SİVAS CUMHURİYET ÜNİVERSİTESİ
Sosyal Bilimler Enstitüsü
Ekonometri Anabilim Dalı

PARAMETRİK OLMAYAN REGRESYON ANALİZİNDE BOOTSTRAP YÖNTEMİ

Yüksek Lisans Tezi

Nilüfer Betül AKMAN

Sivas
Aralık-2018

SİVAS CUMHURİYET ÜNİVERSİTESİ
Sosyal Bilimler Enstitüsü
Ekonometri Anabilim Dalı

PARAMETRİK OLMAYAN REGRESYON ANALİZİNDE BOOTSTRAP YÖNTEMİ

Yüksek Lisans Tezi

Nilüfer Betül AKMAN

Tez Danışmanı
Doç. Dr. Necati Alp ERİLLİ

Sivas
Aralık-2018

KABUL VE ONAY

Üniversite	**: Sivas Cumhuriyet Üniversitesi**
Enstitü	**: Sosyal Bilimler Enstitüsü**
Ana Bilim Dalı	**: Ekonometri**
Bilim Dalı	**: Ekonometri**
Tezin Başlığı	**: Parametrik Olmayan Regresyon Analizinde Bootstrap Yöntemi**
Savunma Tarihi	**: 21/12/2018**
Danışmanı	**: Doç. Dr. Necati Alp Erilli**

	Unvanı-Adı Soyadı	**İmza**
Jüri Başkanı	: Dr. Öğretim Üyesi Şebnem ZORLUTUNA	
Üye	: Doç. Dr. Necati Alp ERİLLİ	
Üye	: Dr. Öğretim Üyesi Özge GÜNDOĞDU	

Oy Birliği ☒

Oy Çokluğu ☐

Nilüfer Betül AKMAN Tarafından hazırlanan "Parametrik Olmayan Regresyon Analizinde Bootstrap Yöntemi" başlıklı tez kabul edilmiştir. ../.../.....

Prof. Dr. Ahmet ŞENGÖNÜL
Enstitü Müdürü

ETİK İLKELERE UGUNLUK BEYANI

Cumhuriyet Üniversitesi Sosyal bilimler enstitüsü bünyesinde hazırladığım bu yüksek lisans tezinin bizzat tarafımdan ve kendi sözcüklerimle yazılmış orijinal bir çalışma olduğunu ve bu tezde;

1- Çeşitli yazarların çalışmalarından faydalandığımda bu çalışmaların ilgili bölümlerini doğru ve net biçimde göstererek yazarlara açık biçimde atıfta bulunduğumu;
2- Yazdığım metinlerin tamamı ya da sadece bir kısmı, daha önce herhangi bir yerde yayınlamamışsa bunu da açıkça ifade ederek gösterdiğimi;
3- Başkalarına ait alıntılanan tüm verileri (tablo, grafik, şekil vb. de dahil olmak üzere) atıflarla belirttiğimi;
4- Başka yazarların kendi kelimeleriyle alıntıladığım metinlerini, tırnak içerisinde veya farklı dizerek verdiğim yine başka yazarlara ait olup fakat kendi sözcüklerime ifade ettiğim hususları da istisnasız olarak kaynak göstererek belirttiğimi;

Beyan ve bu etik ilkeleri ihlal etmiş olmam halinde bütün sonuçlarına katlanacağımı kabul ederim.

25./12/2018

Nilüfer Betül AKMAN

NAkman

TEŞEKKÜR

Çalışmam süresince yapıcı, öğretici eleştirileriyle yol gösteren ve yardımlarını esirgemeyen danışman hocam Doç. Dr. Necati Alp ERİLLİ' ye,

Bu süreç boyunca öneri ve yardımlarına ne zaman ihtiyaç duysam yanımda olan kardeşim Döne' ye ve yakın arkadaşım Seçkin' e,

Son olarak hayatım boyunca iyi ve kötü her anımda yanımda olan, maddi ve manevi desteklerini benden esirgemeyen en değerlim anneme,

Sonsuz teşekkürler…

İÇİNDEKİLER

KISALTMALAR

EKK : En Küçük Kareler

EÇOK : En Çok Olabilirlik

K-NN : En Yakın Komşu

HKOK : Hata Karelerinin Ortalamalarının Karekökü

TABLOLAR LİSTESİ

ŞEKİLLER LİSTESİ

ÖZET

Regresyon analizi aralarında sebep sonuç ilişkisi olan iki veya daha fazla değişkenin aralarındaki ilişkiyi incelemek ve konuyla ilgili tahmin yapabilmek amacıyla oluşturulan matematiksel bir model ile ifade edilen istatistiksel analizdir. Basit doğrusal regresyon analizi bir bağımlı ve bir bağımsız değişken arasındaki fonksiyonel ilişkiyi incelerken, çoklu doğrusal regresyon analizi bir bağımlı ve birden çok bağımsız değişken arasındaki fonksiyonel ilişkiyi inceler. Regresyon analizi, parametrik regresyon, parametrik olmayan regresyon ve yarı parametrik regresyon olmak üzere üç ana başlık altında incelenebilir. Parametrik regresyon analizi bağımlı ve bağımsız değişkenler ile bu değişkenler arasındaki ortalama ilişkinin matematiksel bir fonksiyonla gösterilmesi ve bu fonksiyondaki parametre vektörlerinin açık olarak gösterilmesidir. Parametrik olmayan regresyon analizi bir bağımlı değişken ve bu değişken ile nasıl bir ilişki içerisinde olduğu bilinmeyen bir bağımsız değişkenden oluşmaktadır. Parametrik olmayan regresyon analizinin amacı parametreleri tahmin etmekten çok bilinmeyen yanıt fonksiyonunu tahmin etmektir. Yarı parametrik regresyon modeli ise bağımlı değişkenin bağımsız değişkenlerden bazıları ile aralarındaki ilişkinin belirlenebildiği fakat diğer bağımsız değişken veya değişkenlerle aralarındaki ilişkinin kolay olarak belirlenemediği bir modeldir.

Bootstrap yöntemi ise kısaca verilerin yeniden örnekleme mantığına dayanan bir yöntem olarak ifade edilebilir. Geleneksel parametrik sonucun dağılım varsayımları ve matematiksel analizden daha büyük oranda hesaplamayı içeren bootstrap, istatistiksel sonuca yaklaşım olarak kullanılmaktadır. Bootstrap yöntemi örneklem ortalaması, standart hatalarının hesaplanması ve güven aralıklarının belirlenmesini amaçlayarak geliştirilmiştir.

Bu tez çalışmasında; parametrik olmayan regresyon analizinde, yeniden örnekleme yöntemlerinden olan Bootstrap yöntemi birlikte kullanılmıştır. Yöntem sonuçları, farklı veri yapılarında EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerinden elde edilen sonuçlar ile karşılaştırılmış ve yorumlanmıştır. Ayrıca Bootstrap ile elde edilen farklı veri yapılarına, belirli oranda aykırı değerler eklenerek, uygulanan yöntemin gücü test edilmiştir. Sonuç olarak Bootstrap tekniği ile parametrik olmayan regresyon analizine alternatif bir yöntem sunulmuş, sonuçlar yorumlanarak karşılaştırmalar yapılmıştır.

Anahtar Kelimeler: Doğrusal Regresyon, Parametrik Olmayan Regresyon, Bootstrap Analizi

ABSTRACT

Regression analysis is a statistical analysis which is used to examine the relationship between two or more variables which have causative relationship between them and to make predictions on the subject. Simple linear regression analysis examines the functional relationship between a dependent and an independent variable. Multiple linear regression analysis examines the functional relationship between a dependent and multiple independent variables. Regression analysis can be examined under three main headings. Parametric regression, nonparametric regression and semi-parametric regression. The parametric regression analysis is a mathematical function of the dependent and independent variables and the relationship between these variables. Nonparametric regression analysis consists of a dependent variable and an independent variable that is not associated with this variable. The aim of the nonparametric regression analysis is to estimate the unknown response function rather than predict parameters. The semi-parametric regression model is a model in which the relationship between the dependent variable and some of the independent variables can be determined but the relationship between the other independent variables or variables cannot be easily determined.

The Bootstrap method can be briefly expressed as a method based on the re-sampling logic of the data. The bootstrap, which includes the calculation of the traditional parametric result and the calculation of a larger proportion than the mathematical analysis, is used as the approach to the statistical result. The Bootstrap method was developed with the aim of determining the sample mean, the calculation of standard errors and the determination of confidence intervals.

In this thesis; In the nonparametric regression analysis, Bootstrap method, which is one of the resampling methods, was used together. The results of the method were compared and interpreted with the results obtained from EKT, Theil-Sen, Mood-Brown, Optimum and Hodges-Lehman estimation models in different data structures. In addition, the power of the applied method has been tested by adding certain values to the different data structures obtained with Bootstrap. As a result, an alternative method for nonparametric regression analysis was presented with Bootstrap technique.

Keywords: Linear Regression, Nonparametric Regression, Bootstrap Analysis

GİRİŞ

Regresyon analizi; herhangi bir değişkenin (bağımlı değişken) bir veya birden fazla değişken ile (bağımsız – açıklayıcı değişken) arasındaki ilişkinin matematiksel bir fonksiyon şeklinde yazılmasıdır. Elde edilen bu fonksiyona ise regresyon denklemi adı verilmektedir (Orhunbilge 2000: 12). Regresyon analizinin temeli; ilk olarak Francis Galton tarafından 19. yüzyılın sonlarında atılmıştır. Galton yaptığı çalışmada; anne-babaların boyu ile çocuklarının boyları arasındaki ilişkiyi incelemiş ve kısa boylu anne-babaların çocuklarının boylarının kısa, uzun boylu anne-babaların çocuklarının boylarının uzun olmasına rağmen, çocuklarının boylarının anakitle boy ortalamasına doğru yaklaşma eğiliminde olduğunu görmüştür. Bu eğilimi "ortaya doğru çekilme = regression to mediocrity" olarak adlandırmıştır. (Galton 1886: 246-263). Galton'un çalışmaları bugün, değişkenler arasındaki istatistiki ilişkileri inceleyen "Regresyon Analizi'nin (Regression Analysis – Relationship Analysis)" başlangıcı olmuştur. Günümüzde regresyon analizi için parametrik ve parametrik olmayan olmak üzere iki türlü yaklaşım bulunmaktadır.

Bu tez çalışması; parametrik olmayan regresyon modellerine Bootstrap yöntemi uygulanarak parametrik regresyon modeli olan EKK ve diğer parametrik olmayan tahmin modelleri ile karşılaştırıp en iyi tahmin modelini bulabilmek amacıyla yapılmıştır.

Çalışmanın ikinci bölümünde regresyon analizi ve parametrik, parametrik olmayan ve yarı parametrik regresyon modelleri hakkında teorik bilgiler verilmiştir. Ayrıca parametrik olmayan regresyonun tanımı, özellikleri, avantajları, tahmin yöntemleri, parametrik olmayan regresyonda düzgünleştirme ve parametrik olmayan regresyon yöntemleri hakkında ayrıntılı bilgiler verilmiştir. Üçüncü bölümde bootstrap yöntemi ve çeşitleri teorik olarak anlatılmıştır. Dördüncü bölümde simülasyon ve gerçek hayat verilerine bootstarp yöntemi uygulayıp aykırı değer eklenilerek iki ayrı uygulama yapılmış ve yorumlanmıştır. Beşinci bölümde ise sonuçlar özetlenmiştir.

BÖLÜM 1

1. LİTERATÜR TARAMASI

1.1. Parametrik Olmayan Regresyon Analizi Literatür Taraması

Mood-Brown (1951), a ve b_i katsayılarını belirleyen ve kendi isimleri ile anılan bir yöntem geliştirmişlerdir. Parametrik olmayan basit doğrusal regresyon analizinde medyana göre parametre tahmininde regresyon doğrusunun eğimi için $H_0 : b = b_0$ hipotez testini $H_1 : b \neq b_0$ alternatifine karşı test etmek için incelemeler yapmışlardır.

Griffin (1962), parametrik olmayan metotları, popülasyon hakkında katı varsayımlar öngörmeyen testler olarak ifade etmiştir.

Hill (1962), konveksliğin (dışbükeyliğin) alternatifine karşı bir medyan regresyon eğrisinin doğrusallığının bir testini önermiştir. Doğrusallığın önerilen testi, Mood-Brown metodu vasıtasıyla (X_i, Y_i) noktalarının bir setinde (kümesinde) tahmin edilmiş bir doğru dayanağında tarif edilmiştir.

Adichie (1967), α ve β 'nın nokta tahminlerini hesaplamış ve bu tahminlerin etkinliklerini Mood-Brown medyan tahminler ile mukayese etmiş ve medyan tahminlerinin etkinliğinde bazı kayıpların olduğunu belirtmiştir. Ayrıca $H_0 : \alpha = 0$ ve $H_0 = \beta = 0$ hipotezleri için rank puan testlerinin bir sınıfını önermiştir.

Sen (1968), 2 veya daha fazla eğim parametresinin birbirine eşit olduğunu iddia eden sıfır hipotezlerini test eden bir sıra puanı yöntemini incelemiştir. Kendall Tau katsayısından esinlenerek b_i 'nin basit ve sağlam tahmincileri üzerinde çalışmıştır. Kendall Tau test ölçütü hesabının güç ve etkinliğini incelemiştir. Nokta tahmincisini $x_i \neq x_j$ ile noktaların $(y_j - y_i)/(x_j - x_i)$ eğim çiftleri popülasyonunun medyanı olarak tarif etmiştir. Sen, ileri sürdüğü tahmincilerin çeşitli özelliklerini incelemiş ve kendi ismini verdiği yöntemin EKK ve diğer parametrik olmayan tahminciler ile karşılaştırmasını yapmıştır.

Priestley ve Chao (1972), değişkenler arasındaki fonksiyonel ilişkinin elde edilmesi konusunda çalışmalar yapmışlardır ve elde edilecek fonksiyonun, düzgün olması gerektiğini belirtmişlerdir. Parametrik olmayan regresyon fonksiyonunun tahmini için gerekli temel bilgileri vererek tahminin ortalaması, varyansı üzerinde çalışmalar ve Kernel fonksiyonunun seçimi üzerinde açıklamalar yapmışlardır.

Cryer ve ark.(1972), her bir reel sayı x'in $[0,1]$ aralığında olduğunu farz ederek, ortalama $\mu(x)$ ve medyan $M(x)$'in parametrik olmayan tahmin edicilerini tanımlamış ve $M(.)$'nın monoton olan $\widehat{M}(.)$ tahmincisini tartışmıştır.

Potthoff (1974), iki basit regresyon doğrusunun paralel olup olmadığının bir parametrik olmayan testini incelemiş ve uygulanan parametrik olmayan test iki-örnek Wilcoxon testi ile benzer olarak geliştirilmiştir.

Benedetti (1977), Priestley ve Chao'nun önerdikleri parametrik olmayan tahminlerin tutarlılığını teorem yardımı ile ispatlamışlardır. Ayrıca Kernel seçimi, asimptotik normallik ile parametrik olmayan regresyon fonksiyonunun tahmininin genel yapısı üzerinde teoremler ve ispatlar vermişlerdir.

Maritz (1979), $v_i = {y_i}/{x_i}$ ve $u_i = {1}/{x_i}$ olması durumundan $v_i = \alpha u_i + \beta$ yapısında tekrar $y_i = \alpha + \beta x_i$ nin yazılmasıyla teşekkül eden bir alternatif tahminci önermiştir. Maritz u üzerine v'nin regresyonu ile α eğimini tahmin etmek için parametrik olmayan regresyon parametrelerini tahmin metodu olan Theil metodunu uygulamıştır.

Ibragimov ve Hasminskii (1980), regresyonun parametrik olmayan tahmini üzerine bir araştırma yapmışlardır.

Kildea (1981), en küçük kareler tahmincisinin bir doğal medyan benzerinin elde edilmesi ile parametrik olmayan regresyon parametrelerini tahmin metodu olan Mood-Brown tahmincisinin nasıl değiştiğini göstermiştir.

Rao ve Gore (1982), parametrik olmayan prosedürlerin basit adaptasyonlarıyla test edilebilen bir-örnekli, iki-örnekli ve birkaç-örnekli doğrusal regresyonlarda kesişim parametresi hakkındaki çeşitli hipotezleri açıklamışlardır.

Hussain ve Sprent (1983), çeşitli parametrik olmayan regresyon modelleri arasındaki karşılaştırmaları yapmışlardır. Medyan tahmincileri EKK'de uygun ağırlık yöntemlerini kullanarak, sapan gözlemlerin etkisinin öneminin azaltılmasının mümkün olacağını belirtmişlerdir. a ve b'nin tahminlerinde medyan hesabına dayalı Theil yöntemini incelemişlerdir.

Hussain ve Sprent (1983), bir regresyon doğrusu eğiminin parametrik olmayan tahmincilerinin sağlam olduğunu belirtmişler ve $x_j \neq x_i$ olmak üzere $(x_i, y_i), (x_j, y_j)$ nokta çiftleri ile bileşen doğrularla ilgili bütün eğim çiftleri dizisinin $b_{ij} = \frac{(y_j - y_i)}{(x_j - x_i)}$ şeklinde olduğunu ifade etmişlerdir. Ayrıca b_{ij} eğim çiftleri takımından (kümesinden) β'nın parametrik olmayan tahmincilerinin birkaçını da tanımlayarak bunların birbirinden farkını kısaca belirtmişlerdir.

Rice (1984), parametrik olmayan regresyon için bandwith parametre seçimi incelemiş ve bu çalışmasında Fourier seriler tahmin analizini yapmıştır.

Hardle ve Gasser (1984), Priestley-chao kernel tahmincisini ve Gasser-Müller kernel ağırlıklarını kullanarak sağlam (robust) parametrik olmayan fonksiyon uygunluğunu incelemişlerdir.

Silverman (1984), eğri tahmini ve parametrik olmayan regresyon için spline yumuşatma yaklaşımını incelemiştir ve spline yumuşatma ile kernel yumuşatma arasındaki ilişkiyi araştırmıştır.

Speckman (1985), $Y_i = f(x_i) + e_i, x_i \in (a,b)$ formunun parametrik olmayan regresyon modellerinde lineer tahminini tartışmıştır.

Silverman (1986), verilerde yoğunluk fonksiyonunun parametrik olmayan tahmininden yola çıkarak, düzleştirme parametresinin seçimi, Kernel tahmin yönteminin ortalaması, varyans seçimi üzerinde oldukça geniş kapsamlı çalışmalar yapmıştır. Parametrik olmayan basit doğrusal regresyon fonksiyonun tahmin yöntemlerinin pratik amaçlar için iyi olduğunu, basit ve sezgisel olarak çekici ve matematiksel özelliklerinin iyi olduğunu belirtmiştir. K-NN yöntemi ile elde edilen tahminler çeşitli gözlemlere karşılık gelen bant genişliğinin bir modelini verdiğini ve bu bant genişliklerinin Kernel tahmini oluşturmada kullanılabildiğini belirtmiştir. Parametrik olmayan basit doğrusal regresyon yöntemlerinin özelliklerini karşılaştırmış avantajlı, dezavantajlı ve anlaşılma kolaylığı olan yöntemleri belirtmiştir.

Daniel (1990), parametrik olmayan regresyon yöntemleri konusunda derleme çalışması yapmıştır. Theil (1950: 1397-1412) tarafından eğimin testi için önerilen yöntemin Kendall Tau istatistiğine dayandığını belirtmiştir.

Hall ve Marron (1990), parametrik olmayan regresyon modelinde varyansın bir tahmincisini önermişlerdir.

Aytaç (1991), çalışmasına göre, bilinmeyen dağılım fonksiyonu sürekli olan veya onun üzerinde herhangi bir bilgiye gerek duymayan yöntemler parametrik olmayan yöntemlerdir. Ayrıca parametrik olmayan testlerin, popülasyon parametrelerinin değerleri üzerine hiçbir varsayım yapılmadığı durumlarda uygulandığı belirtilmiştir.

Hardle ve Mammen (1993), parametrik regresyon uygunluğuna karşı parametrik olmayan regresyon analizini karşılaştırmışlardır. Çalışmalarını "Regresyon verilerinin parametrik modelinin uygunluğu, bir parametrik olmayan smoothing tahmincisi ile mukayesesiyle değerlendirilebilir" şeklinde ifade etmişlerdir.

Simonoff (1998), yoğunluk tahmini ve parametrik olmayan regresyon arasındaki bir bağ olarak kategorik veri düzeltme durumunu araştırmıştır.

Erilli ve Alakuş (2014), çalışmalarında eşit değerli verilerde parametrik olmayan regresyon tahmin çalışmaları yapmışlardır. Özellikle Theil-Sen yönteminde eşit değerli gözlemlere ait hesaplamaların sonsuz değeri vermesi problemine karşılık önerilen yöntem, eşit değerlere sahip veri setlerinde başarı ile uygulanmıştır.

Erilli ve Alakuş (2016), çalışmalarında Mood-Brown regresyon analizinde düzeltilmiş ortalama ve Jackknife parametre tahmininin birlikte kullanımı üzerinde çalışmışlardır. Buna göre, düzeltilmiş ortalama ile kullanılan Jackknife yöntemi diğer Jackknife yöntemi sonuçlarına göre daha küçük HKOK değeri vermiştir.

1.2. Bootstrap Literatür Taraması

Hall (1962), Edgeworth açılımlarını içine alan fonksiyonları ele almıştır. Edgeworth açılımları ve ilgili Cornish-Fisher açılımları bootstrap güven aralıklarının doğruluğunu, bootstrap hipotez testlerinin değerini ve parametrik olmayan regresyonda bootstrap kullanımını daha iyi anlamayı sağlamıştır. Bu açılımlar ayrıca Monte Carlo ve bootstrap türevi çalışmalarda çalışanlara yol göstermiştir.

Efron (1979), bootstrap yöntemine ilişkin yaptığı çalışmada, istatistiğin bootstrap dağılımını elde etmede kullanılan üç farklı yöntemi tanıtmıştır. İlk yöntem, doğrudan kuramsal hesabı içeren yöntemdir. Çalışmada, kuramsal hesabı içeren yöntemin tanıtılması iki örnek ile desteklenmiştir. İkinci yöntem Monte Carlo yaklaşımı, üçüncü yöntem ise Taylor serisi açılımının kullanıldığı yöntemdir. Çalışmada ortanca tahminin Monte Carlo yaklaşımı ile bir değerlendirmesi yapılmıştır. Araştırmacı çalışmasında bootstrap yönteminin regresyon modelinde nasıl uygulandığını da açıklamıştır. Yaptığı çalışmada örnek hacmini artırmaksızın popülasyon parametresi ile tahmin edici arasındaki sapmanın azaltılacağını ve teorik olarak elde edilmesi mümkün gibi görünen ama uygulama da söz konusu olmayan tahmin edicilerin örnekleme dağılımının oluşturulabileceğini belirterek örnekleme dağılımlarının bu şekilde oluşturulmasıyla tahmin edicilerin standart hatasının daha sağlıklı olarak elde edilmesinin sağlanabileceğini bildirmiştir.

Bickel ve Freedman (1981), bootstrap yönteminin bazı matematiksel koşullar altında tutarlılığını gösteren sonuçları elde etmişlerdir.

Efron (1982), yayınlandıktan sonra, bootstrap üzerine yapılan araştırmalar katlanarak çoğalmıştır. Bootstrap tahmininin asimptotik tutarlılığı ve bazı karşı örnekler üzerine birçok teorik gelişmeler ortaya çıkmıştır. Staudte ve Sheather (1990), bootstrap yöntemini, tahminlerin standart hatasını tahminlemek için ele almışlardır. Özellikle dayanıklı tahmincilerin standart hatalarıyla ilgilenmişlerdir. Hipotez testi ile ilgilenmelerine rağmen, hipotez testi problemleri için bootstrap yöntemini kullanmamışlardır.

Efron ve Gong (1983), yaptıkları çalışmada, standart sapma tahmininde kullanılan parametrik olmayan tahmin yöntemlerini tanıtmışlar. Popülasyon dağılımının normal ve negatif üstel olduğu durumlar için ortalamanın standart sapmanın bootstrap ve jackknife tahminleri elde edildiğini açıklamışlar.

Beran ve Srivastava (1985), kovaryans matrisinin fonksiyonları için bootstrap testlerini önermişlerdir.

Stine (1985), doğrusal regresyon modelinde, hata değerlerinin dağılımı için $\text{Sapma}_{\varepsilon} = e_i - \bar{e}$ değerine $\frac{1}{n}$ olasılığını vererek deneysel dağılımın tahmin edilebileceğini göstermiştir.

Wu (1986), eşit varyanslılıkta $\hat{\beta}$'nin bootstrap varyans tahmini sapmasız olurken, bu yöntemin heterojen varyanslıkta sapmalı ve tutarsız sonuç verdiğinden heterojen varyans durumunda tahmin yapmaya uygun olmadığını açıklamıştır.

