Karlılığın Belirleyicileri 1997-2000, *Mali Sektör Politikaları Dairesi Çalışma Raporu*, No:2002/1, s.1-16

Keskin Benli, Y. (2005), Bankalarda Mali Başarısızlığın Öngörülmesi Lojistik Regresyon ve Yapay Sinir Ağı Karşılaştırması, *Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi*, Sayı 16, s.31-46

Kim, B. ve Bishu, R.R. (1998), Evaluation of Fuzzy Linear Regression Models by Comparing Membership Functions, *Fuzzy Set and Systems 100*, s.342-352

Kim, H. ve Gu, Z. (2009), Financial Features of Dividend-Paying Firms in the Hospitality Industry: A Logistic Regression Analysis, *International Journal of Hospitality Management 28*, s.359-366

Kim, K. J. ve Chen, H. R. (1997), A Comparison of Fuzzy and Nonparametric Linear Regression, *Computers Ops Research*, Vol. 24, No. 6, s. 505-519

Kleinbaum, D. ve Klein, M. (2002), *Logistic Regression A Self Learning Text*, Second Edition, Springer-Verlag New York Inc, USA

Long, J. S. (1997), *Regression Models For Categorical and Limited Dependent Variables*, Sage Publications, USA

Lu, J. ve Wang, R. (2009), An Enhanced Fuzzy Linear Regression Model with More Flexible Spreads, *Fuzzy Sets and Systems 160*, s.2505-2523

Menard, S. (2002), *Applied Logistic Regression Analysis*, Second Edition, Sage Publications, USA

Ming, M., Friedman, M. ve Kandel, A. (1997), General Fuzzy Least Squares, *Fuzzy Sets and Systems 88*, s.107-118

Modarres, M., Nasrabadi, E. ve Nasrabadi, M.M. (2005), Fuzzy Linear Models with Least Squares Errors, *Applied Mathematics and Computations 163,* s.977-989

Moskowitz, H. ve Kim, K.J. (1993), On Assesing The h Value in Fuzzy Linear Regression, *Fuzzy Sets and Systems*, Volume 58, s.303-327

Nasrabadi, M.M. ve Nasrabadi, E. (2004), A Mathematical-Programming Approach to Fuzzy Linear Regression Analysis, *Applied Mathematics and Computations 155*, s.873-881

Özdamar, K. (2002), *Paket Programlar ile İstatistiksel Veri Analizi*, Kaan Kitabevi, Eskişehir

Özdemir, A. (2008), *Yönetim Biliminde İleri Araştırma Yöntemleri ve Uygulamalar*, Beta Basım Yayım Dağıtım A.Ş., İstanbul

Özer, G. G. (2006), *Türk Bankacılık Sektöründe Lojistik Regresyon Yöntemi Kullanılarak Kriz Olasılığının İncelenmesine İlişkin Bir Araştırma*, Yüksek Lisans Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul

Roy, S. S. ve Guria, S. (2008), Diagnostics in Logistic Regression Models, *Journal of the Korean Statistical Society 37*, s.89-94

Ryan, T.P. (1997), *Modern Regression Analysis*, John Wiley&Sons Inc, USA

Sanchez, J. A. (2006), Calculating Insurance Claim Reserves with Fuzzy Regression, *Fuzzy Sets and Systems 157*, s.3091-3108

Sanchez, J. A. ve Gomez, A. T. (2004), Estmating a Fuzzy Term Structure of Interest Rates Using Fuzzy Regression Techniques, *European Journal of Operations Research 154*, s.804-818

Stahl, C. (2006), A Strong Consistent Last Squares Estimator in a Fuzzy Linear Regression Model with Fuzzy Parameters and Fuzzy Dependent Variables, *Fuzzy Sets and Systems 157*, s.2593-2607

Stölzle, T. R., Koissi M. C. ve Shapiro A. F. (2010), Detecting Fuzzy Relationships in Regression Models: The Case of Insurer Solvency Surveillance in Germany, *Insurance: Mathematics and Economics 46*, s.554-567

Şahin, M. (2005), *Yönetim Bilgi Sistemi*, TC Anadolu Üniversitesi, Eskişehir

Taç, H. K. ve Budak, F. (2007), İşletmelerin TS-ISO 14000 Standartlarını Uygulama Kararlarına Etki Eden Faktörlerin Belirlenmesi, *Çukurova Üniversitesi Çevre Mühendisliği Fakültesi Dergisi*, Cilt 22, Sayı 1, s.379-389

