ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

Savaş OKUR

PARAMETRİK VE PARAMETRİK OLMAYAN BASİT DOĞRUSAL REGRESYON ANALİZ YÖNTEMLERİNİN KARŞILAŞTIRMALI OLARAK İNCELENMESİ

ZOOTEKNİ ANABİLİM DALI

ADANA, 2009

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

PARAMETRİK VE PARAMETRİK OLMAYAN BASİT DOĞRUSAL REGRESYON ANALİZ YÖNTEMLERİNİN KARŞILAŞTIRMALI OLARAK İNCELENMESİ

Savaş OKUR

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOTEKNİ ANABİLİM DALI

Bu Tez 20/03/2009 Tarihinde Aşağıdaki Jüri Üyeleri Tarafından Oybirliği/ Oyçokluğu İle Kabul Edilmiştir.

İmza………….	İmza…………	İmza…………..
Doç.Dr. Nazan KOLUMAN DARCAN	Prof.Dr.G. Tamer KAYAALP	Doç.Dr. Suat ŞAHİNLER
DANIŞMAN	ÜYE	ÜYE

Bu tez Enstitümüz Zootekni Anabilim Dalında hazırlanmıştır.

Kod No:

Prof.Dr. Aziz ERTUNÇ
Enstitü Müdürü
İmza ve Mühür

ÖZ

YÜKSEK LİSANS TEZİ

PARAMETRİK VE PARAMETRİK OLMAYAN DOĞRUSAL REGRESYON ANALİZ YÖNTEMLERİNİN KARŞILAŞTIRMALI OLARAK İNCELENMESİ

Savaş OKUR

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOTEKNİ ANABİLİM DALI

Danışman: Doç.Dr. Nazan KOLUMAN DARCAN
Yıl : 2009, Sayfa: 53
Jüri : Doç.Dr. Nazan KOLUMAN DARCAN
Prof.Dr. G. Tamer KAYAALP
Doç. Dr. Suat ŞAHİNLER

Bu çalışmada, parametrik ve parametrik olmayan basit doğrusal regresyon analiz yöntemlerinin karşılaştırmalı olarak incelenmesi amaçlanmıştır.

Parametrik basit doğrusal regresyon analizinde e.k.k yöntemi, parametrik olmayan basit doğrusal regresyon analizinde medyana göre parametre tahmini yapan, Mood-Brown ve Theil yöntemleri, parametrik olmayan basit doğrusal regresyon fonksiyonunun tahmin yöntemi, en yakın komşu (k-NN) tahmin yöntemleri tanıtılmıştır.

Yaş ve doğum ağırlığı arasında doğrusal bir ilişkinin olup olmadığı hem parametrik hem de parametrik olmayan basit doğrusal regresyon analiz yöntemleriyle incelenmiştir. Parametrik basit doğrusal regresyon analizinde yaş ve doğum ağırlığı arasında ilişki olmazken, parametrik olmayan basit doğrusal regresyon analizinde ise, Mood-Brown ve k-NN yöntemlerinden, yaş ve doğum ağırlığı arasında ilişki elde edilmiş, Theil yönteminden ise ilişki elde edilmemiştir.

Anahtar Kelimeler: Parametrik regresyon, Parametrik olmayan regresyon, Medyan, Fonksiyon tahmin.

ABSTRACT

MSc THESIS

INVESTIGATION OF THE COMPARISONS OF PARAMETRIC AND NON PARAMETRIC LINEAR REGRESSION ANALYSIS METHODS

Savaş OKUR

DEPARTMENT OF ANIMAL SCIENCE
INSTITUTE OF NATURAL AND APPLIED SCIENCES
UNIVERSITY OF ÇUKUROVA

Supervisor: Doç.Dr. Nazan KOLUMAN DARCAN
Year : 2009, Sayfa: 53
Jury : Doç.Dr. Nazan KOLUMAN DARCAN
Prof.Dr. G. Tamer KAYAALP
Doç. Dr. Suat ŞAHİNLER

The aim of this work to investigate the methods of parametric and non-parametric simple linear regression calculus by comparing them.

The methods of Mood Brown and Theil, which is least square methods for parametric simple linear regression calculus and parameter estimate according to median for non-parametric simple linear regression calculus, was introduced. It was also introduced the estimate method of non-parametric simple linear regression function and the nearest neighborhood estimate methods (k-NN).