Efron ve Tibshirani (1986), standart sapmanın bootstrap tahminini ele almışlar; parametre tahmininin de hangi tahmincinin kullanılacağını; bu tahmincinin parametre için ne kadar doğru bilgi taşıdığı sorularından bootstrap tekniği ikinci soruya cevap bulmak için geliştirilmiş bir yöntem olduğunu açıklamışlardır.

Dicicciove ve Tibshirani (1987), bootstrap güven aralıkları ile varyans ve korelasyon katsayısı için ilgili güven aralığı yöntemlerini kullanıp güven aralıkları elde etmişlerdir.

Efron (1990), bootstrap tekrarlama sayısını azaltma amacına yönelik daha etkili bootstrap hesaplamaları yapmış ve tekrarlama büyüklüğünün 50 ile 200 arasında olmasının yeterli olduğunu açıklamıştır.

Hahn ve Meeker (1991), bootstrap güven aralıklarını kısaca ele almışlardır.

Lahiri (1992), M-tahminci (bir çeşit dayanıklı konum tahmincisi) için bootstrap yöntemini kullanmıştır.

Mooney (1993), bootstrap yönteminin örnekleme yöntemlerinden farklı bir yöntem olduğunu belirtmiştir. Bootstrap tekniğinin örnekleme dağılımını oluşturmak için güçlü varsayımlar ve analitik formüller yerine büyük sayıda yinelenmiş hesaplamalar içerdiğini ve bootstrap tekniğinin örnekleme yöntemlerinin yetersizliklerini ortadan kaldırmak, tahmin edicinin örnekleme dağılımının oluşturulmasını sağlamak hem de sapmanın azaltılması amacına yönelik parametrik olmayan bir yöntem olarak geliştirdiğini açıklamıştır.

Shao (1996), regresyon analizinde yöntemin kullanılması sırasında yeniden örnekleme sürecinin, analiz öncesinde gözlemlerin yeniden örneklenmesi veya analiz sonrasında hataların yeniden örneklenmesi biçiminde olacağını açıklamıştır.

Jeremy (1996), yeniden örnekleme yöntemlerinin tanımlı ve bağımsız veriler için kullanılabildiği gibi zaman serilerinde de uygulanabileceğini bildirmiştir.

Strawderman ve Wells (1997), Hazard (Birikimli Tehlike) ve yaşam fonksiyonları için bootstrap güven sınırlarının geçerliliği üzerine çalışmışlardır. Aynı zamanda rasgele olarak azalan veri için bootstrap yönteminin yararlı olacağının belirtmişlerdir.

McLachlan ve Krishnan (1997), bootstrap yöntemini, kovaryans matrisinin dayanıklı tahmin aracı olarak ele almışlardır.

Zeng ve Davidan (1997), bağışıklık sisteminde bootstrap güven aralıklarının kullanımı üzerinde durmuşlardır.

Fox (1997), bootstrap yönteminin hata terimlerinden tekrarlı örnekler seçerek tahmin değerlerine ($\hat{Y}_1$) eklendiği için hata terimlerinin normal dağılıma sahip olduğunu varsaymıştır.

Topuz (2002), Klasik Örnekleme ve Yeniden Örnekleme (Bootstrap ve Jackknife) yöntemi sonuçlarını karşılaştırmıştır. Bootstrap yöntemi ile yapılan tahminlerin standart hataları Klasik Örnekleme yöntemi ile yapılan tahminlerin standart hatalarından daha küçük tahminler elde edilmiştir. Bu şekilde jackknife ve bootstrap yöntemi sonuçları da karşılaştırılmış ve jackknife yöntemi, bootstrap yönteminden daha büyük standart hatalı tahminler vermiştir. Jackknife yöntemi de bootstrap yöntemi gibi parametrik varsayımlar kullanmak yerine verideki değişkenliği kullanmaya yönelik yöntemler olmasına rağmen bootstrap yönteminin jackknife yönteminden daha az hatalı parametre tahminleri yaptığı belirtilmiştir.

Atakan (2003), hata oranı tahmin edicilerinin jackknife ve bootstrap değerlendirmesi çalışmasında, hem gerçekhata oranı için ve hem de bu hata oranına ilişkin tahminediciler için tüm durumlarda genelde jackknife tekniğine göre elde edilen sonuçlar, bootstrap tekniğine göre elde edilen sonuçlara göre gerçek değerlere daha yakın olduğu görülmektedir. Buradan jackknife yönteminin, bootstrapyöntemine tercih edilebileceğini belirtmiştir.

Okutan (2009), bootstrap tekniğini bazı özel dağılımlarının doğrusal regresyon modeli için Uyarlanmış En Çok Olabilirlik ve Klasik En Küçük Kareler tahmin edicisinin varyansını tahmin etmede kullanmıştır ve sonuçlara göre regresyonda bootstrap yönteminin kullanımının daha gerçek sonuçlar vereceğini belirtmiştir.

Özel ve Sezgin (2011), veri sayısı az olduğundan ve sağlam (robust) tahminler üretmek amacıyla çalışmalarında bootstrap tekniği kullanmışlardır.

BÖLÜM 2

2. REGRESYON ANALİZİ

Regresyon analizi aralarında sebep sonuç ilişkisi olan iki veya daha fazla değişken arasındaki ilişkiyi incelemek ve o konuyla ilgili tahmin yapabilmek amacıyla oluşturulan ve matematiksel bir model ile belirtilen istatistiksel bir analizdir. Bu çözümleme yönteminde iki veya daha fazla değişken arasındaki ilişki açıklamak için matematiksel bir model kurulur ve bu model regresyon modeli olarak adlandırılır (Birkes, Dodge 1993: 80-140). Kurulan modellerin geçerliliğini analiz için yararlanılan yönteme ise regresyon analizi adı verilmektedir. Regresyon analizi kullanım amaçlarına veya farklı model yapılarına göre farklı alt başlıklar ile sınıflandırılabilir. Uygun model seçimi kesin kurallarla belirlenmemesine rağmen, aşağıda verilen tablo uygun model yapıları için özet bilgi sunmaktadır.

Tablo 2.1. Regresyon Analizi Ayrımları

Regresyon Analizi Ayrımları	**Regresyon Modelleri**	**Açıklamalar**
Bağımlı Değişken Sayısına Göre	Tek Değişkenli Regresyon	Tek bir nicel bağımlı değişken olmalıdır.
	Çok Değişkenli Regresyon	İki veya daha fazla nicel bağımlı değişken olmalıdır.
Bağımsız Değişken Sayısına Göre	Basit Regresyon	Bir bağımlı ve sadece bir tane bağımsız değişken olmalıdır.
	Çoklu Regresyon	Bir bağımlı ve iki veya daha fazla bağımsız değişken olmalıdır.
Matematiksel Biçimine Göre	Doğrusal Regresyon	Tüm parametreler doğrusal olmalıdır. Değişken dönüştürme ile doğrusal hale getirilen yapılar da bu kısımda yer almaktadır.
	Doğrusal Olmayan Regresyon	Bazı parametrelerin doğrusal olmadığı durumlar. Bağımlı değişken ile bazı bağımsız değişkenlerin doğrusal olmayan ilişkide olmaları. Fakat değişken dönüştürme ile doğrusal hale getirilemeyen yapılar olmalıdırlar.
Bağımsız Gölge Değişkenli	Varyans Analizi Modelleri	Tüm bağımsız değişkenler nitel yapıda olmalıdırlar.
	Kovaryans Analizi Modelleri	Bazı bağımsız değişkenler nitel, bazıları ise nicel yapıda olmalıdırlar.
Varsayımların Sağlanma Durumuna Göre	Parametrik Regresyon	Parametrik regresyon analizinin başarılı bir şekilde uygulanabilmesi için verinin normal dağılması, eşit varyanslı olması, otokorelasyonlu olmaması gibi varsayımların sağlanması gerekmektedir.
	Parametrik Olmayan Regresyon	Parametrik regresyon yöntemler için geçerli olan bazı varsayımların sağlanamaması durumlarında kullanılan yöntemlerdir.

(Kaynak: Erilli 2015: 175)

Varsayımların sağlanma durumlarına göre Regresyon Analizi çeşitleri aşağıda kısaca tanımlanmıştır.

2.1. Parametrik Regresyon Analizi

Parametrik regresyon analizi bağımlı ve bağımsız değişkenler ile bu değişkenler arasındaki ortalama ilişkinin matematiksel bir fonksiyonla ifade edilmesi ve bu fonksiyondaki parametre vektörlerinin açık bir şekilde gösterilmesidir. Parametrik regresyon, regresyon fonksiyonunun $z_1, z_2, ..., z_p$ bağımsız değişkenlerinin bir doğrusal fonksiyonu olarak yazılabildiğini varsayar. $E(y \mid Z)$ koşullu beklenen değeri Z biliniyorken y 'nin ortalama dağılımının Z ile fonksiyonel ilişkisini gösterir. Başka bir ifade ile $Z = (z_1, z_2, ..., z_p)$ bağımsız değişkenlerindeki değişime karşılık y bağımlı değişkeninin ortalama tepkisini ifade eder ve

$$E(y \mid Z) = Z\beta \tag{2.1}$$

Eşitlik 2.1 şeklinde veya,

$$y = Z\beta + \varepsilon \tag{2.2}$$

Eşitlik 2.2 şeklinde yazılır. Burada $\varepsilon = y - E(y \mid Z)$ ifadesi $E(y \mid Z)$ koşullu beklenen değerinin y 'den sapması olarak tanımlanır (Hardle, Muller, Sherlich, Werwatz 2004:). Eşitlik 2.2' de gözlem sayısı n, bağımsız değişken sayısı p olmak üzere; $y(n \times 1)$ boyutlu bağımlı değişken vektörü, $Z(n \times p)$ boyutlu ve p ranklı bağımsız değişkenler matrisi, β $(p \times 1)$ boyutlu bilinmeyen regresyon katsayıları vektörü, ε ise gözlenemeyen 0 ortalamalı ve sabit varyanslı rasgele hataların $(n \times 1)$ boyutlu vektörüdür. Parametrik regresyon modelinde amaç modelin uydurulması ve model uygunluğunun araştırılmasıdır. Doğrusal regresyon modelini belirlemek için bilinmeyen β parametrelerini tahmin etmek gerekir. Eşitlik 2.2'de (y_i, z_i) gözlem değerlerine karşılık gelen nokta ile bu noktanın en küçük kareler yöntemi ile elde edilmiş olan doğru üzerindeki izdüşümleri toplamı yani $\varepsilon_i = y_i - \hat{y}_i$ sıfır olmalıdır. Ayrıca bu farkın kareler toplamı minimum olmalıdır. Bu durumda eşitlik 2.2 ile verilen modelde parametre tahminleri eşitlik 2.3 elde edilir.

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - y_i \right)^2 = \sum_{i=1}^{n} \left(y_i - z_i^{'} \hat{\beta} \right)^2 \tag{2.3}$$

Parametrik yaklaşım, tümüyle varsayımlara dayalıdır. Eşitlik 2.2'de Z bağımsız değişkenler arasındaki fonksiyonel yapının $Z\beta$ biçiminde doğrusal olduğu ve β parametrelerinin sonlu olduğu varsayılmaktadır. Ancak burada değişkenler arasında doğrusal olmayan ilişkiler varsa parametrik yöntemler yerine parametrik olmayan yöntemlerin kullanılması gerekir.

Günümüzde parametrik olan istatistiksel yöntemlere karşılık olarak parametrik olmayan yöntemler geliştirilmiştir. Parametrik olmayan yaklaşımda β parametre vektöründen bahsedilmemekte ve bağımsız değişkenler arasındaki fonksiyonel yapı da bilinmemektedir. Hiçbir varsayım gerektirmediğinden dolayı bu yaklaşımın uygulamalarda daha fazla tercih edilmesi gerekirken, birçok problem ile karşılaşılmasından dolayı nadiren kullanılmaktadır. Bu problemler içinde en fazla dikkat çekeni, bağımsız değişken sayısının fazla olması durumunda tahmin ve yorumlamada güçlük çekilmesidir (Hardle, Muller, Sperlich, Werwatz 2004:).

Bağımlı değişkeninin bağımsız değişkenlerden bazıları ile doğrusal ilişki içinde fakat bazıları ile de ilişkisinin kolayca parametreleştirilemediği durumlar vardır. Bu gibi durumlarda parametrik ve parametrik olmayan regresyon yöntemlerinin yeterli olamayacağı açıktır. Bu nedenle, hem parametrik hem de parametrik olmayan regresyon modellerini içeren ve bu modellerin özel bir durumu olan '*kısmi doğrusal model'* olarak da adlandırılan '*yarı parametrik regresyon modeli'* parametre tahmininde daha uygun bir sonuç verecektir.

2.1.1. Basit Doğrusal Regresyon

Birbiriyle ilişkili olan iki değişkenin olduğu basit doğrusal regresyonda, değişkenlerden birisi bağımlı diğeri ise bağımsız değişken olarak kabul edilir (Baskan 1993). Y bağımlı değişkeni, X bağımsız (tasarım) değişkeni, β_0 ile β_1 bu değişkenin bilinmeyen parametrelerini ve ε hata terimlerini göstermek üzere kitle için basit doğrusal regresyon denklemi eşitlik 2.4 şeklinde yazılır.

$$Y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i, i = 1,2,...,n \qquad \textbf{(2.4)}$$

Doğrusal regresyon yöntemini kullanmak için bazı varsayımların bulunduğu kabul edilmektedir. Varsayımlardan herhangi birinin gerçekleşmemesi durumunda geçerli sonuçlara ulaşmak mümkün olmaz. Berry (1993), bu varsayımları şöyle sıralamaktadır:

Bağımsız değişkenlerin hepsi nicel veya nitel olarak ölçülmüş olması, bağımlı değişken Y 'nin ise nicel ve sürekli olması gerekmektedir. X ve Y değişkenleri doğru olarak ölçülmelidir.

1. Bütün bağımsız değişkenlerin varyansının sıfırdan farklı olması gerekmektedir.
2. Hata terimleri ortalaması sıfırdır: $E(\varepsilon_i) = 0$
3. Bağımsız değişkenler ve hata terimi arasında korelasyon olmamalıdır.
4. Hata terimlerinin varyansı sabit olmalıdır, $E(\varepsilon_i^2) = \sigma^2$
5. Hata terimleri arasında korelasyon olmamalıdır. $E(\varepsilon_j \varepsilon_i) = 0, (i \neq j)$
6. Hata terimleri ε_i, normal dağılmalıdır.

2.1.1.1. En Küçük Kareler (EKK) Yöntemi

En küçük kareler yöntemi, birbirine bağlı olarak değişen iki büyüklük arasındaki matematiksel bağlantıyı, gerçeğe en yakın bir denklem olarak yazmak için kullanılan, bir regresyon tahmin yöntemidir. Hesaplama kolaylığından dolayı, en çok kullanılan yöntemdir. Nokta tahmin yöntemidir; yani örneklem veriyken, her tahmin edici ilgili anakitle katsayısının tek bir tahminini verir (Gujarati 2005: 850).

En küçük kareler yönteminin temeli hata kareler toplamını minimum yapmak üzerinedir. Yani EKKY temel işleyiş mantığı $\min \sum u_i^2$ ' dir. Bir başka ifade ile $\min \sum \left(Y_i - \hat{Y}_i\right)^2$.

En küçük kareler yöntemi iki farklı yöntemle parametre tahminine imkân sağlar. Bunlardan birincisi normal denklemler yöntemi diğeri ise ortalamalardan farklar yöntemidir.

EKKY Normal Denklemler:

$Y_i = \beta_0 + \beta_1 X_1 + u_i$ basit doğrusal regresyon modelindeki parametre tahminlerini elde etmek için gerekli formüllerdir 2.5 ve 2.6'da verildiği gibidir:

$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i \tag{2.5}$$

$$\sum_{i=1}^{n} Y_i X_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 \tag{2.6}$$

Bu formüller; $\sum u_i^2$ formülündeki değerlerin yerlerine yerleştirilerek yani eşitlik 2.7'de gösterildiği gibi fonksiyonunun β_0 ve β_1'e göre kısmi türevleri alınarak ve sıfıra eşitlenmesi ile bulunurlar.

$$\sum u_i^2 = \sum \left(Y_i - \beta_0 - \beta_1 X_1 \right) \tag{2.7}$$

Ortalamalardan Farklar Yöntemi:

$Y_i = \beta_0 + \beta_1 X_1 + u_i$ basit regresyon modelindeki katsayıları elde etmek için gerekli formüller şunlardır:

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \tag{2.8}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \tag{2.9}$$

Eşitlik 2.9'da $x_i = \left(X_i - \overline{X} \right)$ ve $y_i = \left(Y_i - \overline{Y} \right)$olarak tanımlanmaktadır.

2.1.1.2. En Çok Olabilirlik (EÇOK) Yöntemi

EÇOK yönteminde olduğu hem X hem de hata teriminin bir dağılıma sahip olduğu varsayılırsa iki değişkenli olabilirlik fonksiyonu L, x marjinal ve e hata teriminin (koşullu dağılım) olabilirliği şeklinde yazılabilir.

$$L = L_x L_e \tag{2.10}$$

$$InL = InL_x + InL_e \tag{2.11}$$

Eşitlik 2.11' de InL'yi maksimize eden tahmin ediciler InL_x ve InL_e'yi maksimize etmek durumundadır. Literatürde en yaygın olarak kullanılan iki değişkenli dağılım olan iki değişkenli normal dağılım varsayıldığında InL_x ve InL_e'yi maksimize eden EÇOK tahmin edicilerine eşittir.

2.1.2. Çoklu Doğrusal Regresyon

Basit doğrusal regresyon analizinde bir bağımlı ve bir bağımsız değişken arasındaki fonksiyonel ilişki incelenirken, çoklu doğrusal regresyon analizinde bir bağımlı ve birden fazla bağımsız değişken arasındaki fonksiyonel ilişki incelenmektedir.

Çoklu doğrusal regresyonda araştırmacının iki genel amacı vardır. Bunlardan birisi bağımlı değişkeni etkilediği belirlenen değişkenler vasıtasıyla bağımlı değişkenin değerini tahmin etmek, bir diğeri; bağımlı değişkeni etkilediği düşünülen bağımsız değişkenlerden hangisinin veya hangilerinin bağımlı değişkeni daha çok etkilediğini tespit etmek ve aralarındaki ilişkiyi tanımlamaktır (Alpar 2003: 408).

Yapılan araştırmalarda daha sağlam sonuç elde edebilmek için ele alınan faktörü etkileyen bütün etkenleri gözlemek ve bağımsız değişkenlerin bağımlı değişkeni ne yönde ve nasıl etkilediğini tespit etmek önemlidir. Örneğin; yumurta tavuklarında yumurta verimine havanın sıcaklığı, tavuğun yaşı, verilen yem miktarı ve ışıklandırma süresi gibi faktörler etkili olmaktadır. Bu faktörlerden hangisinin daha çok etkili olduğunu ve bu faktörlere göre aylık yumurta verimini tahmin etmek için çoklu regresyon analizi kullanılmalıdır.

Y bağımlı değişkeni ile p sayıda bağımsız değişken arasındaki ilişki doğrusalsa, Y ve X ’lere ait n tane gözlem değeri varsa çoklu doğrusal regresyon modeli; eşitlik 2.12 ve 2.13 şeklindedir.

$$Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + e_i \qquad (i = 1,2,\ldots,n) \qquad \textbf{(2.12)}$$

$$Y = \beta_0 + \sum_{k=1}^{p} \beta_k X_{ik} + e_i \qquad \textbf{(2.13)}$$

Burada $\beta_0, \beta_1, \ldots, \beta_p$ bilinmeyenleri kısmi regresyon katsayıları veya kısaca regresyon katsayılarıdır.

2.1.2.1. Çoklu Doğrusal Regresyon Analizinde Katsayıların Tahmini

Çoklu doğrusal regresyon analizinde genel regresyonda uyulması gerekli şartlara ilave olarak bağımsız değişkenler arasında çoklu bağımlılık olmaması şartı aranır.

Hatalar rasgele değişkenler olduğu için hatalara ilişkin bazı varsayımlarda bulunmak gerekir. Bu varsayımlar:

1. $E(\varepsilon_j)=0$, ε_j'nin beklenen değeri sıfırdır.
2. $Var(\varepsilon_j)=\sigma^2$
3. $Cov(\varepsilon_j,\varepsilon_k)=0 \quad j \neq k$

Burada ε vektöründeki hata terimlerinin normal dağılım gösterdiği varsayılır. Çoklu doğrusal regresyon modelinde bulunan β parametrelerinin tahmin edicisi olan $p+1$ tane β istatistik değerleri EKK yöntemi yardımıyla eşitlik 2.14 kullanılarak bulunur.

$$\beta = \left(X^1X\right)^{-1}\left(X^1Y\right) \tag{2.14}$$

β parametrelerine ilişkin güven aralıklarının oluşturulmasında, yokluk ve alternatif hipotez testlerinde serbestlik derecesi $[n-(p+1)]$ olan t dağılımından faydalanılır.

Çoklu doğrusal regresyon modelinin varsayımları aşağıdaki gibidir (Kılıçbay 1980: 67, Koutsoyiannis 1992: 638, Çömlekçi 1998: 493, Gujarati 2005: 850).

1. Bağımlı değişken bir rasgele değişkendir ve normal dağılım gösterir.
2. Bağımsız değişkenler rasgele değişken değildir. Bağımsız değişkenlerin değerleri araştırmacı tarafından önceden belirlenmiş veya isteğe bağlı olarak seçilmiş ya da sabit değerlerdir.
3. Tahminlerin hataları rasgeledir ve birbirleriyle ilişki göstermez. $\left(Cov(\varepsilon_i,\varepsilon_j)\right)=0$. Aynı zamanda varyansı $V(\varepsilon)=\sigma^2$ ve ortalaması sıfırdır.
4. $Cov(\varepsilon, X_{ji})$, $j=1,2,...,p$ ve $i=1,2,...,n$ dir.
5. Bağımsız değişkenler arasındaki korelasyon katsayıları 0 veya 0'a çok yakın olmalıdır. Bu varsayıma çoklu doğrusal bağlantı olmama adı verilmektedir.
6. Hipotez testlerini yapmak ve güven aralıklarını oluşturabilmek için hataların normal dağılıma sahip olduğu varsayılır.

Çoklu doğrusal regresyon analizinde bağımlı değişken ve bağımsız değişkenler arasındaki ilişkinin gücünü açıklamak için R ile gösterilen çoklu korelasyon katsayısı kullanılır. Ancak çoklu determinasyon (belirleme) katsayısı olan R^2 bu ilişki hakkında yorum yapmak için daha uygundur.

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \qquad \textbf{(2.15)}$$

Teorik olarak bağımlı değişkeni açıklayabilen sonsuz sayıda bağımsız değişken bulunabilir. Fakat uygulamada 1 ya da 2 bazen 3 tane bağımsız değişken bağımlı değişkendeki varyansın büyük bir kısmını açıklar. R^2 değerinin 0,80 dolaylarında olması yeterli sayılır (Ünver, Gamgam 2008: 424).

2.2. Parametrik Olmayan Regresyon Modelleri

Bir y bağımlı değişkeni ve bu değişkenle ne tür bir ilişki içerisinde olduğu bilinmeyen bir x bağımsız değişkeninin yer aldığı basit parametrik olmayan regresyon modeli, eşitlik 2.16 şeklindedir.

$$y_i = f(x_i) + \varepsilon_i, i = 1,...,n \qquad \textbf{(2.16)}$$

Burada $f \in C^2[a,b]$ olan bir düzgün fonksiyon, x_i parametrik olmayan bağımsız değişkenlere ait gözlem değerleri, y_i bağımlı değişkene ait gözlem değerleri, ε_i 0 ortalamalı ve σ^2 sabit varyanslı, bağımsız özdeş olarak dağılan rasgele hata terimleridir.