Tanaka, H., Uejima, S. ve Asai K. (1982), Linear Regression Analysis with Fuzzy Model, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-12, No.6, s.903-907

Tanaka, K. (1997), *An Introduction to Fuzzy Logic for Practical Applications*, Rassel Inc, New York

Tezcan, B. (2006), *Lojistik Regresyon Analizi ve Sigortacılık Sektöründe Bir Uygulama*, Yüksek Lisans Tezi, Marmara Üniversitesi Bankacılık ve Sigortacılık Enstitüsü, İstanbul

Ulupınar, S.D. (2007), *2001 Krizi Öncesi ve Sonrası Türk Ticari Bankalarının Karlılıklarının Lojistik Regresyon Analizi ile İncelenmesi*, Yüksek Lisans Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul

Ünsal, A. ve Güler, H. (2005), Türk Bankacılık Sektörünün Lojistik Regresyon ve Diskriminant Analizi ile İncelenmesi, http://www.ekonometridernegi.org/bildiriler/o14s2.pdf, Erişim: 30.05.2010

Ürük, E. (2007), *İstatistiksel Uygulamalarda Lojistik Regresyon Analizi*, Yüksek Lisans Tezi, Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul

Wang, H.F. ve Tsaur, R.C. (2000a), Insight of a Fuzzy Regression Model, *Fuzzy Sets and Systems 112*, s. 355-369

Wang, H.F. ve Tsaur, R.C. (2000b), Resolution of a Fuzzy Regresion Model, *European Journal of Operational Research 126*, s.637-650

Wang, L., Jiao, L., Shi, G., Li, X. ve Liu, J. (2006), *Fuzzy Systems and Knowledge Discovery,* Third International Conference FSKD Proceedings, Springer-Verlag Berlin Heidelberg, Germany

Wu, B. ve Tseng, N. F. (2002), A New Approach to Fuzzy Regression Models with Application to Business Cycle Analysis, *Fuzzy Sets and Systems 130*, s.33-42

Yücel, İ. L. (2005), *Bulanık Regresyon: Türkiye'de 1980-2004 Döneminde Kayıt Dışı Ekonominin Bulanık Yöntemlerle Tahminine İlişkin Bir Uygulama*, Yüksek Lisans Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul

Zadeh, L.A. (1965), Fuzzy Sets, *Information and Control*, Volume 8, s.338-353, Academic Press, New York

Zadeh, L. A. ve Kacprzyk, J. (1999), *Computing with Words in Information / Intelligent Sysytems:Foundations*, Physica-Verlag Heidelberg, New York

Türkiye Bankalar Birliği, İstatistiki Raporlar, http://www.tbb.org.tr/tr/Banka_ve_Sektor_Bilgileri/Tum_Raporlar.aspx, Erişim: 30.05.2010

Bankacılık Düzenleme ve Denetleme Kurulu, Finansal Piyasalar Raporları, http://www.bddk.org.tr/websitesi/turkce/Raporlar/Finansal_Piyasalar_Raporlari/Finansal_Piyasalar_Raporlari.aspx, Erişim:30.05.2010