We investigate the question “Are there any linear relations between age and weight of born” using parametric and non-parametric simple linear regression methods.

In parametric case, there are no linear relations between them but in non-parametric case, it is obtained a relation from Mood-Brown and k-NN methods. On the other hand it isn’t obtained any relation from Theil methods.

Keywords: Parametric regression, Nonparametric regression, Median, Function estimation.

TEŞEKKÜR

Bu çalışmayı yöneten ve değerli zamanlarını harcayarak yakın ilgi ve yardımlarını esirgemeyen sayın hocam Prof.Dr. G. Tamer KAYAALP'e, Arş.Gör. Gülsen KIRAL ve Arş.Gör. Soner ÇANKAYA' ya teşekkürlerimi sunarım.

İÇİNDEKİLER SAYFA

ÇİZELGELER DİZİNİ **SAYFA**

ŞEKİLLER DİZİNİ **SAYFA**

1.GİRİŞ

Regresyon analizi, aralarında sebep- sonuç ilişkisi bulunan iki veya daha fazla değişken arasındaki ilişkiyi inceler. Regresyon analizinden ekonomi, fizik, kimya, biyoloji, sosyal bilimler gibi bir çok alanda yararlanılmaktadır. Bir değişkenin değerinin diğer değişkenlerdeki değişimlere bağlı olarak nasıl etkilendiğinin istatistik analizlerle incelenmesi çeşitli nedenlerle istenmektedir.

Değişkenler arasındaki ilişki bilindiğinde, bir değişkenin değerine bakılarak diğeri tahminleneceği gibi, etki eden faktörler kontrol altına alınabilirse, ilgili değişkenlerin değerleri optimum düzeye gelebilir. İstatistik tahminin çoğu uygulamasında, bir grup değişken arasındaki ilişkinin değerlendirilebilmesi için örnek verileri kullanılarak değişkenler arasındaki ilişkinin modellenmesi gerekmektedir. Böylece elde edilen model sayesinde, bağımlı değişken olarak seçilen değişkenin gelecekteki herhangi bir değeri tahmin edilir.

Veriler arasındaki ilişkiyi tanımlayan en uygun modelin bulunmasını sağlayan istatistiksel tekniğe regresyon analizi adı verilir. Basit regresyon analizi bağımlı değişken ile bağımsız değişken arasındaki doğrusal ilişkiyi açıklar. Eğer bağımlı değişken ile birden fazla bağımsız değişken arasında doğrusal veya eğrisel bir ilişki varsa bu ilişki çoklu regresyon analizi ile incelenir (Işık, 2006). Basit doğrusal regresyon analizi, bağımlı değişken (Y) ile bir bağımsız değişken (X) arasındaki ilişkinin doğrusal fonksiyonla ifade edilmesidir.

$$\hat{Y}_i = \hat{a} + \hat{b}X_i \tag{1.1}$$

(1.1) eşitliği basit doğrusal regresyon eşitliği olarak izah edilir. $\hat{a}$ doğrusal fonksiyonun sabiti, $\hat{b}$ ise, doğrusal fonksiyonun eğimidir.

X ile Y değerlerinin dağılımını gösteren serpilme diyagramlarında doğrusal bir eğim gözüküyorsa, X’ in Y’ e göre fonksiyonun doğrusal olduğuna karar verilir.

Günümüzde parametrik istatistik yöntemlere karşılık gelen nonparametrik yöntemler geliştirilmiştir. Bu yöntemlerden bazıları da parametrik olmayan regresyon yöntemleridir.

Regresyonda parametre tahminleri genelde En Küçük Kareler (E.K.K.) yöntemine göre yapılmaktadır. En küçük kareler tahmin edicileri matris notasyonuyla,

$$\hat{b} = (X'X)^{-1}X'Y \tag{1.2}$$

eşitliği yardımıyla hesaplanmaktadır.

E.K.K. tahmin edicileri, regresyon modelindeki bir stokastik değişken olan bağımlı değişkenin değerlerinin doğrusal bir fonksiyonu ve parametreleri bakımından sapmasız olduğunda, minimum varyansa sahiptirler, yani en iyidirler. Fakat tahminlerin standart hataların küçük ve dolayısı ile parametrelerin etkin olması, çoklu doğrusal regresyon modelinin varsayımlarının gerçekleşmesine bağlıdır. Parametrik olmayan regresyon tahmin yöntemleri, medyana göre regresyon parametrelerini tahmin yöntemleri ve regresyon fonksiyonunun tahmin yöntemleri olmak üzere iki kısımda toplanmaktadır.