Parametrik olmayan regresyon analizinin amacı parametreleri tahmin etmekten çok bilinmeyen yanıt fonksiyonu (ortalama fonksiyon) olan $f(x)$'i tahmin etmektir. $x = (x_1, x_2, ..., x_n)$ bağımsız değişkenlerinin ve y bağımlı değişkeninin ölçümleri arasındaki ilişkiyi açıklayan en yaygın yöntem, $E(y \mid x) = f(x)$ koşullu beklenen fonksiyonunu tahmin etmektir. Bu ilişkiyi açıklayan genel parametrik olmayan regresyon modeli,

$$\begin{aligned} y &= E(y \mid x) + \varepsilon \\ &= f(x) + \varepsilon \end{aligned} \qquad \textbf{(2.17)}$$

şeklinde ifade edilir. 2.17 eşitliğinde $y = (y_1,..., y_n)'$, $f = (f_1,..., f_n)' = (f(x_1),..., f(x_n))'$ ve $\varepsilon = \varepsilon(\varepsilon_1,...,\varepsilon_n)'$ olup f belirgin bir şekle sahip olmayan x açıklayıcı değişkenlerinin bir fonksiyonudur. Ayrıca f ikinci mertebeden sürekli türeve sahip olan fonksiyonlar uzayının bir elemanıdır ve $f \in C^2[a,b]$ modelin düzeltme kısmıdır. Parametrik olmayan tahmin edicinin yakınsama oranı genellikle parametrik tahmin edicilerden daha yavaştır. Bu nedenle parametrik tahmin ile kıyaslamada parametrik olmayan yöntemler çok büyük örneklem hacimleri gerektirir (Yatchew 1998: 669-721). Ancak parametrik modellerde olduğu gibi bu tür modellerde de istatistiksel doğruluk örneklem hacmine değil tahmin edicilerin varyans ve kovaryanslarına bağlıdır (Heerde, Leeflang, Witting 2001: 197-215, Yatchew 2003: 669-721).

Parametrik olmayan regresyon yönteminin tahmin yaparken kısıtlayıcı varsayımları olmamasına rağmen bazı sakıncaları vardır. Bağımsız değişken sayısı fazla olduğu zaman tahmin yapmak zor olmakta ve elde edilen grafikler karmaşık bir yapıda olmaktadır. Bu durum 'boyutluluk sorunu' olarak adlandırılır. Ayrıca parametrik olmayan yöntemle kesikli bağımsız değişkenleri dikkate almak ve bağımsız değişken sayısındaki artışa bağlı olarak y değişkenine ait bireysel etkileri yorumlamak zor olmaktadır. Parametrik olmayan bu yöntemin sakıncaları yarı parametrik regresyon modeli kullanılarak giderilmektedir. Yarı parametrik regresyon modeli hem parametrik hem de parametrik olmayan regresyon modelinin her ikisini birlikte kullanır. Bu nedenle yarı parametrik regresyon modeli parametrik modellerin kısıtlayıcı varsayımlarından etkilenmemekle birlikte parametrik olmayan yöntemlerin cazip özelliklerini bir araya getirmektedir.

Parametrik olmayan regresyon aykırı gözlemlerin bulunduğu veri setleri için önemli bir analiz yöntemidir. İstatistiksel çalışmalarda aykırı gözlemlerin etkilerini farklı biçimlerde ele alan güçlü parametrik yöntemler bulunmaktadır. Bununla birlikte, aykırı gözlemlerden dolayı parametreler bozulduğu için bu güçlü yöntemler bile uygun çözümler üretemeyebilir ve verinin gerçek yapısı modele yansıtılamaz. Bu durumda parametrik olmayan regresyon, ön bilgi sağlamaktadır (Hardle 1994: 1-23). Parametrik yaklaşımda çok fazla varsayım yapıldığından dolayı sonuçların güvenilirliği giderek azalmaktadır. Parametrik olmayan yaklaşımda ise hiçbir varsayım yapılmamakta fakat bağımsız değişken sayısının fazla olması durumunda model tahmininin elde edilmesi zor olmaktadır. Yarı parametrik yaklaşım, parametrik ve parametrik olmayan yaklaşım arasında bir orta yol bulmayı amaçlamaktadır.

2.2.1. Parametrik Olmayan Regresyon Modellerinin Özellikleri

Parametrik olmayan regresyon modellerinin en büyük özelliği, parametrik modellerin aksine modele ait herhangi bir fonksiyonel formu dikkate almadan bağımlı ve bağımsız değişkenler arasındaki ilişkiyi belirlemeye çalışan bir yöntem olmasıdır. Buna göre parametrik olmama terimi, regresyon eğrisinin (ya da doğrusunun) esnek fonksiyonel formunu kapsamaktadır (Hardle 1994: 1-23). Anakitle regresyon fonksiyonun genel gösterimi olan eşitlik 2.18 aynı zamanda parametrik olmayan regresyon modellerinin de gösterimidir.

$$Y_i = m(X_i) + u_i \qquad \textbf{(2.18)}$$

Parametrik modellerin dezavantajları ışığında birçok parametrik olmayan tahminci, $m(X_i)$'nin "düzgün" tahminlerini üretmek amacıyla geliştirilmiştir (Pagan, Ullah 1999).

Parametrik olmayan yaklaşımla bir regresyon ilişkisini belirlemeye çalışmanın 4 avantajı bulunmaktadır. Bunlar:

- Parametrik olmayan tahmin yöntemi bağımlı ve bağımsız değişken arasındaki
  genel ilişkinin araştırılmasında çok yönlü ve esnek metot sağlamaktadır.
- Herhangi bir fonksiyonel forma ihtiyaç duyan sabit parametrik modeli temel
  almaksızın gözlemlere ait tahminler yapar.
- Uç değer olarak tabir edilen değerlerin etkisinin analizde yer alması sebebiyle
  "sahte" gözlemleri bulmak için bir araç niteliği taşımaktadır.
- Bağımsız değişkenler arasında enterpolasyon yapılabilmesi ve veri seti içerisinde kayıp gözlemlerin yerine konulabilmesi açısından esnek bir metot
  oluşturmaktadır.

Parametrik olmayan bir regresyon modeli eşitlik 2.19 şeklinde gösterilmektedir:

$$Y_i = m(X_i) + u_i \qquad \textbf{(2.19)}$$

Bu regresyon modelinde parametrik olmayan yaklaşım olarak tabir edilen yapı, $m(X_i)$'in herhangi bir fonksiyonel form varsayılmadan tahmin edilmesidir. Parametrik olmayan tahmin bu yapıyla ilgilenmektedir. Parametrik olmayan modellere özgü "düzgünleştirme" teknikleri ile $m(X_i)$ anakitle regresyon fonksiyonuna ait en uygun düzgün tahmini üretmeye çalışmaktadır.

$\{X_i, Y_i\}_{i=1}^{n}$ veri setinin düzgünleştirmesi, regresyon ilişkisindeki m'nin ortalama tepki eğrisinin yaklaşımını kapsamaktadır (Hardle 1999: 1-23). Burada m'nin fonksiyonel formu regresyon eğrisinin kendisi, türevleri ya da uç değer noktaları gibi türevlerinin fonksiyonları olarak belirlenebilmektedir.

Analizde kullanılan verilere göre $m(X_i)$'nin tahmini farklılık göstermektedir. Bağımsız değişken X_i'nin gözlenen değerleri içerisinde tekrarlanan ve tekrarlanmayan değerlere göre ortaya çıkan bu farklılık sebebiyle $m(X_i)$' in tahmini eldeki veri setine göre farklılık gösterir.

Bağımsız değişkenlere ait gözlemlerde, tekrarlanan değerler var ise bu durumda regresyon eğrisinin tahmini, bağımlı değişkenin uygun değerlerinin ortalaması yardımıyla elde edilebilmektedir. Örneğin $X = x_i$ gibi bir değerde tekrar eden gözlemler sabit ise; regresyon eğrisinin tahmini, Y_i bağımlı değişken değerlerinin ortalaması kullanılarak yapılabilmektedir.

Bu durum daha çok kesit verileri ile yapılan analizlerde ortaya çıkmaktadır. Bağımsız değişkene ait gözlenen değerler $X_i (i = 1,2,...,n)$ ise bağımsız değişkene ait tüm gözlem değerleri içerisinde x gibi bir sabit (yinelenen) gözlem belirlendiği durumda regresyon eğrisinin tutarlı tahminci 2.20 eşitliği ile elde edilir:

$$\hat{m}(x_i) = \frac{n^*}{n} \qquad \textbf{(2.20)}$$

n^*, x_i' den x_n' e kadar olan tüm gözlemler içerisinde, x sabit (yinelenen) değere eşit sayıdaki değerleri ifade etmektedir. Örneğin tüm gözlem değerleri arasında söz konusu x değeri 5 kez tekrarlanmışsa, n^* değeri 5 olarak belirlenmektedir. Diğer taraftan n ise toplam gözlem sayısını ifade etmektedir.

Yukarıdaki denklemde yer alan regresyon eğrisi tahmincisi, daha genel bir ifadeyle eşitlik 2.21' deki gibi verilir.

$$\hat{m}(x_i) = \frac{1}{n}\sum_{i=1}^{n} I(x_i = x) \tag{2.21}$$

Eşitlik 2.21' de $I(x_i = x)$, gösterge fonksiyonu olarak tabir edilen fonksiyonu ifade etmektedir. Söz konusu gösterge fonksiyonu, $x_i = x$ olduğu durumda "1" değerini, bu eşitliğin sağlanmadığı durumlarda ise "0" değerini almaktadır. Bu gösterge fonksiyonundan hareketle $m(X_i)$ regresyon eğrisinin tahmini, X_i bağımsız değişkenin sabit (yinelenen) değerleri ile kesit verilerinin kullanıldığı regresyon analizlerinde bu şekilde gerçekleştirilmektedir. Bu yaklaşım "Yerel Histogram Yaklaşımı" olarak adlandırılmaktadır. Belirtilen yöntem parametrik olmayan tahmin yöntemleri içerisinde regresyon eğrisi tahmini için kullanılan en basit ve temel yöntem olup aynı zamanda "Yerel Ortalama Yaklaşımı" olarak da isimlendirilmektedir.

Ancak bu yaklaşım yalnızca analizde kullanılan değişkenlerin kesit verileri olarak gözlendiği durumlarda kullanılmaz. x_i'nin x'e eşit olması ihtimali sıfırdır ve $m(X_i)$, x_i değerlerinin x değerleri etrafındaki ortalaması alınarak tahmin edilmelidir (Pagan, Ullah 1999). Bu sebeple yerel histogram yaklaşımı ile yapılan parametrik olmayan tahminlerde kesit verileri yerine sürekli veriler ile analizler yapmak daha gerçekçidir. Sürekli verilerle yapılan analizlerde bu yaklaşımla elde edilen regresyon eğri tahmincisinin hesaplanması da farklılık göstermektedir.

Burada x_i değerlerinin belirlenen x değeri etrafında ortalaması ile kastedilen, $h/2$ gibi bir değere tekabül eden aralık etrafındaki ortalamadır. h, düzgünleştirme parametresi ya da bant aralığı olarak adlandırılmaktadır. Bu çerçeve içerisinde yerel ortalama yaklaşımına göre elde edilen eğri tahmincisi:

$$\hat{m}(x) = \frac{1}{nh}\sum_{i=1}^{n} I\left(-1/2 \le \frac{x_i - x}{h} \le 1/2\right) \tag{2.22}$$

Eşitlik 2.22 $\hat{m}(x)$, orta noktası x olan $\left(x - \frac{h}{2}, x + \frac{h}{2}\right)$ aralığında değişen birim başına göreli frekans olarak belirlenmektedir. $\hat{m}(x)$ eğri tahmincisinin elde edilmesi, gözlem sayısı n ve düzgünleştirme parametresi olarak ifade edilen h değerlerine bağlı olarak değişmektedir.

Düzgünleştirme parametresi h, söz konusu tahmini gerçekleştirmek için verilerin düzgünleştirildiği (ortalandığı) değeri kontrol etmektedir (Pagan, Ullah 1999). Ortalama değeri, düzgünleştirme parametresi ile uyumlu olan ağırlıklar tarafından kontrol edilir ve bu düzgünleştirme parametresi x etrafındaki komşuluğun ölçüsünü (büyüklüğünü) düzenlemektedir (Hardle 1999: 1-23).

Yerel ortalama yaklaşımı ile sürekli verilerin kullanıldığı analizlerde, bağımsız değişkene ait bir takım ağırlıklar tanımlayarak da regresyon eğrisine ait tahminler yapılabilmektedir.

$$\hat{m}(x) = \frac{1}{n}\sum_{i=1}^{n} w_{ni}(x) y_i \tag{2.23}$$

Eşitlik 2.23'te $w_{ni}(x)$, bağımsız değişkene ait tüm gözlemler için elde edilen ağırlıkları göstermektedir. Buna göre en basit yolla herhangi bir regresyon eğrisini tahmin etmek için bağımsız değişkene ait ağırlıkların hesaplanmasının ardından, bu ağırlıkların bağımlı değişken ile çarpılıp toplanmasıyla eğrinin "düzgün" tahmini elde edilmektedir. Her düzgünleştirme metodu yukarıdaki denklemin asimtotik formu olup $\hat{m}(x)$ regresyon tahmincisi düzgünleştiren (smoother) ve bu düzgünleştirme yaklaşımının ortaya çıkarttığı durum ise düzgün tahmin (smooth) olarak adlandırılmaktadır. Yerel ortalama yaklaşımında, $w_{ni}(x)$ ağırlıklarının pozitif değerler aldığı ve bağımsız değişkene ait tüm değerlerin toplamının 1 olduğu varsayılır:

$$\frac{1}{n}\sum_{i=1}^{n} w_{ni}(x) = 1 \tag{2.24}$$

Böylece $\hat{m}(x)$'in x noktasına göre en küçük kareler tahminine ait minimizasyonu eşitlik 2.25 şeklindedir:

$$\frac{1}{n}\sum_{i=1}^{n} w_{ni}(x)(y_i - \hat{m}(x))^2 \qquad \textbf{(2.25)}$$

Bu minimizasyon durumunda regresyona ait hata terimleri ikinci dereceden (kuadratik) ağırlıklarla elde edilmektedir. Bu nedenle yerel ortalama yaklaşımı, yerel ağırlıklı en küçük kareler yöntemi ile benzerlikler göstermektedir.

Düzgünleştirme ile analizde yer alan bağımsız değişkenlere göre farklılık gösteren ağırlıklar elde edilmektedir. Söz konusu ağırlıkların elde edilmesinden sonra bu ağırlıklar kullanılarak değişkenlere ait veriler düzeltilmektedir. Böylece hiçbir matematiksel kalıba ihtiyaç duyulmadan bağımlı ve bağımsız değişken(ler) arasında nedensellik ilişkisi kurulmaya çalışılmaktadır. Ayrıca yerel ağırlıklı ortalama yaklaşımı genellikle doğrusal olmayan düzgünleştirme tekniklerine uygulanmaktadır.

2.2.2. Parametrik Olmayan Yöntemlerin Avantajları

Parametrik olmayan istatistik süreçlerin son 60 yılda hızlı ve sürekli bir biçimde gelişmesi parametrik olmayan yöntemlerin aşağıda belirtilen avantajlarından ileri gelmektedir (Hollander, Wolfe 1999).

- Parametrik olmayan yöntemler; verinin elde edildiği anakitlenin hangi dağılımdan geldiği hakkında güçlü varsayımlar gerektirmez.
- Parametrik olmayan yöntemler; bu yöntemleri kullananlara, testlerin kesin p değerlerini, güven aralıkları için kesin kapsam olasılıklarını, çoklu karşılaştırma süreçleri için kesin deneysel hata oranlarını ve anakitle dağılımının normal olması varsayımına bağlı olmaksızın güven bandları için
  kesin kapsam olasılıklarını vermektedir.
- Parametrik olmayan yöntemlerin; her zaman olmasa bile, çoğunlukla uygulanmaları parametrik karşılıklarına göre daha kolaydır.
- Parametrik olmayan yöntemleri çoğunlukla anlamak kolaydır.

- İlk bakışta parametrik olmayan yöntemlerde, örneklerdeki temel bilginin çoğu kaybediliyor gibi görünmesine rağmen bu tam anlamıyla doğru değildir. Genellikle bu yöntemler, normal dağılıma uyan parametrik karşılıklarına göre, yalnızca daha az etkindirler.
- Parametrik olmayan yöntemler, aykırı gözlemlerden daha az etkilenirler.
- Parametrik olmayan teknikler, parametrik tekniklerde olduğu gibi birimlerin büyüklüklerinden daha çok onların sıralarıyla ilgilendikleri için normal dağılım varsayımı gerektiren analizlerin uygulanamadığı birçok durumda kullanılabilir.
- Bilgisayar teknolojisinin gelişimi, yalnızca yaklaşık p değerleri veren büyük
  örnek sonuçlarına dayalı sonuçlar yerine, koşullu parametrik olmayan testler için çok sayıda işlem gerektiren kesin p değerlerinin hızlı bir şekilde hesaplanmasını sağlamıştır. Ayrıca jackknife ve yinelemeli örnekleme (bootstrap) gibi yöntemler ile parametrik olmayan süreçler bir arada kullanılmıştır.

Literatürde aynı anlamda kullanılan parametrik olmayan yöntemler ve dağılımdan bağımsız yöntemler arasında fark bulunmaktadır.

Parametrik olmayan yöntemler, örneğin çekildiği anakitle belli bir dağılıma dayanmadığı ve bu anakitleye dayanan örnekleme dağılımından elde edilen test istatistiğine dayalı çıkarımlar elde edildiği için çoğunlukla dağılımdan bağımsız yöntemler olarak adlandırılır. Genellikle klasik teknik olarak kabul edilen birçok yöntem bu tanıma göre gerçekte parametrik olmayandır. Örnek olarak merkezi limit teoremine dayalı yöntemler, Chebyshev eşitsizliği ve birçok Ki-kare testi büyük örnekler için spesifik dağılım varsayımları gerektirmez (Dickinson 1993). Gerçekte parametrik olmayan terimi kesin değildir fakat dağılımdan bağımsız terimi ise kesin bir anlama sahiptir. Dağılımdan bağımsız olma özelliği, çok sayıda parametrik olmayan yöntemin kilit noktasıdır (Hollander, Wolfe 1999). Parametrik olmayan testte; istatistiksel bir yoğunluk fonksiyonu içinde, bir parametrenin değeri hakkında bir varsayımda bulunulmazken, bir dağılımdan bağımsız test; örnek çekilen anakitlenin kesin şekli hakkında bir varsayımda bulunmaz.

Bir istatistik test; teste dayalı istatistiğin örnekleme dağılımı, değişkenin anakitle dağılımından tamamen bağımsız ise dağılımdan bağımsızdır. Bu duruma örnek olarak Irwin-Fisher Kesin Testi verilebilir. (Marascuilo, McSweeney 1993).

Bir yöntemi dağılımdan bağımsız ya da parametrik olmayan olarak tanımlamak, varsayımdan bağımsız demek değildir. Gerçekte anakitle dağılımı hakkında çoğunlukla bazı varsayımlarda bulunulur. Örnek olarak, anakitle hakkında sürekli ve simetrik varsayımı yapılabilir. Bu durumda anakitlenin normal dağılım gibi belli bir dağılımdan geldiği belirtilmez fakat kesikli ve asimetrik dağılımlar analiz dışında bırakılır (Sprent, Smeeton 2007).

2.2.3. Parametrik Olmayan Regresyon Modellerinin Tahmin Yöntemleri

Parametrik olmayan regresyon yöntemlerinde $m(x)$ anakitle regresyon eğrisinin tahmin edilmesiyle bağımlı ve bağımsız değişkenler arasında ilişkinin yönünü ve boyutunu belirlemek için en küçük kareler tahmincileriyle benzerlik gösteren ancak parametrik olmayan yöntemlere dayanan tahminciler literatürde kullanılmaktadır.

Parametrik olmayan tahmincilerin elde edilmesinde öncelikle söz konusu tahminciler için marjinal yoğunluk fonksiyonlarının belirlenmesi gerekmektedir. Parametrik olmayan tahmincilere ait marjinal yoğunluk fonksiyonu:

$$\hat{m}(x) = \frac{1}{\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)} \int y \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{y_i - y}{h}\right) K\left(\frac{y_i - x}{h}\right) dy \qquad \textbf{(2.26)}$$

Eşitlik 2.26’da $K\left(\frac{x_i - x}{h}\right)$, Kernel fonksiyonunu ifade etmektedir. Kernel 1 değerine entegre olan simetrik, sınırlı ve sürekli bir K reel fonksiyonudur (Sprent, Smeeton 2007). h parametresi ise düzgünleştirme parametresidir. Bir önceki denklemde gösterilen parametrik olmayan regresyon tahmincisinin marjinal fonksiyonuna ait integral işlemlerinin çözülmesi ile çeşitli parametrik olmayan tahminciler elde edilmektedir.

2.2.4. Parametrik Olmayan Regresyonda Düzgünleştirme

Parametrik olmayan regresyon analizinde düzgünleştirme kavramı büyük önem taşımaktadır. Düzgünleştirmenin temelinde lokal olarak ortalama alma fikri yatmaktadır. Düzgünleştirici (smoother) ise; bir ya da birden fazla bağımsız değişkenin fonksiyonu olan bağımlı değişkenin sahip olduğu trendi ifade etmek için kullanılan bir araçtır ve bağımlı değişkenin kendisinden daha az değişken bir trend tahmini yapmayı amaçlamaktadır. Bir düzgünleştiricinin en önemli özelliği, değişkenler arasındaki ilişkinin biçimini kesin bir biçimde varsaymamasıdır ve bu özelliğinden dolayı parametrik olmayan regresyonda sık kullanılan bir araçtır (Hastie, Tibshirani 1990: 9). Parametrik regresyonda elde edilen doğru kesin parametrik bir biçime sahip olduğu için düzgünleştici değildir. Bilinen en basit düzgünleştirici ise hareketli ortalamalardır. Bir düzgünleştirici tarafından tahmin edilen eğriye "düzgün " adı verilmektedir. Tek bir bağımsız değişkenin olduğu durumlar ise "serpilme diyagramı düzgünleştirmesi" olarak adlandırılmaktadır. Serpilme diyagramındaki noktalar, herhangi bir olasılıksal model olmaksızın düzlem üzerindeki noktaların birikimi olarak kabul edilmektedir (Ruppert, Wand, Chao 2003: 57).

Serpilme diyagramı düzgünleştiricileri aynı serbestlik derecesinde yaklaşık aynı sonuçlar vermelerine rağmen aralarında değerlendirme yaparak karar vermede aşağıdaki kriterler kullanılabilir.

- Kullanılan bilgisayar programına uygunluk
- İlgilenilen değişkenler arasındaki muhtemel ilişkiyi ortaya çıkarabilme esnekliği
- Kullanılan yöntem sonucunda elde edilen sonuçların anlaşılabilirliği
- Kullanılan yöntemin matematik özelliklerinin analizinin kolaylığı
- Elde edilen cevapların güvenilirliği
- Kullanılan yöntemin, veriyi en etkin biçimde kullanarak sonuç elde etmesi
- Kullanılan yöntemin, kolaylıkla daha karışık düzenlemelere uyarlanabilmesi

Düzgünleştirme yöntemleri arasında basit hareketli ortalamalardan sonra en çok bilinen yöntem üstel düzgünleştirmedir. En temel üstel düzgünleştirici, üstel olarak ağırlıklandırılmış hareketli ortalama yöntemidir. Bu yöntemin özelliği belli bir modelden bağımsız olması ve düzgünleştirme faktörü olarak basit bir matematik formül kullanmasıdır. Buna bağlı olarak üstel düzgünleştirme, kısıtlayıcı bir model olmaksızın serpilme diyagramından mevcut trendi tahmin etmesi bakımından parametrik olmayan regresyonun en basit biçimine benzemektedir. Sonuç olarak; basit üstel ve ikili üstel düzgünleştirme, kernel regresyonun çeşitleridir (Gijbels, Pope, Wand 1999: 39-50).

Düzgünleştirme teknikleri 1980'li yıllardan itibaren gelişmeye başladıysa da temeli, 19. yüzyıla dayanmaktadır. 1857 yılında Saksonyalı ekonomist Engel, günümüzde regressogram olarak anılan Engelsches Gesetz isimli bir eğri oluşturmuştur. Bu tarihten itibaren parametrik olmayan düzgünleştirme yaklaşımı uzun bir süre ihmal edilmiş ve buna bağlı olarak 20. yüzyılın ilk yarısında istatistik teorisinin matematik gelişimi; hesaplamalardaki kolaylık, model varsayımlarına uyumluluk ve matematik uygunlukları nedeniyle parametrik yaklaşıma ağırlık vermiştir (Hardle 1999: 1-23).

Parametrik olmayan düzgünleştirme yöntemlerinin en önemli faydaları şunlardır:

- Veri setinin özelliklerinin gösterilmesinde uygun, az ve öz bir araç olarak kullanılabilir ve böylelikle pratik bir parametrik modelin kurulmasına yardım eder.
- Tahmin edilmiş parametrik bir modelin kontrol edilmesinde kullanılabilir.
- Tamamen parametrik olmayan bir yapıda oldukça zayıf kısıtlamaların varlığı altında çıkarım yapılabilir.
- Parametrik olmayan tahminleyenler, yarı parametrik modellerde Euclidian değerli büyüklüklerin tahminleyenlerinin oluşturulmasında kullanılabilir.