www.deu.edu.tr/userweb/k.yaralioglu/dosyalar/bul_man.doc, Erişim:10.03.2010

www.yapay-zeka.org/modules/icontent/index.php?page=33, Erişim:10.03.2010

EKLER

EK 1

Modelde Kullanılan Değişkenlerin Doğal Logaritması Alınmış Değerleri

	Ln(Y)	Ln(x_1)	Ln(x_2)	Ln(x_3)	Ln(x_4)
1	2,695	3,001	3,086	3,367	2,393
2	1,974	2,674	2,701	3,165	2,008
3	2,001	2,660	3,414	2,588	2,721
4	-4,605	5,228	4,535	1,887	3,842
5	2,493	2,901	3,063	2,721	2,370
6	-0,693	2,646	3,364	2,646	2,671
7	-0,693	2,918	2,879	2,839	2,186
8	0,095	2,688	2,901	2,695	2,208
9	-0,916	2,885	3,292	1,030	2,599
10	-2,303	3,541	4,253	1,932	3,560
11	0,742	2,874	3,453	2,442	2,760
12	2,534	2,779	3,463	2,918	2,769
13	2,625	2,721	3,721	2,773	3,028
14	2,197	2,754	2,573	2,721	1,879
15	0,000	4,176	4,444	2,501	3,751
16	-2,303	3,529	3,434	0,095	2,741
17	-0,223	2,885	3,945	2,322	3,252
18	0,993	2,845	3,199	2,617	2,506
19	-2,303	3,706	4,043	2,313	3,350
20	-0,693	2,885	3,558	1,504	2,865
21	1,335	2,773	2,839	2,549	2,146
22	0,531	2,695	3,472	2,079	2,779
23	0,742	2,734	3,384	2,398	2,691
24	0,833	2,625	3,314	2,163	2,621
25	-1,609	3,105	3,122	0,642	2,429
26	-2,303	3,086	3,453	-0,916	2,760
27	-1,609	3,538	4,397	2,827	3,704
28	-3,219	3,944	3,714	3,235	3,020
29	-4,605	4,464	4,278	2,510	3,585
30	-3,507	4,802	4,577	2,653	3,884
31	-2,303	3,144	4,317	3,859	3,624
32	-2,303	4,274	4,546	2,809	3,853
33	0,000	4,260	2,996	1,740	2,303
34	-0,357	4,653	2,610	2,534	1,917
35	-2,303	4,372	3,484	2,028	2,791
36	-3,507	3,572	3,752	0,000	3,059
37	-3,912	4,171	4,142	2,262	3,448
38	-4,605	3,896	3,500	2,760	2,806
39	-1,609	4,096	4,516	2,809	3,823
40	-4,605	3,100	3,500	1,194	2,806
41	-0,105	3,049	3,440	2,766	2,747
42	-1,609	3,478	3,140	1,974	2,447
43	-4,605	2,632	4,258	0,095	3,565
44	-4,605	4,690	4,377	1,163	3,684
45	-2,303	4,363	3,215	2,766	2,522

EK 2

Tahminde Kullanılan 2009 Yılına Ait Veriler

	Y_i	x_1	x_2	x_3	x_4
1	15,59	23,22	32,72	33,91	3,55
2	7,60	16,03	15,45	28,32	3,33
3	8,11	15,42	37,34	16,95	2,38
4	0,01	185,88	93,58	1,87	1,58
5	11,94	22,50	39,01	19,21	3,46
6	0,45	12,78	8,59	14,16	2,12
7	0,48	19,99	16,19	19,06	3,95
8	1,12	16,30	25,23	12,20	2,22
9	0,27	20,78	20,14	2,79	0,76
10	0,13	28,84	73,15	0,77	0,11
11	1,89	17,70	29,60	12,74	1,71
12	13,21	21,20	42,87	22,25	3,58
13	14,18	18,31	38,57	17,58	2,60
14	8,08	17,78	14,33	16,39	2,65
15	0,10	45,02	85,50	8,61	8,77
16	0,12	31,61	31,67	8,85	3,27
17	0,61	19,94	54,18	8,91	1,97
18	2,66	19,02	24,07	20,22	3,14
19	0,24	49,69	87,65	25,64	7,03
20	0,48	25,99	33,59	4,97	0,73
21	3,67	17,99	36,08	17,92	2,01
22	1,41	15,96	30,67	5,67	1,20
23	1,73	17,32	31,58	9,88	2,18
24	1,91	15,65	24,18	9,56	1,78
25	0,13	19,17	28,68	-14,58	-1,97
26	0,14	19,32	31,00	1,09	0,27
27	0,15	53,48	80,80	19,54	7,30
28	0,05	49,14	40,12	24,82	5,79
29	0,01	94,34	80,59	8,57	5,89
30	0,02	116,38	94,85	13,22	14,35
31	0,05	10,23	52,52	-72,75	-4,44
32	0,07	84,07	91,27	19,23	6,64
33	1,00	66,44	20,17	5,54	5,93
34	0,81	125,72	34,52	9,37	5,28
35	0,16	70,05	25,23	5,13	2,44
36	0,07	21,29	40,08	5,45	2,02
37	0,01	55,38	1,83	8,55	9,18
38	0,01	42,80	13,46	4,00	2,89
39	0,19	51,76	93,25	13,41	3,06
40	0,02	24,72	36,13	4,36	1,73
41	0,86	24,91	34,01	16,80	3,05
42	0,20	28,97	12,32	5,99	2,14
43	0,01	19,84	80,07	18,00	17,91
44	0,03	72,09	76,89	16,12	4,76
45	0,00	60,24	7,34	-19,09	-12,32