Parametrik ve parametrik olmayan basit doğrusal regresyon modeli eşitlik (1.3) ile izah edilir.

$$Y_i = a + bX_i + e_i \tag{1.3}$$

Eşitlikte;

Y: (nx1) boyutlu şans değişkeni vektörünü,

X: (nxp) boyutlu bilinen katsayı matrisini,

β : (nx1) boyutlu bilinmeyen parametre vektörünü,

α : kısmi regresyon katsayısını,

ε_i : i. hata terimini ifade etmektedir.

Regresyon analizinde hata teriminin bağımsız, ortalaması sıfır, varyansı s^2 olan normal dağılım gösterdiği varsayılır. Elde edilen model için bu varsayımlar gerçekleşmediği takdirde o model ile ilgili her türlü yorum şüphe ile karşılanır.

En küçük kareler regresyon analizinde hata teriminin ortalaması sıfır, varyansı sabit olduğu ve birbirleriyle korelasyonsuz olduğu varsayılır (Şahinler, 2000). Yani,

$E(e_i) = 0$,

$V(e_i) = s^2$,

$E\left[\mathrm{cov}(e_{i,} e_j)\right] = 0$ ‘ dır.

Basit regresyon tahmin edicilerinin diğer bir grubunda ise parametre tahminleri veri kümesinden ayırmaksızın örnek birimleri ikişerli olarak ele alındığında tüm durumlardaki eğimlerin hesaplanması ile elde edilmektedir. Bu yönteme Theil Tekniği adı verilmektedir (Gamgam, 1989).

Bu çalışmanın amacı: önceki çalışmaların ışığı altında, parametrik ve parametrik olmayan basit doğrusal regresyon analiz yöntemlerini uygulama yönünden tanıtmak ve yöntemleri karşılaştırmaktır. Bu yöntemlerde karşılaştırma kriteri olarak belirtme katsayısı (R^2), hata kareler ortalaması (H.K.O.), Theil‘ in U_1 ve U_2 katsayı değerleri kullanılmıştır.

2. ÖNCEKİ ÇALIŞMALAR

Theil (1950), b_i eğim katsayılarının nokta tahminini bulmayı sağlayan bir yöntem geliştirmiştir. Eğim bulmaya yönelik olarak önerilen Mood-Brown yönteminin hızlı ancak çok güvenilir bir yöntem olmadığını belirten Theil özellikle eğim katsayısını bulmak için kendi adıyla anılan bir yöntem geliştirmiştir.

Mood (1950), Mood-Brown hipotez testi yöntemine bağlı olarak β katsayısı için güven aralığı tahminini bulan bir deneme yanılma tekniği geliştirmiştir. Mood-Brown yöntemindeki α ve β' ya ilişkin hipotez testinde n_1 ve n_2' nin 0.5 parametresi ile binom dağıldığını belirtmiş ve bu bilgiye dayalı olarak Mood-Brown yöntemindeki test ölçütünü geliştirmiştir.

Akdeniz (2001), doğrusal regrasyon modelinde, iç ilişki durumunda, RR, genelleştirilmiş inverse tahmin edicisi, temel bileşenler regrasyonu, Liu tahmin edicisi ve düzeltilmiş Ridge ve ve Liu tahmin edicisinin, EKK tahmin edicisinin düzeltilmesinde kullanıldığını belirtmiştir. Çalışmada yanlı bir tahmin edici olan Liu tahmin edicisi ve Liu tahmin edicisinin geliştirilmiş bir versiyonu olan yaklaşık sapmasız genelleştirilmiş Liu tahmin edicisi ve bunların kalıntılarını analiz etmiş ve HKO terimleri içindeki EKK kalıntıları ile karşılaştırmıştır. Sonuç olarak Liu tahmin edicisine ait α parametresinin seçiminin β parametresi ve s^2 ‘nin doğruluğuna bağlı olduğunu ve HKO bakımından Liu tahmin edicisinin ve yaklaşık sapmasız genelleştirilmiş Liu tahmin edicisinin EKK tahmin edicisinden daha üstün olduğunu göstermiştir.