Düzgünleştirme yöntemleri kullanılarak, seçilen sürecin verdiği geçerli çıkarımlar altında yapıların sınıfları genişletilebilir. Fakat unutulmaması gereken bir nokta vardır: Merkezileştirilmiş parametrik olmayan tahminleyenler doğru olarak belirlenmiş modelde, parametrik tahminleyenlerde olduğu gibi $\sqrt{n}$ oranından daha yavaş biçimde $\sqrt{nh}$ oranında yakınsarlar. Burada $h \rightarrow 0$ olmak üzere düzgünleştirme parametresidir. Ayrıca, küçük örneklerde asimptotik dağılımları zayıf tahminlerde bulunabilirler. Ne var ki bu problem parametrik modellerin kullanıldığı durumlarda da olabilir. Aradaki fark, kalitatiften çok kantitatiftir. Merkezileştirilmiş parametrik olmayan tahminleyenler, n yerine nh konulduğunda parametrik tahminleyenler gibi davranırlar (Hardle, Linton 1994: 1-23).

2.2.5. Sık Kullanılan Parametrik Olmayan Regresyon Yöntemleri

2.2.5.1. Mood-Brown Yöntemi

Bu yöntem Mood (1950) çalışmasında tanıtılmış, Mood ve Brown (1951) çalışması ile son halini almıştır.

$$Y_i = \hat{\alpha} + \hat{\beta} X_i \tag{2.27}$$

Eşitlik 2.27 ile verilen regresyon doğrusu için α ve β parametrelerinin tahminine yönelik bu yöntemi uygulamak için öncelikle Y değerleri, X ' lerin medyan değerinden daha küçük ya da ona eşit X değerleri ile birlikte olan ve X ' lerin medyan değerinden daha büyük X değerleri ile birlikte olanlar şeklinde iki gruba ayrılır. α ve β ' nin istenen değerleri, iki grubun her birinde regresyon doğrusundan sapmaların medyanının sıfır olduğu tahmindir. α ve β parametrelerinin elde edilme adımları:

1. Örneklem verileri için serpilme diyagramı hazırlanır.
2. X ' lerin medyan değerinden geçen bir dikey doğru çizilir. Eğer bir ya da daha fazla nokta bu medyan doğrusu üzerine düşüyorsa, bu doğru gerektiği kadar sağa veya sola öyle kaydırılır ki medyanın her iki yanındaki nokta sayısı mümkün olduğunca eşit olur.
3. İkinci adımda oluşan her iki grup için X ve Y değerlerinin medyanı bulunur. Yani toplam 4 tane medyan hesaplanır.
4. Gözlemlerin birinci grubunda X ve Y 'nin medyanının kesiştiği nokta belirtilir. Benzer işlem ikinci grup gözlemler için de yapılır.
5. Dördüncü adımda belirlenen iki noktayı birleştiren bir doğru çizilir. Bu doğru istenen doğru tahminine ilk yaklaşımdır.
6. Eğer bu doğrudan dikey sapmaların medyanı her iki grupta da sıfır değilse, bu doğrunun pozisyonu her gruptaki sapmaların sıfır medyanlı olması sağlanana dek değiştirilir. Daha kesin bir doğruluk isteniyorsa, Mood tarafından önerilen alternatif yöntem kullanılabilmektedir.
7. Sonuçta elde edilen doğrunun Y ile kesişimi α katsayısını verirken, β katsayısı $b_{ij} = (Y_2 - Y_1 / X_2 - X_1)$ ile hesaplanır. (X_1, Y_1) ve (X_2, Y_2) doğru üzerindeki herhangi iki noktanın koordinatlarıdır.

Medyana Göre Parametre Tahmininde α ve β'ya ilişkin hipotezlerin testleri: kurulan regresyon eşitliği için ilgilendiği α ve β katsayılarının birisine ya da her ikisine ilişkin hipotezlerin testidir. $\alpha = \alpha_0$ ve $\beta = \beta_0$ hipotezlerinin eşanlı testleri ele alınmıştır.

Mood ve Brown tarafından önerilen bu testlerin işleyişi şöyledir (Daniel 1995: 622).

Gözlemler X ve Y sürekli değişkenlerinin n tane $(X_1, Y_1), \ldots, (X_n, Y_n)$ çiftinden oluşmaktadır. Her bir gözlem çifti (X_i, Y_i), aynı birliktelik biriminde ölçülmüştür.

Hipotezler:

$$H_0 = \alpha = \alpha_0, \beta = \beta_0$$

$$H_1 = \alpha \neq \alpha_0 \; ve/veya \; \beta \neq \beta_0$$

Test Ölçütü:

1. Verilerin serpilme diyagramı hazırlanır.
2. Serpilme diyagramında $Y = \alpha_0 + \beta_0 X$ doğrusu çizilir.
3. X değerlerinin medyanından geçecek şekilde dikey bir doğru çizilir.
4. n_1, belirtilen regresyon doğrusunun üstündeki ve X'lerin meydanından geçen dikey doğrunun solundaki veri noktalarının sayısı, n_2 ise, belirtilen regresyon doğrusunun üzerinde ve X'lerin medyanından geçen dikey doğrunun sağındaki veri noktalarının sayısı olsun. Mood, n_1 ve n_2' nin 0.5 parametresi ile binom dağıldığını belirtmiştir ve bu bilgiye dayalı olarak test ölçütü eşitlik 2.28' deki gibi izah edilmiştir.

$$\chi^2 = \frac{8}{n}\left(\left(n_1 - \frac{n}{4}\right)^2 + \left(n_2 - \frac{n}{4}\right)^2\right) \qquad \textbf{(2.28)}$$

Tate ve Clelland, (1957) n değerinin yaklaşık olarak 10 veya daha büyük olması halinde bu yaklaşımın iyi sonuçlar verdiğini söylemişlerdir.

Bu ölçüt, H_0 doğru olduğunda ve n çok küçük olmadığında iki serbestlik derecesi ile Ki- kare dağılışı gösterir.

Karar Kuralı: Eğer üstteki denklem ile bulunan hesap değeri, ilgili χ^2 tablosundan iki serbestlik derecesi için bulunan tablo değerini aşarsa H_0 hipotezi ret edilir.

$\beta = \beta_0$ Testi: Regresyon analizinde genellikle regresyon doğrusunun eğimi için $H_0: \beta = \beta_0$ testi ile ilgilenilir. Bu hipotezi $H_0: \beta \neq \beta_0$ alternatifine karşı test etmek için, Mood ve Brown' un önerdiği şu adımlar izlenir:

1. Verilerin serpilme diyagramı hazırlanır.
2. X ' lerin medyanından geçen dikey doğru çizilir.
3. $\propto, Y_i - \beta_0 X$ sapma değerlerinin medyanı ve β_0 hipotezde ileri sürülen değer olmak üzere, verilere $Y = \propto + \beta_0 X$ doğrusu uydurulur. Genellikle bu doğru en kolay şekilde $Y = \beta_0 X$ doğrusunu çizerek ve bu doğruya paralel bir doğruyu verileri iki eşit gruba bölecek şekilde çizerek belirlenir.
4. $Y = \propto + \beta_0 X$ doğrusunun üzerindeki ve X 'lerin meydanının solundaki veri noktaları sayılır ve n_1 ile gösterilir. Buna göre test ölçütü eşitlik 2.29 şeklinde açıklanır (Mood-Brown 1951).

$$\mathcal{X}_{bi}^2 = \frac{16}{n}\left(n_1 - \frac{n}{4}\right)^2 \qquad \textbf{(2.29)}$$

Bulunan bu değer tablodan bir serbestlik dereceli Ki- kare tablo değeri ile karşılaştırılır. Tablo değerini aşması durumunda H_0 hipotezi ret edilir. Tate ve Clelland (1957) n değerinin 20 veya daha büyük olması halinde bu yaklaşımın iyi sonuçlar verdiğini söylemişlerdir.

2.2.5.2. Theil-Sen Yöntemi

Theil (1950) tarafından ileri sürülen yöntem araştırmacıların en çok başvurdukları eğim bulma yöntemlerindendir. Sen (1968: 1379-1389) düzeltmesi ile birçok kaynakta Theil-Sen yöntemi olarak adlandırılır. Bir doğrunun eğimi tahmininde kullanılan Theil (1950)' in yöntemi, $(x_i, y_i), (x_j, y_j)$ gözlem çiftlerinden hesaplanan eğim değerlerinin medyanı hesabına dayandırılmaktadır (Hussain, Sprent 1983: 182-191). Sahip olunan $(x_1, y_1), \ldots, (x_n, y_n)$ n tane gözlem çiftinin;

$$Y_i = \alpha + \beta x_i + e_i \qquad i = 1,2,\ldots,n \tag{2.30}$$

eşitlik 2.30' da α ve β bilinmeyen regresyon parametreleridir. x_i değerleri birbirlerinden farklı bilinen sabitler olup $x_1 < x_2 < \cdots < x_n$ şeklinde sıralanmaktadır (Yıldız, Topal 2001: 429-435). Bu modelde e_i'ler, varyansı σ_e^2 ve medyanı sıfır olan simetrik bir sürekli dağılışa sahip bağımsız ve aynı dağılımdan meydana gelen şansa bağlı hatalardan oluşurlar (Rao, Gore 1982: 42-50). Theil yönteminde α ve β öyle tahmin edilmeli ki e_i hata terimlerinin medyanı sıfır olmalıdır (Maritz 1979: 21). β'nın tahmini $\hat{\beta}, i < j(x_i \neq x_j)$ olmak üzere elde edile $N = \binom{n}{2}$ tane $b_{ij} = (y_j - y_i)/(x_j - x_i)$ eğimlerinin tümünün bir ağırlıklı medyanı olur (Daniel 1995: 622, Wang, Yu 2004).

$$\hat{\beta} = medyan\{b_{ij}\}$$

$$\hat{\alpha} = medyan(Y) - (\hat{\beta})medyan(X_i) \tag{2.31}$$

değerlerinin medyanı olduğu belirtilmiştir (Hussain, Sprent 1983). Eğer eğim katsayısı b_{ij}' ler çift sayıda ise β_0'ı bulmak için medyanı oluşturan iki farklı *y* ve *x* ikililerinden *y*'si büyük olanlar (x_i'lerle beraber) alınır. İki farklı β_0 değeri hesaplanır. Bu iki tahminin orta noktası hesaplanarak β_0 bulunur.

Medyana göre parametre tahmininde α ve β' ya ilişkin hipotezlerin testleri kurulan regresyon eşitliği için ilgilenilen α ve β katsayılarının birisine ya da her ikisine ilişkin hipotezlerin testidir. $\alpha = \alpha_0$ ve $\beta = \beta_0$ hipotezlerinin eşanlı testleri ele alınmıştır.

Theil tarafından eğimin testi için önerilen bu yöntem Kendall'ın Tau istatistiğine dayanmaktadır (Daniel 1990: 426-443).

Varsayımlar:

I. Veriler $(x_1, y_1),...,(x_n, y_n)$ şeklindeki n tane gözlem çiftinden oluşmakta olup uygun denklem: $Y_i = \alpha + \beta X_i + e_i$, $i = 1,2,...,n$ dir. Burada X_i' ler bilinen sabitler α ve β ise bilinmeyen parametrelerdir.

II. Her bir X_i değeri için Y değerlerinin bir alt popülasyonu vardır.

III. Y_i, X_i değerindeki sürekli Y rastlantı değişkeninin gözlenen değeridir.

IV. Tüm X_i değerleri farklıdır ve $x_1 < x_2 < ... < x_n$ sırasındadır.

V. e_i 'ler karşılıklı bağımsız olup aynı sürekli popülasyondan gelirler.

Hipotezler:

Çift Yönlü	Tek Yönlü	Tek Yönlü
$H_0 : \beta = \beta_0$	$H_0 : \beta \leq \beta_0$	$H_0 : \beta \geq \beta_0$
$H_1 : \beta \neq \beta_0$	$H_1 : \beta > \beta_0$	$H_1 : \beta < \beta_0$

Test Ölçütü: Yöntem, Kendall'ın Tau ölçütüne dayanmaktadır. $(X_i, Y_i - \beta_0 X_i)$ formundaki tüm olası gözlem çiftleri kıyaslanarak Kendall'ın Tau ölçütü hesaplanmıştır (Oğuzlar 1999). Yöntem' in adımları şöyle özetlenmiştir.

1. $(X_i, Y_i - \beta_0 X_i)$ gözlem çiftleri, X değerlerine göre bir sütunda doğal sırada düzenlenmiştir.
2. Her bir $Y_i - \beta_0 X_i$ değeri, altındaki her bir $Y_j - \beta_0 X_j$ değeri ile karşılaştırılmıştır.
3. $(Y_i - \beta_0 X_i, Y_j - \beta_0 X_j)$ şeklinde doğal sırada sonuçlanan karşılaştırmaların sayısı P, tersine doğal sırada sonuçlanan karşılaştırmaların sayısına ise Q denir.
4. $S = P - Q$ olmak üzere test ölçütü $\tau = \dfrac{S}{n.(n-1)/2}$ eşitliği ile açıklanmıştır.

Hipotezlere Göre Karar Kuralı:

Çift yönlü hipotez testi için:

- Hesaplanan τ değeri pozitif olup, Kendall'ın Tau tablosundan n ve $\alpha/2$ için bulunacak değerden büyükse ya da negatif bulunan τ değeri aynı tablo değerinden küçük ise H_0 hipotezi ret edilir.

Tek yönlü hipotez testi için:

- Hesaplanan τ değeri pozitif olup, Kendall'ın Tau tablosudan n ve α için bulunacak kritik değerden büyük ise H_0 hipotezi red edilir.
- Hesaplanan negatif τ değeri, Kendall' ın Tau tablosun' dan n ve α için bulunacak kritik değerin negatifinden daha küçük ise H_0 hipotezi red edilir.

Tekrarlı Gözlemler: Eğer (x_1, y_1) ve (x_2, y_2) gözlem çifti için $(y_2 - y_1)/(x_2 - x_1) > 0$ ise uyumlu, aksi takdirde uyumsuzdur denir. $x_1 = x_2$ ise karşılaştırma yapılamaz. Eğer $x_1 \neq x_2$ iken $y_1 = y_2$ ise bu oran sıfır olur. Bu durumda gözlem çifti $1/2$ uyumlu ve $1/2$ uyumsuzdur. Bu durum τ katsayısı hesabında bir değişiklik yaratmaz. Ancak tekrar durumunda τ' nin hesabında değişiklik yapar.

Kendall'ın τ katsayısı için analiz edilen test, verilerin sürekliliğini varsayar. Ancak pratikte tekrarlar ya X, ya Y, ya da her ikisinde oluşabilir. Tekrar durumunda en basit yöntem, tekrarlı gözlemlere bunların doğal sıra sayılarının ortalamasını rank olarak atamaktır. Tekrarlar için önerilen bir yaklaşım aşağıdaki gibidir.

1. Gözlemler X 'lerin büyüklüğüne göre artan doğal sırada sıralanır.
2. X 'lerin tekrarlı gözlemler grubuna karşılık gelen Y değerleri artan sırada düzenlenir.
3. Y çiftlerinin doğal ve tersine doğal sırada sayıları tekrarsız durumdaki gibi sayılır ancak tekrarlı X 'e X_a karşılık gelen Y değeri, X_a ile tekrarlı diğer X'e eşlik eden herhangi bir Y değeri ile karşılaştırılmaz.

Tekrar durumunda $(x_i \neq x_j)$ için tüm (x_i, y_i) ve (x_j, y_j) karşılaştırmaları yapıldığında τ katsayısı, eşitlik 2.32 ile hesaplanmıştır.

$$\tau = \frac{P-Q}{P+Q} \qquad \textbf{(2.32)}$$

Bu yaklaşımın avantajı tekrar durumuna rağmen ilişki miktarını 1 veya -1 elde etme şansının olmasıdır. İlk kez Goodman ve Kruskal (1963: 310-364) tarafından tanımlanan bu ölçüte "Gamma Katsayısı'' denilmektedir. Tekrar durumunda P ve Q' nun hesabı şöyle özetlenmiştir.

Hesaplamalarda (X_i, Y_i) gözlem çiftleri X'lerin artan sırasında sıralanırsa işlem kolaylaşır. Böylece her bir Y yalnızca altındaki değerle karşılaştırılmıştır. Verilerde tekrarlar olduğunda sonuçlar kesin değil yaklaşık olur. Bu testin güç ve etkinliği ise Sen (1968: 1379-1389) tarafından incelenmiştir.

Kendall'ın Tau tabloları n'in yalnızca 40'dan küçük veya eşit değerleri için yazılmıştır. Örnek genişliği 40'dan fazla olduğunda test istatistiği eşitlik 2.33 ile hesaplanabilmektedir (Aytaç 1991).

$$Z = \frac{3\tau\sqrt{n(n-1)}}{\sqrt{2(2n+5)}} \sim N(0,1) \qquad \textbf{(2.33)}$$

Bu test istatistiği veri seti içerisinde, eşit gözlem değerleri var olduğu zaman da başarı ile kullanılabilmektedir.

2.2.5.3. Eğim Parametrelerinin Ortak Olduğu ve Sabit Parametrenin Değişkeninin Tahmin Edildiği Yöntemler

$\hat{\beta}_1$; Theil-Sen yönteminden tahmin edilen eğim parametresini göstermek üzere, aşağıda verilen iki ayrı yöntem, β_0 parametresinin tahmini için geliştirilmiş tahmin yöntemleridir:

2.2.5.3.1. Optimum Tahmin Yöntemi

$d_i = y_i - \hat{\beta}_1 x_i$ değerleri hesaplanır ve bu değerlerin ortancası $\hat{\beta}_0$ olarak hesaplanır. Bu yöntemde d_i değerlerinin simetrik dağılmış olması gerekmemektedir. Özellikle aşırı uç değerli verilerde daha başarılı sonuçlar vermektedir.

2.2.5.3.2. Hodges-Lehmann Yöntemi

Burada da, $d_i = y_i - \hat{\beta}_1 x_i$ değerleri hesaplanır ve $\hat{\beta}_0$ değeri d_i değerlerinin aritmetik ortalamasına eşittir. Bu yaklaşım d_i değerlerinin $\hat{\beta}_0$ etrafında simetrik dağıldığı varsayımını gerektirir. Aykırı gözlemlerin çok olması yöntemin gücünü düşürmektedir (Lehmann 2006).

Medyana Göre Parametre Tahmininde Eğim Katsayısı İçin Güven Aralığı

Çoğu durumda araştırmacıların ilgi odağı, eğim katsayısı β için güven aralığı bulmaktır. Mood, β için Mood-Brown hipotez testi yöntemine dayalı güven aralığı bulmada deneme ve yanılma tekniğini önermiştir (Daniel 1990: 426-443). Bu teknik, $\%(1-\alpha)$ güven aralığının, α yanılma düzeyinde ret edilemeyecek β_0 değerlerinden oluştuğu bilgisine dayanmaktadır. Bu yaklaşıma alternatif bir yaklaşım yine Mood tarafından önerilmiştir. (Yıldız, Topal, Bilgin 2004).

Bu test için yapılan varsayımlar güven aralığı araştırması için de geçerlidir. β_1' nın $\%(1-\alpha)$ güven aralığını veren yöntemin adımları şöyledir:

β için bir güven aralığı çift yönlü hipotez $H_0 : \beta = \beta_0$, $H_1 : \beta \neq \beta_0$ testiyle elde edilir. Güven aralığı $i < j$ olmak üzere düzenlenmiş b_{ij} eğimler setinde uygun olarak seçilmiş sınırlı noktalara sahiptir (Sievers 1978: 628-631).

$x_j - x_i \neq 0$ farklarının sayısı N ise, önerilen nokta tahmincisi $x_j \neq x_i$ için N tane b_{ij} eğiminin medyanı olur (Sen 1968: 1379-1389). $i < j$ olmak üzere toplam $N = \binom{n}{2}$ adet b_{ij} değeri küçükten büyüğe doğru sıralanır. β'nın güven aralığının alt sınırı $\hat{\beta}_a$ küçükten büyüğe doğru sıralanan k'ncı b_{ij} değeridir. β'nın güven aralığının üst sınırı $\hat{\beta}_a$ ise büyükten küçüğe doğru sıralanan k'ncı b_{ij} değeridir (Daniel 1990: 426-443). Burada k değeri, eşitlik 2.34' de gösterildiği gibidir.

$$k = \frac{N - S_{\alpha/2} - 2}{2} \qquad \textbf{(2.34)}$$

$S_{\alpha/2}$ değerini belirlemek için, n ve $\alpha/2$' ye göre Kendall' ın Tau test istatistik değerleri tablosuna bakılır. Bulunan değer yukarıdaki eşitlikte yerine konarak k değeri tespit edilir.

Eğer n değeri çok büyük $(n \geq 30)$ ise, dağılış normal dağılışa yaklaşır. Buna göre standart normal dağılım tablosu kullanılarak k değeri eşitlik 2.35'teki gibi hesaplanır (Griffin 1962: 259-273).

$$k = \frac{N - Z_{\alpha/2}\sqrt{\dfrac{n(n-1)(2n+5)}{18}}}{2} \tag{2.35}$$

2.2.5.4. K-En Yakın Komşuluk Tahmin Yöntemi

Parametrik olmayan regresyon analizinde fonksiyon analiz yöntemlerinden en yakın komşuluk (k-nearest neighbor, K-NN) tahmin yöntemi ele alınmıştır. İlk incelenmeye başlanmasından bu yana uzun yıllar geçmiş olmasına karşın son yıllarda tekrar yoğun bir şekilde kullanılmaya başlanmıştır.

Son yıllarda parametrik olmayan yaklaşım, parametrik eşitlik kurmada yardım amacı ile ya da K-NN olmak üzere çeşitli parametrik olmayan tekniklerden tutarlı tahminler elde edildiğinden parametrik regresyon analizine bir alternatif olarak gündeme gelmiştir. (Kıroğlu 2001).

Parametrik olmayan regresyon fonksiyon analizinde en basit gösterimi ile (X_i, Y_i) şeklinde verilen n genişlikli veri seti için değişkenler arasındaki ilişki yapısı eşitlik 2.36' daki düzgün fonksiyonu ile ifade edilmektedir.

$$Y_i = f(X_i) + e_i \qquad (i = 1,2,\ldots,n) \tag{2.36}$$

Burada hata terimi e_i' lerin bağımsız, sıfır ortalamalı olmaları dışında hiçbir kısıtlayıcı varsayım bulunmamaktadır.

Parametrik olmayan regresyon fonksiyon yine bağımlı ve bağımsız değişkenin ilişkisi araştırılmıştır. Parametrik regresyon analizinden farkı, tahminlerin parametrik olmayan yöntemlerle yapılmasıdır.

Parametrik olmayan regresyon fonksiyonunun tahmin yöntemlerinden K-NN tahmin yöntemi parametre tahmin hesabına dayanmaktadır.

Bu çalışmada, parametrik olmayan regresyon fonksiyonunu tahmin yöntemlerinden parametre tahmin hesabı yapan K-NN tahmin yöntemi anlatılmıştır.

Bu yönteme kısaca yakın komşuluk denilmesinin nedeni, yoğunluk tahmini belli bir noktanın en yakın komşularından çok yakın komşu noktalarına dayanır (Silverman 1986: 898-916).

K-NN regresyon fonksiyon tahmini $\hat{m}$ aşağıdaki gibi tarif edilir (Györfi 1981, Devroye 1978: 142-151).

$$\hat{m}_k(x) = n^{-1}\sum_{i=1}^{n} W_{ki}(x) Y_i \qquad \textbf{(2.37)}$$

Eşitlik 2.37'de $\{W_{ki}(x)\}_{i=1}^{n}$, J_x indekslerinin kümesi vasıtasıyla tarif edilmiş bir ağırlık kümesidir.

$$J_x = \{\, i : X_i\ x'\text{e en yakın gözlemlerin bir tanesi}\} \qquad \textbf{(2.38)}$$

Komşu gözlemlerin indeksleri kullanılarak K-NN ağırlık kümesinin düzenlenmiş hali şu şekildedir.

$$W_{ki}(x) = \begin{Bmatrix} (n/k, & i \in J_x ise \\ 0, & (i \notin J_x) \end{Bmatrix} \qquad \textbf{(2.39)}$$

Eşitlik 2.39'da,

n: gözlem sayısını

k: değişken komşulukta bir ağırlık noktasındaki veri sayısını ifade eder.

(Stone 1977: 595-645, Devroye 1978: 142-151)'e göre, K-NN ağırlık fonksiyonları şu şekilde tanımlanabilir:

Üniform K-NN ağırlık fonksiyonu;

$$W_{ki}(x) = \begin{Bmatrix} (1/k, & 1 < i < k \\ 0, & i > k \end{Bmatrix} \qquad \textbf{(2.40)}$$

Üçgen (triangular) K-NN ağırlık fonksiyonu;

$$W_{ki}(x) = \begin{Bmatrix} (k-i+1)/b_k, & 1 < i < k \\ 0, & i > k \end{Bmatrix} \qquad \textbf{(2.41)}$$

Eşitlik 2.41'de $b_k = k(k+1)/2$ dir. Quadratik K-NN ağırlık fonksiyonu;

$$W_{ki}(x) = \begin{cases} (k^2 - (i-1)^2)/b_k, & 1 < i < k \\ 0, & i > k \end{cases} \qquad \textbf{(2.42)}$$

Eşitlik 2.42'de $b_k = k(k+1).(4k-1)/6$'dır.