Tahminde Kullanılan 2009 Yılına Ait Verilerin Ln Değerleri

	$Ln(Y_i)$	$Ln(x_1)$	$Ln(x_2)$	$Ln(x_3)$	$Ln(x_4)$
1	2,75	3,14	3,49	3,52	1,27
2	2,03	2,77	2,74	3,34	1,20
3	2,09	2,74	3,62	2,83	0,87
4	-5,06	5,23	4,54	0,63	0,46
5	2,48	3,11	3,66	2,96	1,24
6	-0,79	2,55	2,15	2,65	0,75
7	-0,73	3,00	2,78	2,95	1,37
8	0,11	2,79	3,23	2,50	0,80
9	-1,32	3,03	3,00	1,03	-0,28
10	-2,05	3,36	4,29	-0,26	-2,25
11	0,63	2,87	3,39	2,54	0,53
12	2,58	3,05	3,76	3,10	1,28
13	2,65	2,91	3,65	2,87	0,96
14	2,09	2,88	2,66	2,80	0,98
15	-2,29	3,81	4,45	2,15	2,17
16	-2,11	3,45	3,46	2,18	1,19
17	-0,49	2,99	3,99	2,19	0,68
18	0,98	2,95	3,18	3,01	1,14
19	-1,43	3,91	4,47	3,24	1,95
20	-0,73	3,26	3,51	1,60	-0,32
21	1,30	2,89	3,59	2,89	0,70
22	0,34	2,77	3,42	1,74	0,18
23	0,55	2,85	3,45	2,29	0,78
24	0,65	2,75	3,19	2,26	0,58
25	-2,01	2,95	3,36	2,68	0,68
26	-1,94	2,96	3,43	0,09	-1,31
27	-1,90	3,98	4,39	2,97	1,99
28	-3,01	3,89	3,69	3,21	1,76
29	-4,69	4,55	4,39	2,15	1,77
30	-3,69	4,76	4,55	2,58	2,66
31	-3,10	2,33	3,96	4,29	1,49
32	-2,66	4,43	4,51	2,96	1,89
33	0,00	4,20	3,00	1,71	1,78
34	-0,21	4,83	3,54	2,24	1,66
35	-1,83	4,25	3,23	1,63	0,89
36	-2,63	3,06	3,69	1,70	0,70
37	-4,44	4,01	0,60	2,15	2,22
38	-4,35	3,76	2,60	1,39	1,06
39	-1,68	3,95	4,54	2,60	1,12
40	-3,82	3,21	3,59	1,47	0,55
41	-0,15	3,22	3,53	2,82	1,11
42	-1,61	3,37	2,51	1,79	0,76
43	-4,82	2,99	4,38	2,89	2,89
44	-3,65	4,28	4,34	2,78	1,56
45	-5,76	4,10	1,99	2,95	2,51

EK 3

İterasyon Geçmişi

İterasyon		-2 Log likelihood	Katsayılar				
			Sabit	x_1	x_2	x_3	x_4
Step 1	1	35,474	-,604	-,009	-,011	,058	-,078
	2	26,955	-,914	-,016	-,025	,146	-,185
	3	23,106	-,845	-,034	-,037	,212	-,308
	4	20,376	-,644	-,070	-,040	,293	-,574
	5	18,403	-,572	-,127	-,035	,421	-1,055
	6	16,910	,870	-,230	-,039	,528	-1,625
	7	15,717	4,795	-,413	-,069	,643	-2,506
	8	15,608	6,225	-,478	-,088	,699	-2,849
	9	15,606	6,403	-,486	-,091	,709	-2,901
	10	15,606	6,406	-,486	-,091	,709	-2,902
	11	15,606	6,406	-,486	-,091	,709	-2,902

a. Metod: Enter

b. Sabit modelde yer almaktadır.

c. İlk -2 Loglikelihood değeri: 47,674

d. 11 iterasyon sayısına ulaşıldığında iterasyona son verilmiştir, çünkü parametre tahminleri 0,001’den küçük bir oranda değişmektedir.