Mood-Brown (1951), α ve β_i katsayılarını belirleyen ve kendi isimleri ile anılan bir yöntem geliştirmiştir. 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 yapmıştır.

Sen (1968), iki veya daha fazla eğim parametresinin birbirine eşit olduğunu iddia eden sıfır hipotezlerini test eden bir sıra puanı yöntemini incelemiştir. Kendall' ın Tau' sundan esinlenerek b_i ' nin basit ve sağlam tahmincileri üzerinde çalışmıştır.

Kendall' ın Tau test ölçütü hesabının güç ve etkinliğini incelemiştir. Nokta tahmincisini, $x_i \neq x_j$ ile noktaların $(y_j\text{-}y_i)/(x_j\text{-}x_i)$ eğim çiftleri populasyonunun medyanı olarak tarif etmiştir. Sen, ileri sürdüğü tahmincilerin çeşitli özelliklerini incelemiş ve kendi ismini verdiği yöntemi E.K.K. 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.

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..

Sievers (1978), linear regresyon modelinde β_i parametresinin tahminini ve güven aralığını incelemiştir. Theil yönteminde güven aralığının i<j olmak üzere düzenlenmiş b_{ij} eğimler setinde uygun olarak seçilmiş sınırlı notalara sahip olduğunu belirtmiştir.

Györfi (1981), k_n –NN regresyon tahminini işlemiş ve k_n –NN regresyon tahminlerinin yakınsama oranını göstermiştir.

Hussain ve Sprent (1983), çeşitli parametrik olmayan regresyon modelleri arasındaki karşılaştırmaları yapmışlardır. Medyan tahmincileri e.k.k.' de uygun ağırlık yöntemlerini kullanarak, kesinlikle sapan gözlemlerin etkisinin öneminin azaltılmasının mümkün olacağını belirtmişlerdir. α ve β' nın tahminlerinin medyan hesabına dayalı Theil yöntemini incelemişlerdir.

Aytaç (1984), bilinmeyen dağılım fonksiyonu sürekli olan veya onun üzerinde herhangi bir bilgiye gerek duymayan yöntemler parametrik olmayan yöntemlerdir.. Ayrıca parametrik olmayan test, populasyon parametrelerinin değerleri üzerine hiçbir varsayım yapılmadığı durumlarda uygulanır.

Rice (1984), regresyon doğrusunun tahmini için düzleştirme parametresinin seçilmesi problemi ile ilgili olarak yaptığı çalışmada, fourier serisi ile Kernel fonksiyonu arasındaki ilişkiyi araştırmıştır. Ayrıca hata kareler ortalamasının minimum ve düzleştirme parametresinin seçiminde asimptotik optimallik durumlarını teorik olarak incelemiştir.

Silverman (1986), verilerde yoğunluk fonksiyonunun parametrik olmayan tahmininden yola çıkarak, düzleştirme parametresinin seçimi, Kernel tahmin yönteminin ortalaması, varyansı, seçimi ve yanı 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 belirtmek için çalışmalar yapmıştır.

Daniel (1990), parametrik olmayan regresyon yöntemleri konusunda derleme çalışması yapmıştır. Theil tarafından eğimin testi için önerilen yöntemin Kendall’ ın Tau istatistiğine dayandığını belirtmiştir.

Lesaffre ve Marx (1993), genelleştirilmiş doğrusal regrasyonda kötü koşulluluk problemini incelemiştir. Bağımsız değişkenler arasındaki iç ilişkiyi, kovaryans matrisi üzerinde benzer zararlı etkiye sahip olan ML- iç ilişki diye adlandırılan kötü koşulluluğun diğer tipini tanımlamıştır. Ml- iç ilişkide orijinal bağımsız değişkenler arasında iç ilişki olmadığını fakat değişken ve model seçimi ile bağımsız değişkenlerin bileşimi yüzünden bu durumun ortaya çıktığını bildirmiştir. Bu yüzden kötü koşulluğu tanımlamak için bilgi matrisinden yararlanıldığı ve iç ilişki ile ML- iç ilişki arasındaki en önemli farklılığın bu olduğunu göstermiştir. ML- iç ilişki durumuna örnek olarak lojistik regrasyon modelini vermiştir. Fakat örnek olarak genelleştirilmiş doğrusal regrasyon modelinin herhangi bir üyesinin de seçilebileceğini söylemiştir. Sonuç olarak ML- iç ilişki problemini gidermede birkaç alternatif yöntem önermiştir. Bunlardan birisi değişken seçim yöntemidir. Fakat bu yöntemin bilgi matrisinin kötü koşulluluğunun azaltılması için daima en iyi yöntem olmadığını bildirmiştir. Bir diğer yöntem Ridge lojistik tahmin edicisi ile adımsal temel bileşenler yöntemidir. Temel bileşenler modeli ile kötü koşululuk probleminin azaldığını ve ML-iç ilişki varken bu yöntemin iyi bir yöntem olduğunu söylemiştir.