$W_{ki}(x)$, Uniform ağırlıklar ile tayin edilen komşuluk vasıtasıyla EKK doğrusunun α ve β_i parametreleri aşağıdaki şekilde hesaplanmıştır (Green, Jennison, Seheult 1985: 299-315, Hardle 1997).

$$\alpha_{xi} = \hat{m}_k(x_i) - \beta_i \overline{\mu}_{xi} , \qquad \textbf{(2.43)}$$

$$\beta_{xi} = \frac{C_{xi} - \overline{\mu}_{xi}.m_k(X)}{V_{xi} - \overline{\mu}_{xi}^2} , \qquad \textbf{(2.44)}$$

$$\overline{\mu}_x = k^{-1} \sum_{i=j_x} X , \qquad \textbf{(2.45)}$$

$$C_x = \sum_{i=j_x} X_i Y_i , \qquad \textbf{(2.46)}$$

$$V_s = \sum_{i=j_x} X^3 \qquad \textbf{(2.47)}$$

Yukarıdaki eşitliklerde;

$\alpha_{xi} = W_{ki}(x)$ ağırlıklara göre elde edilen regresyon doğrusunun y eksenini kestiği nokta,

$\beta_{xi} = W_{ki}(x)$ ağırlıklara göre elde edilen regresyon doğrusunun eğimi,

$M_x = i \in j_x$ şartına uygun x_i değerlerinin ortalaması

$C_x = i \in j_x$ şartına uygun x_i ve y_i değerlerinin çarpımlı toplamı

$V_x = i \in j_x$ şartına uygun x_i değerlerinin karelerinin toplamıdır.

K-NN yöntemi, gözlemleri ayrıştırmaya yönelik diskriminant analizi için etkin olarak kullanılmıştır. k değeri hem değişken sayısına hem de olasılık fonksiyonunun düzgünlüğüne dayanmalıdır. Bu tahminci yardımı ile ayrıştırma pratikte denemeye ve farklı k değerleri için hata değeri ve gözlemlerin yanlış sınıflandırılma oranına bağlıdır. Bu durumda tüm gözlemler için uzaklıklar hesaplanmalıdır. Hesaplama miktarını azaltmanın bir yolu gözlem sayısını azaltmaktır. Diğer bir yol ise (Fukunaga, Narendra 1975: 750-753) tarafından önerilen ‘’branch ve bound ‘’ algoritmasıdır (Hand 1986).

Parametrik Basit Doğrusal Regresyonun Varyans Analizi: parametrik basit doğrusal regresyon analiz yöntemlerinin varyans analiz tablosu çizelgede belirtildiği gibi yapılmıştır.

Çizelge 1. Varyans Analizi Tablosu

Varyans Kaynağı	**Serbestlik Derecesi**	**Kareler Toplamı**	**Kareler Ortalaması**	**F**
Regresyon	$k-1$	$\sum_{i=1}^{n}\left(\hat{Y}_i-\overline{Y}\right)^2$	$\sum_{i=1}^{n}\left(\hat{Y}_i-\overline{Y}\right)^2 / k-1$	
Hata	$n-k$	$\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2$	$\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2 / n-k$	*R.K.O/H.K.O*
Genel	$n-1$	$\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2$		

Çizelge 2. Varyans Analizi Tablosu (Medyan Testine Göre)

Varyans Kaynağı	**Serbestlik Derecesi**	**Kareler Toplamı**	**Kareler Ortalaması**	**F**
Regresyon	$k-1$	$\sum_{i=1}^{n}\left(\hat{Y}_i-Y_{med.}\right)^2$	$\sum_{i=1}^{n}\left(\hat{Y}_i-Y_{med.}\right)^2 / k-1$	
Hata	$n-k$	$\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2$	$\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2 / n-k$	*R.K.O/H.K.O*
Genel	$n-1$	$\sum_{i=1}^{n}\left(Y_i-Y_{med.}\right)^2$		

n: Gözlem sayısını
k: Parametre sayısını ifade etmektedir.

Parametrik olmayan basit doğrusal regresyonun medyana göre parametre tahmininin varyans analizi: parametrik olmayan basit doğrusal regresyon analizinin medyana göre parametre tahmin yöntemlerinin varyans analiz tablosu Çizelge 2'de belirtildiği gibi yapılmıştır.

2.3. Yarı Parametrik Regresyon Analizi

Yarı parametrik regresyon modelleri bağımlı değişkenin bağımsız değişkenlerden bazıları ile ilişkisinin belirlendiği fakat diğer bağımsız değişken veya değişkenlerle ilişkisinin kolayca belirlenemediği modellerdir. Değişkenlerin skaler olduğu ve f fonksiyonunun parametrik bir aile içerisinde bulunmadığı 'yarı parametrik regresyon modeli',

$$y = E(y \mid Z, x) + \varepsilon \qquad \textbf{(2.48)}$$

$$y = Z\beta + f(x) + \varepsilon \qquad \textbf{(2.49)}$$

biçimindedir. Eşitlik 2.49'daki y, Z, f ve ε önceden tanımlandığı gibidir. z parametrik değişkenler vektörünün parametrik olmayan $x = (x_1, x_2, ..., x_n)$ değişkeni ile düzgün bir regresyon ilişkisine sahip olduğu varsayılır (Yatchew 2003) ve bu modelle parametre tahmini yapılırken normallik varsayımına gerek duyulmaz (Zeger, Diggle 1994).

Parametrik olmayan model tahmininde yorumlanabilir sonuçlar elde etmek için en fazla iki açıklayıcı değişken ile çalışmak mümkün iken, yarı parametrik yöntemde k tane açıklayıcı değişkenin bağımlı etkisini incelemek mümkündür. Ayrıca parametrik modeldeki kadar varsayım yapılmaması nedeni ile bu yaklaşımın uygulamalı çalışmalarda kullanılması önerilmektedir (Horowitz 1993: 58).

BÖLÜM 3

3. BOOTSTRAP YÖNTEMİ VE ÇEŞİTLERİ

3.1. Bootstrap Yöntemi

Yeniden örnekleme yöntemi olarak ortaya çıkan bootstrap yöntemi, bir başka yeniden örnekleme yöntemi olan Jacknife yöntemine alternatif olarak söz konusu yöntemden daha kolay uygulanabilir ve çok daha güvenilir olduğu belirtilerek ileri sürülmüştür Efron (1979: 1-26). Geleneksel parametrik sonucun dağılım varsayımlarını ve matematiksel analizden daha büyük oranda hesaplanmayı içeren bootstrap, istatistiksel sonuca bir yaklaşım olarak kullanılmaktadır (Mooney 1996: 570). Bootstrap, örneklemden elde edilen ortalama ve standart hatalarının hesaplanması ve güven aralıklarının oluşturulması amacıyla geliştirilmiştir (Schenkar 1985: 360). Yöntemin önemi, gözlenen örnek verilerinden hareketle, tahminin standart hatasını minimuma indirerek popülasyon parametrelerinin tahmin edildiği istatistiksel yöntemler aşamasında ortaya çıkmaktadır.

Örneklem verilerinin yeniden örneklenmesi mantığına dayanan bu yaklaşımda; örneklemin elde edildiği anakitlenin dağılımı konusunda bilgi varsa, örneğin normal dağıldığı biliniyorsa bu durumda "Parametrik Bootstrap" söz konusu olur. Ancak; özellikle, anakitlenin normal dağıldığı ya da örneklem hacminin en az 30 olması halinde Merkezi LimitTeoremi'nin geçerli olduğu biçimindeki varsayımlar (Aytaç 1991:) ihlal edilmişse bu durumda dağılımlar üzerine varsayımlar gerektirmeyen "Parametrik Olmayan Bootstrap" kullanılır (Stoffer, Wall 1991: 1024).

Parametrik bootstrap, bir istatistiğin örnekleme için önsel varsayımları gerektirirken (örneğin bir istatistiğin normal olarak dağıldığı varsayılırsa, onun örnekleme dağılımını belirlemek için iki parametre gerekir.), parametrik olmayan bootstrap, örneklem değerlerini kullanarak bir istatistiğin örnekleme dağılımının bir tahminini verir (Mooney 1993: 570).

Literatüre bakıldığı zaman birkaç farklı bootstrap yöntemine rastlanılabilir. Bunlar, basit bootstrap, çift bootstrap, ağırlıklı bootstrap, tekrarlamalı bootstrap, doğal (wild) bootstrap, ardışık bootstrap ve daha birçoğu sayılabilmektedir (Şengün 1999: 1030).

Model seçim kriteri olarak tahmin hatasının hesaplanmasında da kullanılan bootstrap yöntemi, regresyon modellerinde parametrelerin tahmin edilmesinde uygulanan hata terimlerinin yeniden örneklenmesi ve gözlem değerlerinin yeniden örneklenmesine dayanmaktadır. Bu çalışmada sadece hata terimlerinin yeniden örneklenmesine dayanan bootstrap konusuna değinilecektir.

$$F \rightarrow \left(x_1, x_2, ..., x_n\right) \tag{3.1}$$

Her bir gözlemin $1/n$ olasılığıyla seçilebileceği dağılıma ampirik dağılım denir ve F olarak eşitlik 3.1'deki gibi gösterilecektir.

Bootstrap yöntemi bootstrap örneklerine bağlıdır. F seçilmiş bir ampirik dağılım olduğu varsayılırsa bootstrap örneği gözlemlenmiş veri setinden iadeli olarak rasgele seçilmiş olan bir örnek olur ve eşitlik 3.2'deki gibi gösterilir.

$$X^* = \left(X_1^*, X_2^*, ..., X_n^*\right) \tag{3.2}$$

Sağ üst köşede olan yıldız işareti gözlemlenmiş gerçek değerlerin içerisinden iadeli seçim yapılarak oluşturulan örneği temsil etmektedir. Şekil 3.1'de bootstrap yönteminin mantığını göstermektedir.

GERÇEK HAYAT	BOOTSTRAP
Bilinmeyen Anakitle Dağılımı — Gözlenen Rasgele Örnek	Ampirik Dağılım — Bootstrap Örneği
$F \rightarrow x = \left(x_1, x_2, ..., x_n\right)$	$\hat{F} \rightarrow x^* = \left(x_1^*, x_2^*, ..., x_n^*\right)$
$\downarrow$	$\downarrow$
$\hat{\theta} = s(x)$	$\hat{\theta}^* = s(x^*)$
İlgilenilen istatistik	Bootstrap tekrarı

Şekil 3.1. Bootstrap Tekniği

Şekil 3.1'de tek örnekli bir problem için bootstrap tekniğinin nasıl uygulandığı gösterilmektedir. Sol tarafta gerçek hayatta karşılaşılan ve dağılımının ne olduğu bilinmeyen F dağılımı için gözlemlenmiş olan veri seti kullanılarak elde edilmeye çalışılan istatistik değeri gösterilmektedir. Sağ tarafta ise bootstrap tekniği kullanılarak gözlemlenmiş olan değerler kullanılarak yaratılan veri setlerinden elde edilen istatistik değeri gösterilmektedir.

Gerçek hayatta tek bir veri seti ile karakteristik tahmin edilmeye çalışılırken, bootstrap tekniğinde iadeli rasgele seçim uygulanarak elde edilen istenilen sayıdaki veri setinden karakteristik tahmin edilmeye çalışılmaktadır (Efron 1979: 1-26).

Populasyon parametresinin tahmin edicisi olan $\hat{\theta}$'nın örnekleme dağılımının oluşturulmasının amacı, söz konusu populasyon parametresinin tahmin edilmesi ya da test edilmesidir. Ancak, teorik olarak mümkün olan bu yöntemin uygulanabilirliği konusunda kuşkular bulunmaktadır. Tahmin edicinin örnekleme dağılımını oluşturmak imkânsız olmasa da son derece güç ve zaman alıcı bir iştir. Ancak, tahmin edicinin deneysel örnekleme dağılımını oluşturmak amacıyla ortaya atılan bootstrap yöntemi bu sakıncayı ortadan kaldırmaktadır. Bu mantık doğrultusunda gerekli bootstrap algoritması, aşağıdaki biçimde tanımlanabilir (Fox 1997: 501).

1) Populasyondan n hacimlik bir örneğin elde edilmesi ve bu örnek kullanılarak populasyon parametresinin tahmin edicisinin hesaplanması.
2) Her bootstrap örneği için ilgilenilen tahmin edicinin hesaplanması.
3) Belirli sayıda örnekten hareketle bu tahmin edicilerin örnekleme dağılımının elde edilmesi.
4) Elde edilen bu dağılımdan, dağılımla ilgili ortalama, standart sapma ve standart hata gibi önemli tahmin ediciler ile parametre tahmin değerlerinin elde edilmesi.
5) Sonuçta bu tahminler kullanılarak populasyon hakkında yorumların yapılması.
6) Elde edilen bu örnek populasyon ile ilgili başka hiçbir bilgi olmadığından bu populasyonun tek en iyi tahmin edicisi kabul edilir. Bu nedenle bu örnek populasyon gibi kabul edilerek her defasında iadeli seçimle her bir gözlemin örneğe girme olasılığını $1/n$ alarak n hacimlik bir örneğin yeniden elde edilmesi ve bu sürecin B kez tekrarlanması;

Örnek $\Rightarrow$

1.Bootstrap örneği

2.Bootstrap örneği

.

.

.

B.Bootstrap örneği

Şekil 3.2. Bootstrap Algoritması

Şekil 3.2'deki algoritma bootstrap yönteminin mantığını genel olarak açıklayan bir algoritmadır. Yeniden örnekleme sayısı olan B, uygulamaya bağlıdır. Aslında n hacimlik bir örnekten teorik olarak n^n sayıda bootstrap örneği oluşturmak mümkünse de bu hem gereksizdir hem de zaman kaybına neden olmaktadır (Topuz 2002).

3.1.1. Tek Değişken İçin Standart Hata Takdiri

Elde dağılımı bilinmeyen rasgele bir örnek olduğu düşünülürse $X = (x_1, x_2, ..., x_n)$ dağılımında bilinmeyen bu gözlemler için standart sapmanın hesaplanmasında bootstrap yöntemi kullanılırken aşağıdaki algoritma takip edilmektedir (Efron 1979: 1-26).

(1) B tane birbirinden bağımsız bootstrap örneği , $x^{*1}, x^{*2}, ..., x^{*B}$, her bir örneklem n tane gözlem içermektedir.

(2) Her bir örnek için standart sapma hesaplanmaktadır. $\hat{\theta}^*(b) = s(x^{*b})$ $b = 1,2,...,B$ standart sapmadır.

(3) Standart hata ise her bir standart sapma kullanılarak eşitlik 3.3'deki gibi hesaplanacaktır.

$$se'_B = \left\{ \sum_{b=1}^{B} \left[\hat{\theta}^*(b) - \hat{\theta}^*(.) \right]^2 \Big/ B-1 \right\}^{1/2} \qquad \textbf{(3.3)}$$

3.1.2. İki Örnekli Veri Setinde Bootstrap Tekniği

Tek değişken olduğu durumda bootstrap tekniğini kullanmak oldukça kolay iken iki değişkenin olduğu veri setinde bootstrap tekniğini kullanırken daha dikkatli olunması gerekecektir. $P \rightarrow x$ gösteriminden bilinmeyen bir ihtimal modeli olan P

’nin gözlenen veri seti X anlaşılmalıdır. P’nin F ve G gibi iki tane ihtimal dağılımından oluştuğu düşünülecek olursa $P(F,G)$ olarak ifade edilecek ve $P \to x$ ifadesindeki X veri setini (z,y) şeklinde düşünülmesi gerekecektir. z veri seti F dağılımından y veri seti ise G dağılımından gelecektir. Bootstrap tekniğini bu tür bir veri setinde kullanmak için veri setleri (z,y) ayrı ayrı düşünülüp kendi bootstrap örnekleri oluşturularak daha sonradan birleştirilmesi sonucu X veri seti elde edilecektir.

$P:(F,G)$ olarak ifade edilecek ve $X^* : (z^*, y^*)$ böylece oluşturulacak her bir veri seti aşağıdaki gibi ifade edilecektir.

Doğrusal regresyon denklemi elde edilerek, $x = x_0$ olmak üzere Y’nin tahmin değeri, $\hat{Y}(x_0) = x_0\hat{\beta}$ şeklinde ifade edilsin. Y değeri ve $\hat{Y}$ tahmin değeri arasındaki hata, $Q[Y,\hat{Y}] = (Y-\hat{Y})^2$ ile hesaplanır.

$Y_i - \hat{Y}_i$, $Q[Y_i,\hat{Y}_i] = 0$, $Y_i = \hat{Y}_i$ ise $Q[Y_i,\hat{Y}_i] = 1$ olmaktadır (Efron 1983: 316).

Gerçek hata oranı , $\hat{Y}(Y_0)$ için, $S(X,F) = E_F\left[Q\left(Y_0,\hat{Y}(x_0)\right)\right]$’dir.

Gerçek hata oranı, eşitlik 3.4’teki gibi tahmin edilebilir.

$$S(X,\hat{F}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} Q[Y_i,\hat{Y}(x_i)]} = \sqrt{\sum_{i=1}^{n}\left[Y_i - \hat{Y}(x_i)\right]^2 \Big/ n} \qquad \textbf{(3.4)}$$

Tahmin hatasının bootstrap yöntemi ile elde edilmesi hata terimlerinin yeniden örneklenmesine dayandığından, $X^* = \{(X_1,Y_1^*),\dots,(X_n,Y_n^*)\}$ bir bootstrap örneklemi ile hata oranı, eşitlik 3.5’teki gibi elde edilir.

$$S(X^*,\hat{F}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} Q[Y_i,\hat{Y}^*(x_i)]} \qquad \textbf{(3.5)}$$

Eşitlik 3.5’te $\hat{Y}^*(x_i), x = x_i$ değerine karşılık gelen tahmin değeridir. Bu yaklaşım, sadece tek bir bootstrap örneklemi ile ilgili olduğundan oldukça değişkendir. Bunun yerine ortalama tahmin hatası hesaplanabilir.

$$E_{\hat{F}}\left[S\left(X^{*},\hat{F}\right)\right]=E_{\hat{F}}\sqrt{\sum_{i=1}^{n}Q\left[Y_i,\hat{Y}^{*}(x_i)\right]\Big/ n} \tag{3.6}$$

Eşitlik 3.6'daki ifade, sonsuz sayıda bootstrap örneklemine ilişkin bir ideal bootstrap tahminidir. Sonlu sayıda (B) bootstrap örneklemi ile bu değer, eşitlik 3.7 ile hesaplanır (Efron 1983: 316).

$$E_{\hat{F}}\left[S\left(X^{*},\hat{F}\right)\right]=\sqrt{\frac{1}{B}\sum_{b=1}^{B}\ \sum_{i=1}^{n}Q\left[Y_i,\hat{Y}^{*}(x_i)\right]^{2}\Big/ n} \tag{3.7}$$

3.2. Bootstrap Yönteminin Çeşitleri

3.2.1. Çift (Double) Bootstrap Yöntemi

Çift bootstrap yöntemi, güven aralıklarının hesabında uygulanan bir iterasyondur.

Güven aralıkları hesaplanırken kullanılan yüzdeler yönteminde, bulunan bootstrap güven aralıklarına tekrar bootstrap uygulanarak daha doğru tahminlerin elde edildiği görülmüştür (Davison, Hinkley 1997: 223). Efron (1983: 316) tarafından geliştirilen yöntem, bir iterasyon temeline dayanır. B, orijinal örnekten yapılan bootstrap örneklerinin sayısını göstermek üzere; bootstrap iterasyonu toplam olarak β^2 kadar bootstrap örneğinin yaratılmasını gerektirir (Chernick 1999).

Yaratılan bu örnekler üzerinden uygulanan bootstrap yöntemine göre güven aralıkları elde edilir. Güven aralıklarını elde etmede kullanılan bootstrap örneklerine, tekrar bootstrap yöntemi uygulanarak bir adım daha ileri gidildiğinde çift bootstrap yöntemine ulaşılır.

Böylece daha doğru güven aralıkları elde edilir. Çift bootstrap yöntemini açıklamak için öncelikle iterasyon işleminden bahsetmek gerekir. Iterasyon teorisinin temel yaklaşımları Hall (1990) tarafından verilmiştir. Bootstrap iterasyonu için öncelikle n tane gözleme sahip $x=\left(x_1,x_2,\ldots,x_n\right)$ orijinal örneğinden n büyüklüğünde ve $\left(x_1^{*},x_2^{*},\ldots,x_n^{*}\right)$ ile gösterilen bir bootstrap örneği çekildiği düşünülsün. Örneklenen anakitlenin θ parametresi için elde edilen $(1-\alpha)$ düzeyindeki güven aralığını I_0 ile gösterilsin. Örneğin, I_0 Efron (1983) yüzdeler metoduna göre θ parametresi için elde edilen $(1-\alpha)$ düzeyindeki güven aralığı olabilir.

I_0'ın orijinal örnek x ve güven düzeyi $(1-\alpha)$'ya bağımlı olduğu $I_0(\alpha/x)$ şeklindeki bir ifade ile gösterilebilir. $I_0(\alpha/x)$ aralığının gerçek ilişkisi $\pi_0(\alpha)$ ile ifade edilebilir.

$$\pi_0(\beta_\alpha) = P\{\theta \in I_0(\beta_\alpha / x^*)\} = 1 - \alpha \tag{3.8}$$

eşitlik 3.8 ifadesinin çözümü ve $I_0(\beta_\alpha/x^*)$, orijinal örnek yerine bootstrap yeniden örneğini kullanarak hesaplanan I_0'ın bir versiyonu olmak üzere, Hall (1990: 415-419) yeniden örnekleme prensibine göre, $I_0(\beta_\alpha/x^*)$ kullanılarak I_0 ile verilen ifadeden daha iyi güven aralıkları elde etmek mümkündür. Burada β_α, bir önceki formüldeki β_α' nın tahminidir ve θ yerine θ^*, x yerine x^* yazmak suretiyle elde edilir. İterasyonu tekrarlamak için, I_0' ın yerine elde edilen yeni aralık kullanılır ve aynı yöntem bu yeni aralığa uygulanır. Bootstrap iterasyonu tek bir adım içerdiğinde de çift bootstrap adını alır (Chernick 1999).

3.2.2. Bayesyan Bootstrap Yöntemi

Deneysel dağılımı $\hat{F}$ olan ve F dağılımına sahip X tesadüfi değişkeninin n tane bağımsız ve aynı şekilde dağılmış $(x_1, x_2, ..., x_n)$ değerlerinin gözlemlendiğini, ayrıca F deneysel dağılımından iadeli olarak bootstrap örneklerinin oluşturulduğunu varsayılsın. F dağılımının bir parametresini de φ ile gösterilsin. x_i tek boyutlu ve φ 'da tek bir parametre olarak ele alındığında ikisi birlikte çok boyutlu olacaklardır. φ, φ'nın $x_1, x_2, ..., x_n$ 'e dayalı bir tahminidir. Her bir x_i 'nin iadeli olarak $1/n$ olasılığı ile örneklemeye tabi tutulması yerine, x_i için sonsal olasılık dağılımı, her bir x_i için $(1/n)$'de yoğunlaşır, fakat bir Bayesyan bootstrap tekrarından diğerine değişiklik gösterir.

Bayesyan bootstrap tekrarları tanımlanırken izlenecek yol aşağıdaki gibi sıralanabilir:

- $[0,1]$ aralığından $n-1$ tane uniform tesadüfi değişken çekilir.
- Bu değişkenlerin değerleri artan bir sıra içinde $u_1, u_2, \ldots, u_{n-1}$ ile gösterilir.
- $u_0 = 0$ ve $u_1 = 1$ olsun. Bu durumda $i = 1,2,3,\ldots,n$ için $g_i = u_{(i)} - u_{(i-1)}$ şeklinde eşitlik 3.9 uniform sıralı istatistikleri arasındaki farkı gösteren bir değişken tanımlanır.

$$g = \begin{bmatrix} g_1 \\ g_2 \\ \bullet \\ \bullet \\ \bullet \\ g_n \end{bmatrix} \quad \textbf{(3.9)}$$

$x_1, x_2, \ldots, x_n$ örneğinden iadeli seçimle n gözlemden oluşan örnek oluşturulur. Bir çekim işleminde $1/n$ olması gereken olasılık bu kez x_1 için g_1, x_2 için g_2 olarak belirlenir. Bu işlem bu şekilde sürdürülür. İkinci bir Bayesyan bootstrap tekrarı aynı şekilde yaratılır fakat bu kez hem $n-1$ tane yeni bir uniform tesadüfi değişken kümesi hem de yeni bir g_i kümesi kullanılır.

φ'nin tahmin edilmiş sonsal dağılımına dayalı φ parametresi hakkında genel Bayesyan türünde bilgiler elde etmek için kullanılabiliyor olması Bayesyan bootstrap yönteminin avantajıdır.

g_i'nin ilk Bayesyan bootstrap tekrarındaki değeri $g_i^{(1)}$, ikinci tekrardaki değeri $g_i^{(2)}$ ile olarak gösterildiğinde uniform sıradaki istatistikler için aşağıdaki sonuçlara ulaşmak mümkündür:

$$E\left(g_i^{(1)}\right) = E\left(g_i^{(2)}\right) = 1/n \quad \textbf{(3.10)}$$

$$Var\left(g_i^{(1)}\right) = Var\left(g_i^{(2)}\right) = (n-1)/n^3 \quad \textbf{(3.11)}$$

$$C\left(g_i^{(1)}, g_j^{(2)}\right) = C\left(g_i^{(2)}, g_j^{(1)}\right) = -1/(n-1) \quad \textbf{(3.12)}$$

Yukarıdaki eşitliklerde $E(.)$ $Var(.)$ ve $C(.)$ sırasıyla; beklenen değer, varyans ve korelasyonu gösterir. Yukarıda gösterilen özelliklerden dolayı, φ' nin bootstrap dağılımı ve φ' nin Bayesyan bootstrap sonsal dağılımı birçok uygulamada benzerlikler göstermektedir (Chernick 1999).

3.2.3. Düzeltilmiş (Smoothed) Bootstrap Yöntemi

F dağılımı hakkında hiçbir varsayım yapılmadığı takdirde, $\hat{F}$ 'nın F 'nin En Çok Olabilirlik tahmin edicisi olduğunu vurgulamak gerekir. Yani, F 'nin parametrelerinin bootstrap tahminleri, o parametrelerin parametrik olmayan En Çok Olabilirlik tahminleri olarak düşünülebilir. Birçok uygulamada, $\hat{F}$ yerine F Gausyan Kernel Yoğunluk tahminine dayalı düzeltilmiş dağılıma ilişkin değerler konularak işlem yapılır.

Daha düzgün tahminler elde etmek için Gausyan Kernel Yoğunluk tahmininin kullanılması uygun olur. $x_1, x_2, ..., x_n$ orijinal veri kümesini göstermek üzere, Gausyan Kernel Yoğunluk tahmini aşağıdaki gibidir.

$$\hat{f}(t;h) = \frac{1}{nh}\sum_{1}^{n} \varphi\left(\frac{t-x_i}{h}\right) \quad \textbf{(3.13)}$$

Eşitlik 3.13'te $\varphi(t)$, standart normal yoğunluğu göstermekte olup h parametresi, pencere büyüklüğü olarak adlandırılır ve verilere uygulanan düzeltme sayısı olarak tanımlanır. h değerleri arttıkça, daha düzgün bir yoğunluk tahmini elde etmek mümkündür. Yukarıdaki formülasyon, her bir x_i noktası etrafında odaklanan Gausyan yoğunluk fonksiyonu olarak düşünülebilir. Belirlenen pencere büyüklüğüne göre ortaya çıkan yoğunluk fonksiyonu grafiklerine bakılarak, verilerin modu belirlenir. Verilere ilişkin yapılacak yorum ise tamamen seçilen h değerine bağlı olacaktır.

Hipotez testi yaparken h değerini belirlemede oldukça sık kullanılan bir yaklaşım söz konusudur. Bu yaklaşım şu şekilde açıklanabilir: h'nin değeri artarken, Gausyan Kernel Yoğunluk tahminine göre elde edilen modların sayısı artmaz. h artarken, modların sayısı azalacağı için, h değerine ilişkin bir tek minimum vardır. Bu durumda

$\hat{f}(t;h)$'ın tek modlu olduğu söylenir ve $\hat{h}_1$ ile gösterilir. Uygulamalarda $\hat{h}_1$ 'nın yaklaşık değerinin 0.0068 olduğu kabul edilmektedir. Bir üst formülde küçük bir düzenleme yapıldığında h'nin minimum değerine ilişkin gösterim $\hat{f}(.;\hat{h}_1)$ şeklinde olacaktır.

Yukarıdaki formül yeniden ölçeklendirildiğinde, ölçeklendirilmiş tahmin $\hat{g}(.;\hat{h}_1)$ ile gösterilebilir. Bu noktada bir test istatistiğine ihtiyaç duyulur. Bir moda sahip yoğunluk tahmininden ortaya çıkan en küçük pencere büyüklüğünün $\hat{h}_1$ olduğu hatırlanabilir. $\hat{h}_1$'in aldığı büyük değer ise, düzeltmenin bir tek moda sahip tahminini elde etmek için yapılması gerektiğini gösterir. Ayrıca H_0 :modların sayısı =1 bootstrap hipotez testi, elde edilen anlamlılık düzeyine dayalı olarak aşağıdaki gibi ifade edilir:

$$ASL_{BOOT} = \Pr ob_{\hat{g}(.;\hat{h}_1)}\left\{\hat{h}_1^* > \hat{h}_1\right\} \quad \textbf{(3.14)}$$

Eşitlik 3.14'te $\hat{h}_1$'nın değeri 0.0068'de sabitlenmiştir. Bootstrap örneği olan $x_1^*, x_2^*, ..., x_n^*$ $\hat{g}(.;\hat{h}_1)$' dan çekilir ve $\hat{h}_1$, $x_1^*, x_2^*, ..., x_n^*$ bootstrap örneğinin bir tek moda sahip yoğunluk tahmininden ortaya çıkan h'nin en küçük değeridir. ASL_{BOOT} değerini hesaplayabilmek için yeniden ölçeklenmiş yoğunluk tahmini $\hat{g}(.;\hat{h}_1)$'dan bootstrap örnekleri çekmek gereklidir. Yani, verilerden iadeli örnekleme yapmak yerine, anakitlenin düzgün bir tahmininden örnekleme yapılmalıdır. Yapılan bu işleme düzeltilmiş bootstrap denir. Gausyan Kernel tahminine uygunluğundan dolayı, $\hat{g}(.;\hat{h}_1)$ 'den örnek çekmek oldukça kolaydır.

Düzeltilmiş bootstrap, parametrik ve parametrik olmayan bootstrap arasında ortak bir çözüm olarak da düşünülebilir (Holmes 1998: 208).

3.2.4. Parametrik Bootstrap Yöntemi

F dağılımına sahip bir anakitleden alınan X tesadüfi değişkeninin birbirinden bağımsız ve aynı şekilde dağılmış değerlerinin oluşturduğu $x_1, x_2, ..., x_n$ örneğine sahip olduğumuzu varsayalım.

X 'in değerleri gerçek sayılar veya sayıların oluşturduğu vektörler olabilir. Anakitleyi tanımlamanın en genel yolu, onun kümülatif dağılım fonksiyonunu yazmaktır.

$$F(x) = \Pr ob(X \leq x) \quad \textbf{(3.15)}$$

Eşitlik 3.15'teki fonksiyonun X'e göre türevi olasılık yoğunluk fonksiyonunu verir ve $f(x)$ eşitlik 3.16'daki gibi gösterilir.

$$f(x) = \frac{dF(x)}{dx} \quad \textbf{(3.16)}$$

Gözlemlerimiz için tanımladığımız olasılık yoğunluk fonksiyonu eşitlik 3.17' deki gibi gösterilir;

$$X : f_{(\theta)}(x) \quad \textbf{(3.17)}$$

Eşitlik 3.16'daki bu ifadede θ, daha önce de ifade edildiği gibi X dağılımından elde edilen bir veya birden fazla bilinmeyen parametreyi gösterir. Bu ifade ayrıca, X için parametrik model olarak adlandırılır. θ 'nın elemanlarının sayısını ρ ile gösterirsek; örneğin X , μ ortalama ve σ^2 varyanslı normal dağılıma sahip ise bu durumda, eşitlik 3.18'deki gibi olacak ve $\rho = 2$ elde edilecektir.

$$\theta = \left(\mu, \sigma^2\right) \quad \textbf{(3.18)}$$

Dağılımın olasılık yoğunluk fonksiyonu, eşitlik 3.19'da gösterildiği gibidir.

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-1/2\left(\frac{x-\mu}{\varsigma}\right)^2} \quad \textbf{(3.19)}$$

Maksimum olabilirlik, olabilirlik fonksiyonuna dayandırılır ve eşitlik 3.20.' deki gibi gösterilir:

$$L(\theta; x) = \prod_{1}^{n} f_\theta(x_i) \quad \textbf{(3.20)}$$

$L(\theta; x)$, θ'nın bir fonksiyonu olarak düşünülebilir. Veri kesikli olduğunda $L(\theta; x)$, gözlemlenen örneğin olasılığıdır. Süreklilik durumunda $L(\theta; x)\Delta, [x.x + \Delta]$ gibi küçük bir aralıkta yer alan örneğin olasılığı şeklinde açıklanabilir. $1(\theta; x)$'nın algoritmasını eşitlik 3.21'deki gibi yazmak mümkündür.

$$1(\theta; x) = \sum_{1}^{n} 1(\theta; x_i) \qquad \textbf{(3.21)}$$

Bu $1(\theta)$ şeklinde özetlenebilir. Bu ifade logaritmik olabilirlik olarak anılır ve her bir değer $1(\theta; x_i) = \log f_\theta(x_i)$, bir logaritmik olabilirlik bileşeni olarak adlandırılır. Ayrıca maksimum olabilirlik yönteminde $1(\theta; x)$'i maksimize etmek için $\theta = \hat{\theta}$ olduğu düşünülür. $\hat{\theta}$'nın örnekleme dağılımını ve varyansını tahmin etmenin en doğru şekli olarak tanımlanan parametrik yönteme göre, öncelikle $\hat{f}_\theta(x)$ olasılık yoğunluk fonksiyonuna göre n büyüklüğünde B tane örnek çekilir ve her bir örnek varyansı, $\hat{\theta}$'nın varyansını tahmin etmeye yarar. Bu süreç ise parametrik bootstrap olarak adlandırılır. Standart hatanın parametrik bootstrap tahmini eşitlik 3.22'deki gibidir:

$$se_{\hat{F}_{par}} \left(\hat{\theta}^*\right) \qquad \textbf{(3.22)}$$

Eşitlik 3.22'de $\hat{F}_{par}$ veriler için elde edilen parametrik modelden türetilen F'nin tahminidir (Efron, Tibshirani 1993: 208).

F dağılımının sürekli olduğu varsayıldığında, düzeltilmiş bootstrap dönüşümü yapılmaktadır. Bu aşama bir adım daha ilerletildiğinde, örneğin F dağılımının Gausyan dağılım gibi parametrik bir formda olduğu varsayıldığında, bu durumda F'nin yaklaşık tahmin edicisi de μ ve σ 'nın maksimum olabilirlik tahminlerine sahip Gausyan dağılım gösterecektir. F'nin parametrik tahmininden yapılan iadeli örnekleme sayesinde, Fisher'in teorisi ile ilişkili maksimum olabilirlik tahminlerinin yapıldığı bootstrap tahminlerine ulaşılacaktır. Parametrik bootstrap yöntemine Monte Carlo simülasyonu ile bakıldığında ise, bunun maksimum olabilirlik tahminine ulaşmaktan başka bir şey olmadığı gözlenir (Chernick 1999).

3.2.5. Parametrik Olmayan Bootstrap Yöntemi

Parametrik ve parametrik olmayan bootstrap arasındaki en önemli fark, parametrik bootstrap için bir parametrik modelin varolması iken parametrik olmayan bootstrap için böyle olmamasıdır. Bilinmeyen bir F dağılımından alınan bağımsız ve aynı şekilde dağılmış $x_1, x_2, ..., x_n$ değerlerine sahip olduğumuzu ve herhangi bir parametrik modelin var olmadığını düşünelim. Bilinmeyen F dağılımının kümülatif dağılım fonksiyonunu elde etmek için F deneysel dağılımını kullanır. Ancak daha önceki açıklamalara dayanarak, F 'i sadece parametrik bir model var olduğu takdirde kullanmanın mümkün olduğu söylenebilir. Aksi takdirde, verilerin simülasyonu ve gerekli özelliklerin deneysel hesaplamaları yapılmalıdır. Parametrik olmayan bootstrap yöntemi ile ilgili bir örnek açıklamaları anlaşılır kılacaktır. Ortalama hesaplarken deneysel dağılım fonksiyonundan yapılan örnekleme yardımıyla eşitlik 3.23'teki gibi momentler kolayca bulunabilir.

$$E^*\left(\overline{X}^*\right) = E^*\left(X^*\right) = \sum_{j=1}^{n} \frac{1}{n} x_j = \overline{x} \tag{3.23}$$

Benzer şekilde eşitlik 3.24 şeklinde de yazılır.

$$\operatorname{var}^*\left(\overline{X}^*\right) = \frac{1}{n}\operatorname{var}^*\left(X^*\right) = \frac{1}{n} E^*\left\{X^* - E^*\left(X^*\right)\right\}^2$$

$$= \frac{1}{n}\sum_{j=1} \frac{1}{n}\left(x_i - \overline{x}\right)^2 = \frac{(n-1)}{n}\frac{1}{n(n-1)}\sum_{j=1}^{n}\left(x_i - \overline{x}\right)^2 \tag{3.24}$$

Eşitlik 3.23'te verilen ifadede, ilk çarpan hariç, $\overline{X}$ 'nın tahmin edilmiş varyans sonucudur. Bu noktada, deneysel dağılım fonksiyonuna sahip simülasyon uygulaması yapılır. Deneysel dağılım fonksiyonu orijinal veri grubu olan $x_1, x_2, ..., x_n$ kümesindeki değerlerin her birine eşit olasılık verdiğinden; her bir X^*, orijinal örnekten tesadüfi olarak örneklenmiş bağımsız değerler olacaktır. Bu nedenle, simülasyon örneği olan $x_1^*, x_2^*, ..., x_n^*$, orijinal verilerden iadeli olarak alınan tesadüfi bir örnek olacaktır. Burada kolaylık sağlayan, verilerin homojen olmasıdır. Bu yeniden örnekleme yöntemi, parametrik olmayan bootstrap olarak bilinir (Efron, Tibshirani 1993: 208).

BÖLÜM 4

4. UYGULAMA

Uygulamada 20 ve 15 gözleme sahip 2 veri seti kullanmıştır. Her veri setine EKK ile Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman parametrik olmayan regresyon yöntemleri uygulanarak model tahminleri elde edilmiştir. Her veri seti için %10, %20, %30, %40 ve %50 aykırı değer olması durumunda ve %25, %50, %75 ve %100 oranlarında bootstrap yöntemi uygulandığında elde edilen sonuçların karşılaştırılması amaçlanmıştır. Uygulanan her özel durum ayrı ayrı tablolar ile belirtilmiş ve yorumlanmıştır. Parametrik Olmayan Regresyon analizleri Microsoft Excel programı ile tek tek yapılmıştır. EKKY ile regresyon analizi tahminleri SPSS.21 ve Bootstrap yöntemi için veri üretme GRETL paket programı ile yapılmıştır.

Uygulamanın birinci kısmında 20 gözlem ve 2 değişkene ait veriler kullanılmıştır. Veriler Spaeth (1991: 304) çalışmasından alınmıştır ve şunlardır:

20 Gözlemli Veri Seti

X	17	19	20	21	25	29	34	36	39	39
Y	114	124	116	120	125	130	110	136	144	120
X	42	42	44	45	45	46	47	47	48	50
Y	124	128	160	138	135	142	220	145	130	142

Bu değişkenlere ait modele EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman yöntemleri uygulanarak parametre tahminleri ayrı ayrı elde edilmiştir. Tahmin sonuçlarının ayrı ayrı HKOK değerleri hesaplanmış ve yorumlanmıştır.

$$\text{HKOK}=\sqrt{\frac{\sum(Y_i-\widehat{Y})^2}{n}}=\sqrt{\frac{\sum u^2}{n}}$$

Daha sonra veriden %10, %20, %30, %40 ve %50 oranında aykırı değerler oluşturulmuş ve yukarıda belirtilen yöntemler bu yeni modellere uygulanmıştır. Daha sonra 20 gözlemli veri setimiz bootstrap metodu ile %25, %50, %75 ve %100 oranında artırılarak 25, 30, 35 ve 40 gözlem elde edilerek aynı işlemler tekrar edilmiştir.

Tablo 4.1'de 20 gözlemli veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri HKOK değerleri verilmiştir.

Tablo 4.1. 20 Gözleme sahip modele ait sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=94,07034+1,117814(X)	19,48
THEİL-SEN	Y=101,9036364+0,894545455(X)	19,63
MOOD-BROWN	Y=111,4459459+0,594594595(X)	20,36
OPTİMUM	Y=99,82181817+0,894545455(X)	19,78
HODGES-LEHMAN	Y=102,274545+0,894545455(X)	19,63

Tablo 4.1'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.2'de 20 gözlemli ve %10 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.2. 20 Gözlemli ve %10 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=40,01342+2,691538(X)	25,74
THEİL-SEN	Y=65,1375+1,775(X)	138,49
MOOD-BROWN	Y=109,2567568+0,675675676(X)	295,61
OPTİMUM	Y=76,65+1,775(X)	134,81
HODGES-LEHMAN	Y=114,57375+1,775(X)	129,36

Tablo 4.2'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.3'de 20 gözlemli ve %20 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.3. 20 Gözlemli ve %20 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=23,51594+3,224889(X)	171,38
THEİL-SEN	Y=72+2(X)	303,04
MOOD-BROWN	Y=109,2567568+0,675675676(X)	553,92
OPTİMUM	Y=72+2(X)	303,04
HODGES-LEHMAN	Y=177,95+2(X)	283,91

Tablo 4.3'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.4'de 20 gözlemli ve %30 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.4. 20 Gözlemli ve %30 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=28,97565+3,151487(X)	180,20
THEİL-SEN	Y=29,58477842+2,669006331(X)	223,15
MOOD-BROWN	Y=48,12898752+2,939667129 (X)	186,57
OPTİMUM	Y=55,43682906+2,669006331(X)	214,99
HODGES-LEHMAN	Y=111,6391453+2,669006331(X)	207,52

Tablo 4.4'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Mood-Brown yönteminden elde edilmiştir.

Tablo 4.5'de 20 gözlemli ve %40 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.5. 20 Gözlemli ve %40 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=33,13782+3,170455(X)	191,72
THEİL-SEN	Y=58,92558799+2,736640534(X)	226,91
MOOD-BROWN	Y=48,01530612+2,943877551(X)	201,69
OPTİMUM	Y=60,09104247+2,736640534(X)	226,56
HODGES-LEHMAN	Y=127,4880503+2,736640534(X)	216,30

Tablo 4.5'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Mood-Brown yönteminden elde edilmiştir.

Tablo 4.6'da 20 gözlemli ve %50 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.6. 20 Gözlemli ve %50 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=38,53322+3,127748(X)	200,87
THEİL-SEN	Y=63,78146864+2,883001608(X)	212,79
MOOD-BROWN	Y=47,91044776+2,947761194(X)	208,87
OPTİMUM	Y=57,60059318+2,883001608(X)	214,01
HODGES-LEHMAN	Y=102,9381269+2,883001608(X)	209,15

Tablo 4.6'daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Mood-Brown yönteminden elde edilmiştir.

Tablo 4.7'de 20 gözlemli orijinal veri setine %25 bootstrap uygulanarak artırılmış veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.7. 20 Gözlemli %25 Bootstrap Uygulanmış Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=86,48706+1,331349(X)	19,40
THEİL-SEN	Y=96,64+1,44(X)	24,06
MOOD-BROWN	Y=78+1,333333333 (X)	21,14
OPTİMUM	Y=87,84+1,44(X)	20,15
HODGES-LEHMAN	Y=82,4496+1,44(X)	19,43

Tablo 4.7'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.8'de 20 gözlemli orijinal veri setine %25 bootstrap uygulanarak artırılmış ve %10 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.8. 20 Gözlemli %25 Bootstrap Uygulanmış ve %10 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=21,1982+3,16109(X)	52,26
THEİL-SEN	Y=23,45454545+2,545454545(X)	109,60
MOOD-BROWN	Y=73,29310345+1,482758621 (X)	247,41
OPTİMUM	Y=48,00000002+2,545454545(X)	100,72
HODGES-LEHMAN	Y=73,75927277+2,545454545(X)	97,37

Tablo 4.8’deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.9’da 20 gözlemli orijinal veri setine %25 bootstrap uygulanarak artırılmış ve %20 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.9. 20 Gözlemli %25 Bootstrap Uygulanmış ve %20 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=20,53014+3,219164(X)	66,16
THEİL-SEN	Y=39,42259632+2,927449196(X)	84,09
MOOD-BROWN	Y=11,71875+3,4375 (X)	77,74
OPTİMUM	Y=31,19223538+2,927449196(X)	86,05
HODGES-LEHMAN	Y=55,49622957+2,927449196(X)	82,54

Tablo 4.9’daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Mood-Brown yönteminden elde edilmiştir.

Tablo 4.10’da 20 gözlemli orijinal veri setine %25 bootstrap uygulanarak artırılmış ve %30 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.10. 20 Gözlemli %25 Bootstrap Uygulanmış ve %30 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=58,19446+3,26526(X)	174,61
THEİL-SEN	Y=6,140292276+3,054697286(X)	197,29
MOOD-BROWN	Y=-54,06128887+4,631816133(X)	315,82
OPTİMUM	Y=24,86680585+3,054697286(X)	190,11
HODGES-LEHMAN	Y=89,78565679+3,054697286(X)	178,69

Tablo 4.10’daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.11’de 20 gözlemli orijinal veri setine %25 bootstrap uygulanarak artırılmış ve %40 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.11. 20 Gözlemli %25 Bootstrap Uygulanmış ve %40 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=71,90987+3,16867(X)	183,93
THEİL-SEN	Y=6,192+3,053176471(X)	204,99
MOOD-BROWN	Y=2,95864486+3,069626168 (X)	206,39
OPTİMUM	Y=24,92611763+3,053176471(X)	197,72
HODGES-LEHMAN	Y=93,63644228+3,053176471(X)	185,40

Tablo 4.11’deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.12’de 20 gözlemli orijinal veri setine %25 bootstrap uygulanarak artırılmış ve %50 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.12. 20 Gözlemli %25 Bootstrap Uygulanmış ve %50 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=114,303+3,125266(X)	224,56
THEİL-SEN	Y=0,43117797+3,065867232(X)	258,78
MOOD-BROWN	Y=13,0719488+3,04943721(X)	254,77
OPTİMUM	Y=24,43117795+3,065867232(X)	247,79
HODGES-LEHMAN	Y=22,03237366+3,065867232(X)	248,81

Tablo 4.12’deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Optimum yönteminden elde edilmiştir.