Mackinnon ve Puterman (1989), genelleştirilmiş doğrusal modeller(GLM) için iç ilişkiyi, tanımlamış, bunun sonuçlarını incelemiş ve teşhis kriterlerini sunmuştur. Sıradan korelasyon matrisinin iç ilişkiyi tanımlamada daima yeterli olmayacağını ve bunun yerine iç ilişki teşhisinde bilgi matrisinin kullanılması gerektiğini bildirmiştir. GLM için dağılımların ağırlıklarını vermiştir. Teorem 1'de ağırlıklar arasında fark yoksa o zaman GLM ve standart doğrusal modelin(SLM) iç ilişki bakımından birbirine benzediğini ortaya koymuştur. Fakat ağırlıklar farklıysa bu durumda SLM' nin iç ilişkili olmamasına rağmen GLM' nin iç ilişkili olabildiğini bildirmiştir. Son olarak, yanlı tahmin yöntemlerinden biri olan Ridge yöntemleri üzerine çalışmıştır. Sonuçta eğer birkaç yüksek iç ilişkili veri varsa Ridge metodu ile HKO azaldığını bildirmiştir.

Ergüneş (2004), çoklu bağlantı problemi olan bir örneğe öncelikle en küçük kareler yöntemi(EKK) daha sonra da RR yöntemi uygulamıştır. Karşılaştırma kriteri olarak $E\lfloor L_1^2 \rfloor$ ve R^2 kullanmıştır. Sonuç olarak, çoklu bağlantı varlığında, $E\lfloor L_1^2 \rfloor$'yı azaltması, R^2'nin fazla azalma göstermemesi, çoklu bağlantı problemini gidermesi ve dolayısıyla EKK yönteminden daha etkin parametre tahmini yapması sebebiyle RR yöntemi, EKK yöntemi yerine önermiştir.

3. MATERYAL VE METOT

3.1. Materyal

Bu çalışmada, Çukurova Üniversitesi, Ziraat Fakültesi, Araştırma Uygulama Çiftliğindeki farklı yaştaki koyunların yaş ve doğum ağırlığı verileri kullanılmıştır. Bağımsız değişken olarak koyunların yaşı (yıl), bağımlı değişken olarakta koyunların doğum ağırlığı (kg) olarak ele alınmıştır. Bu çalışma, koyunların büyük veya küçük yaşta olmasının doğum ağırlığı ile ilişkisinin nasıl olduğunu incelemek amacıyla yapılmıştır.

Çizelge 3.1' de yer alan yaş faktöründe hesaplamalar yıl, ay ve gün olarak alınmıştır. Bu nedenle ondalıklar; örneğin 1. Sıradaki 4.1 yaş, 4 yıl 1 ay olarak değerlendirilmelidir.