Tablo 4.13’de 20 gözlemli orijinal veri setine %50 bootstrap uygulanarak artırılmış veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.13. 20 Gözlemli %50 Bootstrap Uygulanmış Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=105,858+0,6382219(X)	8,37
THEİL-SEN	Y=98,66666665+0,866666667(X)	8,73
MOOD-BROWN	Y=78,94736842+1,842105263(X)	21,72
OPTİMUM	Y=98,96666666+0,866666667(X)	8,77
HODGES-LEHMAN	Y=97,71777777+0,866666667(X)	8,68

Tablo 4.13'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.14'de 20 gözlemli orijinal veri setine %50 bootstrap uygulanarak artırılmış ve %10 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.14. 20 Gözlemli %50 Bootstrap Uygulanmış ve %10 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=9,527941+3,531733(X)	84,25
THEİL-SEN	Y=93,75+1,25(X)	277,00
MOOD-BROWN	Y=105,8333333+0,833333333(X)	324,18
OPTİMUM	Y=88,625+1,25(X)	278,47
HODGES-LEHMAN	Y=170,9833333+1,25(X)	266,01

Tablo 4.14'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.15'de 20 gözlemli orijinal veri setine %50 bootstrap uygulanarak artırılmış ve %20 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.15. 20 Gözlemli %50 Bootstrap Uygulanmış ve %20 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=27,3645 +3,285872(X)	123,67
THEİL-SEN	Y=54,5+1,75(X)	306,09
MOOD-BROWN	Y=113,4285714+0,571428571(X)	492,50
OPTİMUM	Y=79,25+1,75(X)	295,45
HODGES-LEHMAN	Y=196,6483333+1,75(X)	271,13

Tablo 4.15'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.16'da 20 gözlemli orijinal veri setine %50 bootstrap uygulanarak artırılmış ve %30 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.16. 20 Gözlemli %50 Bootstrap Uygulanmış ve %30 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=44,51173+3,194585(X)	159,54
THEİL-SEN	Y=61,66666666+2,716666667(X)	190,86
MOOD-BROWN	Y=(-16,66666667)+5,057471264(X)	441,24
OPTİMUM	Y=54,14999999+2,716666667(X)	193,16
HODGES-LEHMAN	Y=116,5117777+2,716666667(X)	182,81

Tablo 4.16'daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.17'de 20 gözlemli orijinal veri setine %50 bootstrap uygulanarak artırılmış ve %40 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.17. 20 Gözlemli %50 Bootstrap Uygulanmış ve %40 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=71,9435+3,133543(X)	198,03
THEİL-SEN	Y=9,470491803+2,822131148(X)	240,58
MOOD-BROWN	Y=44,04432133+2,96398892(X)	209,77
OPTİMUM	Y=51,302459+2,822131148(X)	222,64
HODGES-LEHMAN	Y=130,7152622+2,822131148(X)	207,99

Tablo 4.17'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.18'de 20 gözlemli orijinal veri setine %50 bootstrap uygulanarak artırılmış ve %50 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.18. 20 Gözlemli %50 Bootstrap Uygulanmış ve %50 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=106,8571+3,071667(X)	231,39
THEİL-SEN	Y=60,66055046+2,825688073(X)	238,68
MOOD-BROWN	Y=10,35431655+2,96398892(X)	266,60
OPTİMUM	Y=98,96666666+2,825688073(X)	245,99
HODGES-LEHMAN	Y=163,3386314+2,825688073(X)	237,41

Tablo 4.18'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.19'da 20 gözlemli orijinal veri setine %75 bootstrap uygulanarak artırılmış veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.19. 20 Gözlemli %75 Bootstrap Uygulanmış Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=92,39882+1,185624(X)	20,52
THEİL-SEN	Y=105+1(X)	21,43
MOOD-BROWN	Y=104,7619048+0,80952381 (X)	20,99
OPTİMUM	Y=97+1(X)	20,74
HODGES-LEHMAN	Y=99,17142857+1(X)	20,62

Tablo 4.19'daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.20'de 20 gözlemli orijinal veri setine %75 bootstrap uygulanarak artırılmış ve %10 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.20. 20 Gözlemli %75 Bootstrap Uygulanmış ve %10 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=5,311082+3,598755(X)	112,39
THEİL-SEN	Y=86,52631579+1,473684211(X)	323,31
MOOD-BROWN	Y=102,2727273+ 0,909090909(X)	401,64
OPTİMUM	Y=86,52631577+1,473684211(X)	323,31
HODGES-LEHMAN	Y=183,2463157+1,473684211(X)	308,51

Tablo 4.20'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.21'de 20 gözlemli orijinal veri setine %75 bootstrap uygulanarak artırılmış ve %20 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.21. 20 Gözlemli %75 Bootstrap Uygulanmış ve %20 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=13,87563+3,396215(X)	128,73
THEİL-SEN	Y=40+2,666666667(X)	190,58
MOOD-BROWN	Y=119,5652174+0,217391304 (X)	629,22
OPTİMUM	Y=58,33333333+2,666666667(X)	185,46
HODGES-LEHMAN	Y=101,5923809+2,666666667(X)	180,35

Tablo 4.21'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.22'de 20 gözlemli orijinal veri setine %75 bootstrap uygulanarak artırılmış ve %30 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.22. 20 Gözlemli %75 Bootstrap Uygulanmış ve %30 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=43,76471+3,273211(X)	196,90
THEİL-SEN	Y=30,11320755+2,94129979(X)	219,22
MOOD-BROWN	Y=0+3,235294118 (X)	203,30
OPTİMUM	Y=30,58280924+2,94129979(X)	219,08
HODGES-LEHMAN	Y=98,6078587+2,94129979(X)	208,25

Tablo 4.22'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Mood-Brown yönteminden elde edilmiştir.

Tablo 4.23'de 20 gözlemli orijinal veri setine %75 bootstrap uygulanarak artırılmış ve %40 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.23. 20 Gözlemli %75 Bootstrap Uygulanmış ve %40 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=63,94081+3,221987 (X)	225,13
THEİL-SEN	Y=50,57471264+2,977011494(X)	239,94
MOOD-BROWN	Y=26,98113208+3,028301887(X)	241,47
OPTİMUM	Y=50,57471265+2,977011494(X)	239,94
HODGES-LEHMAN	Y=113,4076848+2,977011494(X)	231,57

Tablo 4.23'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.24'de 20 gözlemli orijinal veri setine %75 bootstrap uygulanarak artırılmış ve %50 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.24. 20 Gözlemli %75 Bootstrap Uygulanmış ve %50 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=107,4009+3,15282(X)	267,59
THEİL-SEN	Y=8,501845018+2,985239852(X)	304,73
MOOD-BROWN	Y=25,61992136+3,035386632(X)	290,95
OPTİMUM	Y=50,53284153+2,985239852(X)	287,78
HODGES-LEHMAN	Y=148,9761519+2,985239852(X)	270,42

Tablo 4.24’deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.25’de 20 gözlemli orijinal veri setine %100 bootstrap uygulanarak artırılmış veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.25. 20 Gözlemli %100 Bootstrap Uygulanmış Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=98,68892+0,9078654(X)	15,59
THEİL-SEN	Y=98,4+0,88(X)	15,65
MOOD-BROWN	Y=120,9375+0,594594595(X)	16,77
OPTİMUM	Y=98,2+0,88(X)	15,67
HODGES-LEHMAN	Y=99,729+0,88(X)	15,59

Tablo 4.25’deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.26’da 20 gözlemli orijinal veri setine %100 bootstrap uygulanarak artırılmış ve %10 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.26. 20 Gözlemli %100 Bootstrap Uygulanmış ve %10 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=20,61687 +3,088723 (X)	49,09
THEİL-SEN	Y=89,4 +1,4 (X)	216,24
MOOD-BROWN	Y=59,45945946 + 2,432432432 (X)	93,22
OPTİMUM	Y=88+1,4 (X)	216,64
HODGES-LEHMAN	Y=149,534+1,4 (X)	207,71

Tablo 4.26'daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Mood-Brown yönteminden elde edilmiştir.

Tablo 4.27'de 20 gözlemli orijinal veri setine %100 bootstrap uygulanarak artırılmış ve %20 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.27. 20 Gözlemli %100 Bootstrap Uygulanmış ve %20 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=20,61687+3,088723(X)	49,09
THEİL-SEN	Y=72+2(X)	229,74
MOOD-BROWN	Y=88,875+1,0625(X)	154755,09
OPTİMUM	Y=72+2(X)	229,74
HODGES-LEHMAN	Y=160,08+2(X)	212,19

Tablo 4.27'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.28'de 20 gözlemli orijinal veri setine %100 bootstrap uygulanarak artırılmış ve %30 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.28. 20 Gözlemli %100 Bootstrap Uygulanmış ve %30 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=53,72278 +3,02627(X)	156,71
THEİL-SEN	Y=61,43396226+2,728301887(X)	171,55
MOOD-BROWN	Y=(-44,20153341)+4,210295728(X)	291,86
OPTİMUM	Y=50,87924528+2,728301887(X)	174,26
HODGES-LEHMAN	Y=100,4232453+2,728301887(X)	167,06

Tablo 4.28'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.29'da 20 gözlemli orijinal veri setine %100 bootstrap uygulanarak artırılmış ve %40 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.29. 20 Gözlemli %100 Bootstrap Uygulanmış ve %40 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=66,09629+3,028795(X)	175,69
THEİL-SEN	Y=60,0045124+2,829048175(X)	187,14
MOOD-BROWN	Y=(-15,51515152)+3,474747475(X)	201,46
OPTİMUM	Y=35,5218803+2,829048175(X)	194,76
HODGES-LEHMAN	Y=84,63327953+2,829048175(X)	182,50

Tablo 4.29'daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.30'da 20 gözlemli orijinal veri setine %100 bootstrap uygulanarak artırılmış ve %50 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.30. 20 Gözlemli %100 Bootstrap Uygulanmış ve %50 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=86,78581+2,993377(X)	197,09
THEİL-SEN	Y=58,98784846+2,850607577(X)	210,00
MOOD-BROWN	Y=(-13,44969615)+3,421787081(X)	221,50
OPTİMUM	Y=56,36132947+2,850607577(X)	210,82
HODGES-LEHMAN	Y=123,2356326+2,850607577(X)	199,93

Tablo 4.30'daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Uygulamanın bu bölümünde 15 gözleme sahip gerçek hayat veri seti kullanılmıştır. Bu veri(kaynak) çalışmasından alınmış ve aşağıda verilmiştir.

15 Gözleme Ait Veri Seti

X	3	5	7	8	9	3,2	4,7	6,8	8,5	24	28	31	29,8	3,2	26,4
Y	12	16	15,5	12,1	8,8	3,6	4,1	4,2	2,9	16,8	17,8	18,8	14,5	2,9	19

Birinci uygulamada yapılan tüm tahmin çalışmaları, ikinci uygulamada da yapılmıştır. Yani değişkenlere ait modele EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman yöntemleri uygulanarak parametre tahminleri ayrı ayrı elde edilmiştir. Tahmin sonuçlarının ayrı ayrı HKOK değerleri hesaplanmış ve yorumlanmıştır.

Daha sonra veriden %10, %20, %30, %40 ve %50 oranında aykırı değerler oluşturulmuş ve yukarıda belirtilen yöntemler bu yeni modellere uygulanmıştır. Daha sonra 15 gözlemli veri setimiz bootstrap metodu ile %25, %50, %75 ve %100 oranında artırılmış 18, 23, 27 ve 31 gözlem elde edilerek aynı işlemler tekrar edilmiştir.

Tablo 4.31'de 15 gözlemli veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.31. 15 Gözleme Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y= 5,928953+0,4051908(X)	4,24
THEİL-SEN	Y=1,744444444+0,361111111(X)	6,40
MOOD-BROWN	Y=0,87281106+0,686635945(X)	5,36
OPTİMUM	Y=7,605555556+0,361111111(X)	4,41
HODGES-LEHMAN	Y=6,509632963+0,361111111(X)	4,27

Tablo 4.31'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.32'de 15 gözlemli ve %10 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.32. 15 Gözlemli ve %10 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=2,643215+0,7037414 (X)	5,39
THEİL-SEN	Y=(-1,729107981)+0,544600939(X)	14,90
MOOD-BROWN	Y=16,30241935+(-0,060483871)(X)	14,49
OPTİMUM	Y=6,529489299+0,544600939(X)	11,65
HODGES-LEHMAN	Y=7,621978091+0,544600939(X)	11,60

Tablo 4.32’deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.33’de 15 gözlemli ve %20 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.33. 15 Gözlemli ve %20 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=5,157797+0,5749436(X)	12,89
THEİL-SEN	Y=(-1,272193437)+0,490846287(X)	18,45
MOOD-BROWN	Y=15,46153846+0,107692308(X)	47,32
OPTİMUM	Y=4,38238342+0,490846287(X)	15,89
HODGES-LEHMAN	Y=9,325658031+0,490846287(X)	15,10

Tablo 4.33’deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.34’de 15 gözlemli ve %30 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.34. 15 Gözlemli ve %30 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y= 5,771995 +0,5871999(X)	13,48
THEİL-SEN	Y=10,44148244+0,598423277(X	14,69
MOOD-BROWN	Y=13,05454545+0,589090909(X)	15,41
OPTİMUM	Y=3,414190507+0,598423277(X)	13,63
HODGES-LEHMAN	Y=4,796040962+0,598423277(X)	13,56

Tablo 4.34’deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.35’de 15 gözlemli ve %40 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.35. 15 Gözlemli ve %40 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=6,397301+0,5973553(X)	14,60
THEİL-SEN	Y=1,268164935+0,602518099(X)	15,32
MOOD-BROWN	Y=13,04989309+0,590021383(X)	15,78
OPTİMUM	Y=7,279855209+0,602518099(X)	14,69
HODGES-LEHMAN	Y=5,865863645+0,602518099(X)	14,62

Tablo 4.35'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.36'da 15 gözlemli ve %50 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.36. 15 Gözlemli ve %50 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y= 5,424834+0,5960282 (X)	16,62
THEİL-SEN	Y=3,434254144+0,596193984(X)	16,73
MOOD-BROWN	Y=13,04112406+0,591775188(X)	18,09
OPTİMUM	Y=7,330448128+0,596193984(X	16,73
HODGES-LEHMAN	Y=5,405599345+0,596193984(X)	16,62

Tablo 4.36'daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.37'de 15 gözlemli orijinal veri setine %25 bootstrap uygulanarak artırılmış veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.37. 15 Gözlemli %25 Bootstrap Uygulanmış Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=6,502176+0,3226085(X)	4,42
THEİL-SEN	Y=2,152589641+0,414342629(X)	5,51
MOOD-BROWN	Y=2,152589641+0,414342629(X)	5,51
OPTİMUM	Y=2,152589641+0,414342629(X)	5,51
HODGES-LEHMAN	Y=5,283078213+0,414342629(X)	4,54

Tablo 4.37'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.38'de 15 gözlemli orijinal veri setine %25 bootstrap uygulanarak artırılmış ve %10 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.38. 15 Gözlemli %25 Bootstrap Uygulanmış ve %10 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y= 4,150369 +0,5321637(X)	7,69
THEİL-SEN	Y=1,154782609+0,447826087(X)	12,71
MOOD-BROWN	Y=2,152589641+0,414342629(X)	14,96
OPTİMUM	Y=2,166956522+0,447826087(X)	12,22
HODGES-LEHMAN	Y=7,726549199+0,447826087(X)	10,88

Tablo 4.38'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.39'da 15 gözlemli orijinal veri setine %25 bootstrap uygulanarak artırılmış ve %20 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.39. 15 Gözlemli %25 Bootstrap Uygulanmış ve %20 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y= 13,37043+0,5505091 (X)	33,66
THEİL-SEN	Y=(-1,270082634)+0,490597957(X)	38,42
MOOD-BROWN	Y=2,152589641+0,414342629(X)	40,29
OPTİMUM	Y=1,794189602+0,490597957(X)	37,11
HODGES-LEHMAN	Y=16,36150474+0,490597957(X)	34,13

Tablo 4.39'daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.40’da 15 gözlemli orijinal veri setine %25 bootstrap uygulanarak artırılmış ve %30 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.40. 15 Gözlemli %25 Bootstrap Uygulanmış ve %30 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y= 14,68909 +0,5428524 (X)	34,86
THEİL-SEN	Y=10,27269439+0,575768535(X)	35,09
MOOD-BROWN	Y=1,285939968+0,598736177(X)	36,71
OPTİMUM	Y=2,847739602+0,575768535(X)	36,38
HODGES-LEHMAN	Y=12,78019378+0,575768535(X)	35,00

Tablo 4.40’daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.41’de 15 gözlemli orijinal veri setine %25 bootstrap uygulanarak artırılmış ve %40 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.41. 15 Gözlemli %25 Bootstrap Uygulanmış ve %40 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y= 23,16328+0,4927738 (X)	41,60
THEİL-SEN	Y=1,700473934+0,593601896(X)	45,30
MOOD-BROWN	Y=12,37861094+0,445912723(X)	44,31
OPTİMUM	Y=1,31007109+0,593601896(X)	45,42
HODGES-LEHMAN	Y=2,900816912+0,593601896(X)	44,94

Tablo 4.41’deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Mood-Brown yönteminden elde edilmiştir.

Tablo 4.42’de 15 gözlemli orijinal veri setine %25 bootstrap uygulanarak artırılmış ve %50 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.42. 15 Gözlemli %25 Bootstrap Uygulanmış ve %50 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=30,66596+0,462119(X)	44,75
THEİL-SEN	Y=11,47827567+0,574532048 (X)	47,65
MOOD-BROWN	Y=5,320902613+1,45415677(X)	149,38
OPTİMUM	Y=1,399699376+0,574532048(X)	50,41
HODGES-LEHMAN	Y=3,154245156+0,574532048 (X)	49,79

Tablo 4.42’deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Theil yönteminden elde edilmiştir.

Tablo 4.43’de 15 gözlemli orijinal veri setine %50 bootstrap uygulanarak artırılmış veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.43. 15 Gözlemli %50 Bootstrap Uygulanmış Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=6,04727176+0,4286162(X)	3,98
THEİL-SEN	Y=9,1+0,375(X)	4,67
MOOD-BROWN	Y=0,761538462+0,668269231(X)	5,20
OPTİMUM	Y=7,3+0,375(X)	4,06
HODGES-LEHMAN	Y=6,711413043+0,375(X)	4,01

Tablo 4.43’deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.44’de 15 gözlemli orijinal veri setine %50 bootstrap uygulanarak artırılmış ve %10 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.44. 15 Gözlemli %50 Bootstrap Uygulanmış ve %10 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y= 3,78837+0,6447297 (X)	5,05
THEİL-SEN	Y=10,2+0,6(X)	7,76
MOOD-BROWN	Y=0,679310345+0,693965517(X)	6,34
OPTİMUM	Y=3,16+0,6(X)	6,34
HODGES-LEHMAN	Y=5,263478261+0,6(X)	5,98

Tablo 4.44’deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.45’de 15 gözlemli orijinal veri setine %50 bootstrap uygulanarak artırılmış ve %20 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.45. 15 Gözlemli %50 Bootstrap Uygulanmış ve %20 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=11,33154+0,6626364+(X)	29,45
THEİL-SEN	Y=3,173006135+0,625221541(X)	31,13
MOOD-BROWN	Y=0,097520661+0,603305785(X)	32,77
OPTİMUM	Y=2,494151329+0,625221541(X)	31,35
HODGES-LEHMAN	Y=12,86753552+0,625221541(X)	29,58

Tablo 4.45’deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.46’da 15 gözlemli orijinal veri setine %50 bootstrap uygulanarak artırılmış ve %30 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.46. 15 Gözlemli %50 Bootstrap Uygulanmış ve %30 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=16,18325+0,6475505(X)	35,56
THEİL-SEN	Y=2,420947075+0,627994429(X)	38,50
MOOD-BROWN	Y=(-29,256)+4,92(X)	368,87
OPTİMUM	Y=2,420947075+0,627994429(X)	38,50
HODGES-LEHMAN	Y=17,08563098+0,627994429(X)	35,60

Tablo 4.46’daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.47'de 15 gözlemli orijinal veri setine %50 bootstrap uygulanarak artırılmış ve %40 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.47. 15 Gözlemli %50 Bootstrap Uygulanmış ve %40 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=22,87849+0,664645 (X)	42,54
THEİL-SEN	Y=0,761538462+0,668269231(X)	47,86
MOOD-BROWN	Y=0+0,617647059(X)	49,69
OPTİMUM	Y=1,357692308+0,668269231(X)	47,59
HODGES-LEHMAN	Y=22,68477843+0,668269231(X)	42,54

Tablo 4.47'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.48'de 15 gözlemli orijinal veri setine %50 bootstrap uygulanarak artırılmış ve %50 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.48. 15 Gözlemli %50 Bootstrap Uygulanmış ve %50 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=27,58683+0,6310668 (X)	44,56
THEİL-SEN	Y=10,02282878+0,659057072(X)	47,22
MOOD-BROWN	Y=(-2,628436019)+0,650404238(X)	53,04
OPTİMUM	Y=2,868486352+0,659057072(X)	50,01
HODGES-LEHMAN	Y=25,37723638+0,659057072(X)	44,65

Tablo 4.48'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.49'da 15 gözlemli orijinal veri setine %75 bootstrap uygulanarak artırılmış veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.49. 15 Gözlemli %75 Bootstrap Uygulanmış Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=5,81501+0,3941052(X)	3,99
THEİL-SEN	Y=1,504511278+0,436090226(X)	5,42
MOOD-BROWN	Y=(-0,162264151)+0,641509434(X)	5,23
OPTİMUM	Y=5,281203008+0,436090226(X)	4,01
HODGES-LEHMAN	Y=5,154135338+0,436090226(X)	4,01

Tablo 4.49'daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.50'de 15 gözlemli orijinal veri setine %75 bootstrap uygulanarak artırılmış ve %10 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.50. 15 Gözlemli %75 Bootstrap Uygulanmış ve %10 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y= 11,82733+0,513428(X)	31,83
THEİL-SEN	Y=(-1,729107981)+0,544600939(X)	34,30
MOOD-BROWN	Y=15,80701754+(-0,043859649)(X)	45,92
OPTİMUM	Y=2,551173709+0,544600939(X)	32,96
HODGES-LEHMAN	Y=10,93467919+0,544600939(X)	31,87

Tablo 4.50'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.51'de 15 gözlemli orijinal veri setine %75 bootstrap uygulanarak artırılmış ve %20 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.51. 15 Gözlemli %75 Bootstrap Uygulanmış ve %20 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y= 16,24591+0,5488829(X)	40,50
THEİL-SEN	Y=0,077836412+0,606200528(X)	43,03
MOOD-BROWN	Y=11,21926606+0,110091743(X)	57,86
OPTİMUM	Y=1,660158311+0,606200528(X)	42,54
HODGES-LEHMAN	Y=13,88255839+0,606200528(X)	40,75

Tablo 4.51'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.52'de 15 gözlemli orijinal veri setine %75 bootstrap uygulanarak artırılmış ve %30 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.52. 15 Gözlemli %75 Bootstrap Uygulanmış ve %30 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=18,59723 +0,5639522 (X)	42,69
THEİL-SEN	Y=(-0,553677345)+0,637760706(X)	45,72
MOOD-BROWN	Y=(-41,88979592)+5,269387755(X)	527,13
OPTİMUM	Y=1,493743054+0,637760706(X)	45,11
HODGES-LEHMAN	Y=14,04679736+0,637760706(X)	43,33

Tablo 4.52'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.53'de 15 gözlemli orijinal veri setine %75 bootstrap uygulanarak artırılmış ve %40 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.53. 15 Gözlemli %75 Bootstrap Uygulanmış ve %40 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y= 22,76224+0,5442758(X)	45,44
THEİL-SEN	Y=(-2,577384669)+0,644398196(X)	49,78
MOOD-BROWN	Y=(-3,161987041)+0,713174946(X)	51,02
OPTİMUM	Y=1,537925772+0,644398196(X)	48,60
HODGES-LEHMAN	Y=13,64829219+0,644398196(X)	47,07

Tablo 4.53'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.54'de 15 gözlemli orijinal veri setine %75 bootstrap uygulanarak artırılmış ve %50 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.54. 15 Gözlemli %75 Bootstrap Uygulanmış ve %50 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=26,71315+0,5317474(X)	47,93
THEİL-SEN	Y=0,055516938+0,606345584(X)	52,25
MOOD-BROWN	Y=(-3,15146771)+0,711937378(X)	54,44
OPTİMUM	Y=2,992476593+0,606345584(X)	51,29
HODGES-LEHMAN	Y=18,33454681+0,606345584(X)	48,95

Tablo 4.54'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.55'de 15 gözlemli orijinal veri setine %100 bootstrap uygulanarak artırılmış veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.55. 15 Gözlemli %100 Bootstrap Uygulanmış Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=3,38328+ 0,4793066(X)	3,37
THEİL-SEN	Y=1,069784173+0,571942446(X)	3,72
MOOD-BROWN	Y=0,679310345+0,693965517(X)	4,06
OPTİMUM	Y=1,411870504+0,571942446(X)	3,63
HODGES-LEHMAN	Y=2,324541657+0,571942446(X)	3,51

Tablo 4.55’deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.56’da 15 gözlemli orijinal veri setine %100 bootstrap uygulanarak artırılmış ve %10 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.56. 15 Gözlemli %100 Bootstrap Uygulanmış ve %10 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=(-0,9223173)+0,89568 (X)	5,94
THEİL-SEN	Y=0,87281106+0,686635945(X)	7,82
MOOD-BROWN	Y=1,504511278+0,436090226(X)	13,25
OPTİMUM	Y=0,87281106+0,686635945(X)	7,82
HODGES-LEHMAN	Y=2,789900401+0,686635945(X)	7,58

Tablo 4.56’daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.57’de 15 gözlemli orijinal veri setine %100 bootstrap uygulanarak artırılmış ve %20 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.57. 15 Gözlemli %100 Bootstrap Uygulanmış ve %20 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=(-2,990037)+1,065701(X)	10,11
THEİL-SEN	Y=0,34+0,8(X)	12,53
MOOD-BROWN	Y=1,850359712+0,54676259(X)	18,20
OPTİMUM	Y=0,34+0,8(X)	12,53
HODGES-LEHMAN	Y=2,981032258+0,8(X)	12,25

Tablo 4.57’deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.58'de 15 gözlemli orijinal veri setine %100 bootstrap uygulanarak artırılmış ve %30 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.58. 15 Gözlemli %100 Bootstrap Uygulanmış ve %30 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=(-3,198597)+ 1,026505 (X)	11,45
THEİL-SEN	Y=9,375+0,875(X)	14,83
MOOD-BROWN	Y=(-0,185010267)+0,911704312(X)	11,90
OPTİMUM	Y=0,1+0,875(X)	12,25
HODGES-LEHMAN	Y=0,967983871+0,875(X)	12,22

Tablo 4.58'deki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Mood-Brown yönteminden elde edilmiştir.

Tablo 4.59'da 15 gözlemli orijinal veri setine %100 bootstrap uygulanarak artırılmış ve %40 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.59. 15 Gözlemli %100 Bootstrap Uygulanmış ve %40 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y=(-1,766601)+0,9288729 (X)	15,22
THEİL-SEN	Y=0,845201669+0,883866481(X)	15,32
MOOD-BROWN	Y=9,435549525+0,866350068(X)	17,86
OPTİMUM	Y=0,07162726+0,883866481(X)	15,29
HODGES-LEHMAN	Y=(-0,286033469)+0,883866481(X)	15,28

Tablo 4.59'daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Tablo 4.60'da 15 gözlemli orijinal veri setine %100 bootstrap uygulanarak artırılmış ve %50 aykırı değere sahip veri setinin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminleri ve HKOK değerleri verilmiştir.

Tablo 4.60 15 Gözlemli %100 Bootstrap Uygulanmış ve %50 Aykırı Değere Sahip Modele Ait Sonuçlar

YÖNTEM	DENKLEM	HKOK
EKK	Y= 1,089627+0,920521 (X)	23,39
THEİL-SEN	Y=0+0,90625(X)	23,46
MOOD-BROWN	Y=(-8,76227758)+1,051601423(X)	24,20
OPTİMUM	Y=0+0,90625(X)	23,46
HODGES-LEHMAN	Y=1,650181452+0,90625(X)	23,40

Tablo 4.60’daki sonuçlara baktığımızda, her iki ölçüte göre en iyi tahmin modelinin EKK yönteminden elde edildiği görülürken ikinci en iyi tahmin Hodges-Lehman yönteminden elde edilmiştir.

Sonuçları genel olarak değerlendirecek olursak Bootstrap ile %25, %50, %75, %100 olarak artırdığımız verilerin EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerine ait parametre tahminlerinin HKOK değerleri orijinal veriye ait tahminler ile yakın sonuçlar verirken bu artırmış olduğumuz verilere %10, %20, %30, %40 ve %50 aykırı değer eklediğimizde HKOK değerlerinin arttığı görülürken en iyi tahmin modelinin ise EKK olduğu sonucunun değişmediği sonucuna ulaşılmıştır. Parametrik olmayan tahmin modelleri içerisinden ise genel olarak EKK’ ya en yakın sonuca Hodges-Lehman yönteminden ulaşılmıştır.

SONUÇ

Regresyon analizi, regresyon fonksiyonu hakkında istatistiksel çıkarımda bulunan bir analizdir ve temelde iki değişken arasındaki ilişkinin büyüklüğünü ölçmek için kullanılır. Bu iki değişken arasındaki fonksiyonel ilişki biliniyor ise parametrik regresyon fakat bu iki değişken arasındaki fonksiyonel ilişki bilinmiyor ise parametrik olmayan regresyon olarak adlandırılır.

Bootstrap yöntemi, yeni anakitle olarak kabul edilen n büyüklüğünde orijinal veri kümesinden iadeli seçim yapmak üzere *n* birimden oluşan örneklemler türetmeyi, oluşturulan yeni örneklemlerin tahminlerinin standart hatasını küçültmek ve bunun sonucu olarak daha güvenilir ve doğru tahminlere ulaşmayı amaçlamaktadır.

Uygulama kısmında iki farklı veri seti ile çalışılmıştır. İlk veri setimiz gerçek hayat verilerinden oluşurken ikinci veri setinde ise simülasyon veriler kullanılmıştır. İlk olarak veri setlerimize EKK, Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin yöntemleri ile parametre tahminleri yapılmıştır ve HKOK değerleri hesaplanmıştır. Daha sonra veri setlerimize sırasıyla (%10, %20, %30, %40, %50) aykırı değerleri eklenmiş ve yeniden HKOK tahmin sonuçlarını alıp karşılaştırması yapılmıştır. Verileri Bootstrap metodu ile sırasıyla (%25, %50, %75, %100) HKOK değerleri hesaplanmış ve bu verilerimize sırasıyla (%10, %20, %30, %40, %50) aykırı değerleri eklenip aynı işlemler tekrar uygulanmıştır.

Parametrik olmayan regresyon modelleri örneklem büyüklğnün az olduğu durumda daha iyi sonuçlar vermektedir. Parametrik regresyon modellerinde ise EKKY en iyi sonucu vermektedir. Bu çalışmada da EKKY'e alternatif olacak en yakın tahmin modeli bootstrap ile araştırılmıştır.

Uygulama sonucunda en iyi tahmin sonucunu EKK yönteminden alırken Theil-Sen, Mood-Brown, Optimum ve Hodges-Lehman tahmin modellerinden de yakın sonuçlar elde edilmiştir. Fakat parametrik olmayan regresyon modelleri içerisinde genel olarak EKK'ya en yakın sonuca Hodges-Lehman yönteminden ulaşılmıştır.

Çalışmanın uygulama kısmında 20 gözlemli gerçek hayat verisi ve 15 gözlemli simülasyon veri kullanılmıştır. Bootstrap sonuçlarına göre 15 gözlemli simülasyon verilerinin gerçek hayat sonuçlarına göre daha başarılı olduğu, EKK sonuçlarına en yakın değerlerin az gözleme sahip ve Bootstrap ile elde edilen sonuçlar olduğu da görülmüştür.

Literatürde parametrik olmayan regresyon yöntemlerinde, aykırı değere en iyi sonuç veren yöntem olarak Optimum yöntemi bilinmektedir. Bu çalışmada ise gerek 20 gözlemli gerekse 15 gözlemli verilerde Bootstrap ile elde edilmiş sonuçlar arasında Hodges-Lehman yönteminin daha başarılı olduğu görülmüştür. Ayrıca çalışmada aykırı değer ile doğru orantılı olarak standart hatanın artmış olduğu ve sonraki çalışmalar da aykırı değerin minimum düzeyde tutulması halinde daha iyi tahmin sonuçlarına ulaşabileceği sonucuna ulaşılmıştır.

İleriki çalışmalarda farklı Bootstrap yöntemleri ile (Wild Bootstrap, Çift Bootstrap vs.) parametrik olmayan ve yarı parametrik regresyon yöntemleri ile tahminler yapılarak literatüre katkılar sağlanabilir.

KAYNAKÇA

Adichie James Nwoye (1967). Asymtotic Efficiency of a Class of Non-Parametric Tests for Regression Parametres. Annals of Mathematical Statistics, 38, 884-893.

Alpar Reha (2003). *Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş I*, Nobel Yayın Dağıtım, Ankara, s 408.

Atakan Cemal (2003). Anadolu Üniversitesi Bilim ve Teknoloji Dergisi, Cilt No:4 - Sayı/No: 1: 59-66.

Aytaç Mustafa (1991). Uygulamalı Parametrik Olmayan İstatistik Testleri, Uludağ Üniversitesi Basımevi, Bursa.

Baskan Şanslı (1993). Uygulamalı İstatistik, Ege Üniversitesi Basımevi, İzmir.

Benedetti Jacqueline K. (1977). On the Nonparametric Estimation of the Regression Function, Journal of the Royal Statistical Society, Series B, 39 (2), 248-253.

Beran R. J. ve Srivastava Muni S. (1985). Bootstrap Tests and Confidence Regions For Functions of a Covariance Matrix, The Annals of Statistics, 13, 95-115.

Berry William Dale (1993). Understanding Regression Assumptions.

Bickel Peter John ve Freedman David Amiel (1981). Some Asymptotic Theory For the Bootstrap, The Annals of Statistics, 9, 1196-1217.

Birkes David ve Dodge Yadolah (1993). Alternative Methods of Regression, John Wiley Sons, New York, 80–140, ABD.

Chernick Michael R. (1999). Bootstrap Methods, John Wiley and Sons, Kanada.

Clark R. M. (1977). Non-Parametric Estimation of a Smooth Regression Function. Journal of the Royal Statistical Society, Series B, 39, 107-113.

Cryer Jonathan D., Robertson Tim, Wright F. T. ve Casady R. J. (1972). Monotone Median Regression. Annals of Mathematical Statistics, 43, 1459-1469.

Çömlekçi Necla (1998). Temel İstatistik İlke ve Teknikleri Bilim Teknik Yayınevi, s: 493, Eskişehir.

Daniel W. W. (1990). Applied Nonparametric Statistics. Georgia State University, Boston, 2nd edition, p: 18-20, 426-443, ABD.

Daniel Wayne W. (1995). Biostatistics: A Foundation for Analysis in The Health Sciences. John Wiley & Sons, Inc Canada, 6nd edition. p:622.

Davison Anthony C., Hinkley David Victor (1997). Bootstrap Methods and Their Application, Cambridge University Pres, s.223.

Devroye Luc P. (1978). The Uniform Convergence of Nearest Neighbor Regression Function Estimators and Their Application in Optimization. IEEE Transactions on Information Theory, 24(2), 142-151.

Dickinson Gibbons Jean (1993). Nonparametric Statistics: An Introduction, Quantitative Applications in the Social Sciences, Sage University Paper Series

Efron Bradley (1979). Bootstrap Methods: Another Look at the Jackknife, The Annals of Statistics, 7, 1-26.

Efron Bradley (1982). The Jackknife, the Bootstrap and Other Resampling Plans, Societyfor Industrial and Applied Mathematics, Philadelphia, ABD.

Efron Bradley (1983). Estimating the Error Rate of a Prediction Rule: Improvement onCross – Validation, JASA, Vol:78, No:382, s.316.

Efron Bradley ve Gong Gail (1983). Aleisurely Look at The Bootstrap, The Jackknife and CrossValidation " The American Statistician, February, JASA, Vol. 37, No. 1, 06-48.

Efron Bradley ve Tibshirani Robert J. (1986). Bootstrap Methods for Standart Errors, Confidence Intervals, and Other Measures of Statistical Accuracy, Statistical Science,1,54-77.

Efron Bradley (1990). More Efficient Bootstrap Computations. JASA, Vol. 85, No. 409.

Efron Bradley ve Tibshirani Robert J. (1993). An Introduction to the Bootstrap, Chapman and Hall, New York, s.208.

Erdinç Altay (2001). Varlık Fiyatlama Modelleri: FVFM ve AFT ve İMKB'de Uygulaması, Yayınlanmamış Doktora Tezi, İstanbul, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü

Erilli Necati Alp ve Alakuş Kamil (2014). Non-Parametric Regression Estimation for Data With Equal Values. European Scientific Journal, February 2014 edition vol.10, No.4, p.70-82.

Erilli Necati Alp (2015). İstatistik 2, Seçkin Yayıncılık, 2015, s.175.

Erilli Necati Alp ve Alakuş Kamil (2016). Parameter Estimation in Theil-Sen Regression Analysis With Jackknife Method. Eurasian Academy of Sciences, Eurasian Econometrics, Statistics & Emprical Economics Journal, Volume: 5 S: 28 - 41.

Ertek Tümay (2000). Ekonometriye Giriş, Beta Yayınları, No. 652, 334 s, Ankara.

Fox John (1997). Applied Regession Analysiss Linear Models and Related Methods. London, s.494-520, İngiltere.

Fox John (1997). Applied Regression Analysis, Linear Models and Related Methods, London, s.501, İngiltere.

Frangos Christos C. ve Schucany William R. (1990). Jackknife Estimation of the Bootstrap Acceleration Constant, Comp. Statist. Data Anal. , 9, 271-282.

Fukunaga Keinosuke, Narendra Modi (1975). A Branch and Bound Algorithm for Computing k-Nearest Neighbors, IEEE Transactions On Computing, 24, 750-753.

Galton Francis (1886). Regression Towards Mediocrity in Hereditary Stature, Journal of Anthropological Institute of Great Britain and Ireland, Vol. 15, 246–263

Gasser Thomas, Müller Hans Georg ve Mammıtzsch Volker (1985). Kernels for Nonparametric Curve Estimation. Journal of the Royal Statistical Society, Ser. B,47(2),238-252.

Gijbels Irene, Pope Andrew, Wand Matt (1999). Understanding Exponential Smoothing via Kernel Regression, Journal of the Royal Statistical Society, Series B (Statistical Methodology) , Vol 61, No.1 pp. 39-50.

Goodman Leo A. , Kruskal Wallis H. (1963). Measures of Association for Cross Classifications. III: Approximate Sample Theory. Journal of the American Statistical Association, 58, 310-364.

Green Peter, Jennison Christopher, Seheult Allan (1985). Analysis of Field Experiments by Least Squares Smoothing. Journal of the Royal Statistical Society, Series B, 47 (2), 299-315.

Griffin James (1962). Statistics Methods an Aplication Rinehart and Winston, s259-273.

Gujarati Damodar N. (2005). Temel Ekonometri. (Çev. Ümit Şenesen ve Gülay Şenesen), Literatür Yayıncılık, 850 s, İstanbul.

Hahn Gerald J. ve Meeker William Q. (1991). Statistical Intervals: A Guide for Practitioners, Wiley, New York, ABD.

Hall Peter ve Marron Stephen (1990). On Variance Estimation in Nonparametric Regression. Biometrika,77(2),415-419.

Hand David J. (1986). Discrimination and Classification, Wiley & Sons, New York, ABD.

Hardle Wolfgang ve Gasser Theo (1984). Robust Non-Parametric Function Fitting. Journal of the Royal Statistical Society, Ser. B, 46(1), 42-51.

Hardle Wolfgang ve Mammen Enno (1993). Comparing Nonparametric Versus Parametric Regression Fits. The Annals of Statistics, 21(4),1926-1947.

Hardle Wolfgang (1994). Applied Nonparametric Regression. Cambridge University Press, Cambridge, İngiltere.

Hardle Wolfgang, Linton Oliver (1994). Applied Nonparametric Methods, Cowles Foundation Discussion, 1069, s. 1-23.

Hardle Wolfgang, Muller Marlene, Sperlich Stefan, WERWATZ Axel (2004). Nonparametric and Semiparametric Models. Springer, New York, ABD.

Hastie Trevor, Tibshirani Robert (1990).Generalized Additive Models, Chapman-Hall, 1990, s.9, ABD.

Heerde Harald J., Leeflang Peter S. H., Wittink Dick R. (2001). Semiparametric Analysis to Estimate the Deal Effect Curve. Journal of Marketing Research, 38:2, 197-215.

Hill Bruce McGraw (1962). A Test of Linearity Versus Convexity of a Median Regression Curve. Annals of Mathematical Statistics, 33, 1096-1123.

Hollander Myles, Wolfe Dougles (1999). Nonparameric Statistical Methods, Second Edition, Wiley Series in Probability and Statistics, ABD.

Holmes Susan (1998). Course Notes, Stanford University www.stat.standford.edu/~susan/scgn/issues/back/v62.pdf , (26.05.2006), s.208.

Horowitz Joel L. (1993). Semiparametric Estimation of a Work-Trip Mode Choice Model, Journal of Econometrics, 58, 49-70.

Ibragimov Ildar Abdulovich ve Has'minskii R. Z. (1980), On Nonparametric Estimation of Regression. Soviet Math. Dokl.,21(3), 810-814.

Jeremy Bentham (1996). Recent Developments in Bootstrapping Times Series, Federal Reserve Board, International Finance Discussion Paper Series, September 25, s.2-34.

Kılıçbay Ahmet (1980). Ekonometrinin Temelleri. İstanbul Üniversitesi İktisat Fakültesi Yayını, 67 s, İstanbul.

Kıroğlu Gülay (2001). Uygulamalı Parametrik Olmayan İstatistiksel Yöntemler. Mimar Sinan Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, İstanbul.

Kildea D. G. (1981). Brown-Mood Type Median Estimators for Simple Regression Models. The Annals of Statistics, 9(2), 438-442.

Koutsoyiannis Anna (1992). Ekonometri Kuramı. (Çev. Ümit Şenesen ve Gülay Şenesen), İstanbul Teknik Üniversitesi Matbaası, s 638, İstanbul.

Krieger Abba M. ve Bickel Peter J. (1989). Confidence Bands for a Distribution FunctionUsing the Bootstrap, J. Am. Statist. Assoc., 84, 95-100.

Lahiri Soumendra Nath (1992). On Bootstrapping M-estimators, Sankhya A, 54, 157-170.

Marascuilo Leonard ve MCSWEENEY Maryellen (1993). Nonparametric and Distribution-Free Methods for the Social Sciences.

Maritz J. S. (1979). On Theil's Method in Distribution-Free Regression. Australian Journal of Statistics, 21,30-35.

Mclachlan Geoffrey John ve Krishnan T. (1997). The EM Algorithm and Extensions, Wiley, New York, ABD.

Mood Alexander McFarlane (1950). Introduction to the Theory of Statistics, New York: McgrawHill.

Mood Alexander McFarlane ve Brown G. Wayne (1951). ’’On Median Tests for Linear Hypotheses’’. Proceedings of the Second Berkeley Symposium On Mathematical Statistics and Probability, Berkeley and Los Angeles: The Universitey of California Pres.

Mooney Christopher Z. ve Duval Robert D. (1993). Bootstrapping: A Nonparametric Approach to Statictical Inferencv -ge Un: Papers.

Mooney Christopher Z. (1996). Bootstrap Statistical Inference; Examples Anad Evaluation for Political Science, American Journal of Political Science, Vol:40, No:2, Mayıs, s.570.

Oğuzlar Ayşe (1999). Parametrik Olmayan Basit Doğrusal Regresyon Analizi, Uludağ Üniversitesi IIBF Dergisi, Cilt:17, Sayı:1-2, Mayıs.

Okutan Duygu (2009). Bootstrap Yönteminin Regresyon Analizinde Kullanımı ve Diğer Yöntemlerle Karşılaştırılması, Yayınlanmamış Yüksek Lisans Tezi, Ege Üniversitesi.

Orhunbilge Neyran (2000). Uygulamalı Regresyon ve Korelasyon Analizi, 2.Baskı, İstanbul, s.12

Özel Hasan Alp ve Sezgin Fuat (2011).Ticari Serbestleşme-Ekonomik Büyüme İlişkisinin Bootstrap Kantil Regresyon Yardımıyla Analizi. Çanakkale Onsekiz Mart Üniversitesi& İstanbul Üniversitesi

Pagan Adrian, Ullah Aman (1999). Nonparametric Econometrics, Cambridge University Press, İngiltere.

Potthoff Richard F. (1974). A Non-Parametric Test of Whether Simple Regression Lines are Parallel. The Annals of Statistics, 2 (2), 295-310.

Priestley Maurice B. ve Chao Min Te (1972). Nonparametric Function Fitting, Journal of the Royal Statistical Society, Series B, 34, 385-392.

Rao KS. Madhava ve Gore Anil P. (1982). Nonparametric Tests for İntercept in Linear Regression Problems. Australian Journal of Statistics, 24 (1), 42-50.

Rice John (1984). Banswith Choice for Nonparametric Regression. The Annals of Statistics, 12(4), 1215-1230.

Ruppert David, Wand M.P, Carroll R. C. (2003). Semiparametric Regression, Cambridge University Press, 2003, s.57, ABD.

Schenkar Nathaniel (1985). Qualms About Bootstrap Confidence Intervals, JASA, Haziran, Vol 80. N.390, s.360.

Sen P. K. (1968). Estimates of the Regression Coefficient Based on Kendall's Tau. J. Amer.Statist.Ass., 63, 1379-1389.

Shao Jun (1996). Bootstrap Model Selection. JASA, Voı. 91, No. 434."55-665.

Sievers Gerald L. (1978). Weighted Rank Statistics For Simple Linear Regression. J.Amer.Statist.Ass., 73, 628-631.

Silverman Bernard Walter (1984). Spline Smoothing: The Equivalent Variable Kernel Method. The Annals of Statistics, 12(3), 898-916.

Silverman Bernard Walter (1986). Density Estimation for Statistics and Data Analysis, Chapman & Hall, London, İngiltere.

Simonoff Jeffrey S. (1998). Three Sides of Smoothing: Categorical Data Smoothing, Nonparametric Regression, and Density Estimation. J. International Statistical Review, 66(2),137-156.

Singh Kesar (1981). On the Asymptotic Accuracy of Efron's Bootstrap, The Annals of Statistics, 9, 1187-1195.

Spaeth Helmuth (1991). Mathematical Algorithms for Linear Regression. London: Academic Press, s:304, İngiltere.

Speckman Paul (1985). Spline Smoothing and Optimal Rates of Convergence in Nonparametric Regression Models. The Annals of Statistics, 13(3), 970-983.

Sprent Peter, Smeeton Nigel C. (2007).Applied Nonparametric StatisticalMethods, Text in Statistical Science, Fourth Edition, Chapman&Hall, ABD.

Staudte Robert G. ve Sheather Simon J. (1990). Robust Estimation and Testing, Wiley, New York, ABD.

Stine Robert (1985). Bootstrap Prediction Intemʳals for Regression. Jour. Of the Amer. Stat. Assoç., Voı. 80, No. 392, 1027p.

Stoffer David S., Wall K. D. (1991). Bootstrapping State Space Models, Gaussian Maximum Likelihood Estimation and Kalman Fitler, JASA, Aralık, Vol.56. N.416, s.1024.

Stone Charles J. (1977). Consistent Nonparametric Regression, The Annals of Statistics, Vol:5, No: 4, 595-645.

Strawderman Robert L., Wells Martin T. (1997). Accurate Bootstrap Confidence Limits for the Cumulative Hazari and Survivor Functions Under Random Censoring, Journal of the American Statistical Association, 92, 1356-1374.

Şengün Meltem (1999).“Yeniden Örnekleme Metoduna Nonparametrik Yaklaşım” IV. Ekonometri ve İstatistik Sempozyumu, Antalya, 14-16 Mayıs, s.1030.

Tate Merle Wesley, Clelland Richard C. (1957). Nonparametric and Shorulcut Statistics in the Social. Biological and Medical Sciences. Danvill, III.: Interstate Printers and Publishers.

Theil Henri (1950). ‘’A Rank Invariant Method of Linear and Polynomial Regression Analysis. III.’’ Nederl. Akad. Wetensch.Proc., Series A, 53, 1397-1412.

Topuz Derviş (2002). Regresyonda Yeniden Örnekleme Yöntemlerinin Karşılaştırmalı Olarak İncelenmesi Yayınlanmamış Yüksek Lisans Tezi, Niğde Üniversitesi.

Ünver Özkan ve Gamgam Hamza (2008). Uygulamalı Temel İstatistik Yöntemler. Seçkin Kitapevi, 424 s, Ankara.

Wang X. Yu, Q. (2004). Unbiasedness of the Theil-Sen Estimator. http://home.olemiss.edu/xuegin/papers/tsun.pdf

Wu C. F. J. (1986). Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. Am. Of Stat., Vol. 14pp, No. 4, 1261-1295.

Yatchew Adonis (1998). Nonparametric Regression Techniques in Economics. Journal of Economic Literature, Vol. 36, No.2. 669-721.

Yatchew Adonis (2003). Semiparametric Regression for the Applied Econometrican. Cambridge University Press, İngiltere.

Yıldız Nesrin ve Topal M. (2001). Nonparametrik Regresyon Metodlarının İncelenmesi. Atatürk Üniversitesi, Ziraat Fakültesi Dergisi, 32(4), 429-435.

Yıldız Nesrin, Topal Mehmet ve Bilgin Ömer Cevdet (2004). Farklı Dağılış Gösteren Verilerde Parametrik ve Nonparametrik Regresyon Metotlarının İncelenmesi. 4 Ulusal Zootekni Bilim Kongresi Sözlü Bildiriler Programı, Atatürk Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Erzurum.

Zeng Q. ve Davidian M. (1997). Bootstrap-Adjusted Calibration Confidence Intervals for Immunoassay, Joumal ofthe American Statistical Association, 92, 278-290.

ÖZ GEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı	**:** Nilüfer Betül AKMAN
Uyruğu	**:** T.C.
Doğum Tarihi ve Yeri	**:** 12/04/1993 Kırıkkale
e-posta	**:** niluferbetulakman@gmail.com

EĞİTİM

Derece	**Kurum**	**Mezuniyet Yılı**
Lisans	Cumhuriyet Üniversitesi	2011-2015
Yüksek Lisans	Cumhuriyet Üniversitesi	2015-2018

İŞ TECRÜBESİ

Tarih	**Kurum**	**Görev**

YABANCI DİL BİLGİSİ

KPDS ( ) **ÜDS ( )** **TOEFL ( )** **EILTS ( )**