Çizelge 3.1. Koyunların Yaşı (X) ve Doğum Ağırlığı (Y) Değerleri Tablosu

Sıra No:	Yaş (X)	Doğum Ağırlığı (Y)	Sıra No:	Yaş (X)	Doğum Ağırlığı (Y)	Sıra No:	Yaş (X)	Doğum Ağırlığı (Y)
1	4.1	6.2	**21**	1.8	2.4	**41**	2.18	4.3
2	4.2	3.1	**22**	1.9	3.5	**42**	2.19	5.0
3	3.1	4.0	**23**	3.11	5.0	**43**	1.11	6.0
4	3.2	4.1	**24**	3.12	3.65	**44**	1.12	5.5
5	3.3	4.5	**25**	3.13	5.5	**45**	1.13	4.5
6	3.4	4.0	**26**	3.14	3.5	**46**	1.14	4.0
7	2.1	4.9	**27**	3.15	5.5	**47**	1.15	6.0
8	2.2	5.2	**28**	3.16	4.75	**48**	1.16	3.8
9	2.3	5.0	**29**	3.17	4.0	**49**	1.17	4.0
10	2.4	6.0	**30**	3.18	4.5	**50**	4.10	5.0
11	2.5	5.1	**31**	3.19	4.5	**51**	3.10	4.0
12	2.6	5.0	**32**	3.21	3.5	**52**	2.10	3.4
13	2.7	3.9	**33**	3.22	4.55	**53**	2.10	4.0
14	1.1	4.15	**34**	2.11	3.0	**54**	2.10	3.9
15	1.2	3.7	**35**	2.12	5.2	**55**	2.10	3.5
16	1.3	3.4	**36**	2.13	4.5	**56**	1.10	3.5
17	1.4	4.75	**37**	2.14	4.5	**57**	1.10	3.3
18	1.5	5.2	**38**	2.15	4.8			
19	1.6	5.5	**39**	2.16	3.0			
20	1.7	2.5	**40**	2.17	5.0			

3.2. Metot

Bu çalışmada ele alınacak, parametrik ve parametrik olmayan basit doğrusal regresyon modelinde parametrelerin tahmin edilmelerinde kullanılan yöntemler aşağıda verilmiştir.

- En Küçük Kareler Yöntemi
- Mood-Brown Yöntemi
- Theil Yöntemi
- En yakın komşu Yöntemi (k-NN)

3.2.1. En Küçük Kareler Yöntemi

En küçük kareler (E.K.K.) yöntemi, parametrik basit doğrusal regresyon analiz modellerinde parametre tahminlerinde kullanılır. Regresyon analiz modeli, ilgilenilen problemle ilgili örnek olarak alınmış gözlem değerleri kullanılarak hesaplanılmaktadır. Kurduğumuz analiz modelindeki değerler yöntemlerden elde edilen tahmini değerlerdir. Tahmin edilmeye çalışılan sonuç değişkeni (Y) ve sebep değişkeni (X) ile izah edilir. Çalışmada kullanınlan regresyon modeli (3.1) eşitliğindeki gibidir.

$$\hat{Y}_i = \hat{a} + \hat{b}X_i \tag{3.1}$$

(3.1) eşitliğinde

$\hat{a}$: X değerinin sıfır olduğu durumda Y' nin alacağı değer

$\hat{b}$: X değeri 1 birim arttığı zaman Y değişkeninin kendi birimi cinsinden değişeceği miktar olarak adlandırılmaktadır.

Regresyon analizi için kurulan modelde, bağımlı ve bağımsız değişkenin yanı sıra hata terimi olarak isimlendirilen değişken yer alır. Hata teriminin modele alınma nedenleri:

- Modele alınan Y ve X_i değişkenlerinin yapılan araştırmalarda yanlış ölçülmüş olabilmesi ve,
- Seçilen değişkenler Y ve X_i' lerin hatalı sayıda alınmış örnekler olabilmesidir.
- İster basit regresyon, ister çoklu regresyon modeline bakılıyor olsun, kurulacak modelde bağımlı değişken ile ilişkisi olan model dışında da bağımsız değişkenlerin olabilmesidir.

Bu unsurlar genel olarak e_i hata terimi olarak alınır ve minimum yapılmaya çalışılır. Bunu yaparken de E.K.K. yönteminden yararlanılır.

3.2.1.1. Parametrik Basit Doğrusal Regresyonda Kullanımı

Basit regresyon analiz modelinde bağımlı değişkeni açıklayan bir bağımsız değişken modelde yer almaktadır. Basit regresyon analiz modelleri e.k.k.. yöntemi kullanılarak çözümlenebilmektedir.
Bir bağımsız değişkeni içeren model aşağıdaki gibi gösterilir.

$$Y_i = a + bX_i + e_i \tag{3.2}$$

buradaki e_i hata terimi (3.3)' deki eşitlikten aşağıdaki gibi elde edilir.

$$e_i = Y_i - \hat{Y}_i \tag{3.3}$$

(3.1) eşitliğindeki parametrelerin E.K.K. tahmininin amaç fonksiyonunu minimum yapan $\hat{a}$ ve $\hat{b}$ tahminleri elde edilmek istenir. Burada amaç fonksiyonu:

$$e_i = \sum e_i^{\,2} = \sum (Y_i - \hat{Y})^2$$
$$= \sum [Y_i - (\hat{a} + \hat{b}X_i)]^2 \tag{3.4}$$
$$= \sum_{i=1}^{n} (Y_i - \hat{a} - \hat{b}X_i)^2$$

minimum yapılmaya çalışılır.

$\hat{a}$ ' ya göre kısmi türev alınıp sıfıra eşitlenir. Böylece (3.5a) eşitliği elde edilir

$$\frac{\partial e}{\partial \hat{a}} = 2\sum (Y_i - \hat{a} - \hat{b}X_i)(-1) = 0 \tag{3.5a}$$

$\hat{b}$ ' e göre kısmi türev alınıp sıfıra eşitlenir. Böylece (3.5b) eşitliği elde edilir

$$\frac{\partial e}{\partial \hat{b}} = 2\sum (Y_i - \hat{a} - \hat{b}X_i)(-X_i) = 0 \tag{3.5b}$$

eşitliklerin çözümünden (3.5a) ve (3.5b) eşitlikleri bulunur.

$$n\hat{a} + \hat{b}\sum X_1 = \sum Y_i \tag{3.5a}$$
$$\hat{a}\sum X_i + \hat{b}\sum X_i^2 = \sum X_i Y_i \tag{3.5b}$$

Bu eşitliklerin çözümünden de eşitlik (3.6) elde edilir.

$$\hat{a} = \frac{\sum Y_i}{n} - \frac{\hat{b}\sum X_i}{n}$$
$$\hat{a} = \overline{Y} - \hat{b}\overline{X}_i \tag{3.6}$$

diğer eşitliklerde $\hat{a}$ yerine eşitliği koyulur.

$$\hat{b} = \frac{\sum_{i}^{n} X_i Y_i - \frac{\sum_{i}^{n} X_i \sum_{i}^{n} Y_i}{n}}{\sum_{i}^{n} X_i^{\,2} - \frac{(\sum_{i}^{n} X_i)^2}{n}} \tag{3.7}$$

olarak bulunur. Buna göre basit doğrusal regresyon analiz modeli (3.8) eşitliği ile elde edilir.

Regresyon modeli tahmin edildikten sonra eşitliğin uygunluğuna, parametrelerin önem testine bakmak gerekmektedir. Eşitliğin uygunluğu, belirtme katsayısı olarak tanımlanır ve R^2 ile gösterilir.

Parametrik Basit Doğrusal Regresyon Analiz Modelinde Belirtme Katsayısının Hesaplanması: Belirtme katsayısı kullanılan X_i değişkenlerinin Y' deki toplam varyansı açıklayabilme oranıdır ve $0 < R^2 < 1$' dir (Newton, 1996). Bu katsayı (3.8) eşitliği ile verilir.

$$R^2 = \text{Re}\,g.K.T./G.K.T. \tag{3.8}$$

R^2' nin büyük çıkması her zaman modelin iyi olduğu sonucunu göstermez. Çünkü, modele konu ile ilgili veya ilgisiz bir değişkenin eklenmesi R^2' nin değerini arttırır. Dolayısıyla da büyük R^2' si olan modeller her zaman tahmin yapmada en iyi model olmayabilir (Montgomery ve Peck, 1992). Ancak modele giren değişkenler yönünden bir problem yoksa pratikte iyi bir ölçüdür.

Düzeltilmiş Belirtme Katsayısı: Bu katsayı, belirtme katsayısının R^2 serbestlik derecesine göre düzeltilmişidir.

$$R_d^2 = 1 - \{[(H.K.T./(n-k))]/[(G.K.T./(n-1))]\} = 1 - [(n-1)/(n-k)](1-R^2) \tag{3.9}$$

Yukarıdaki bilgilere ek olarak pratikte, modele giren bağımsız değişkenler ve gözlem sayısının yeterliliği konusunda ön bilgiler verebilir (Levine ve ark., 1997).

Eğer R^2 ile R_d^2 değerleri çok farklı değilse basit olarak kullanılan gözlem sayısının yeterli olduğu, aksi durumunda ise anlamlı katkıları olmayan değişkenlerin modele dahil edildiği anlamını taşır. Modelin yeterliliği konusunda bilgi vermez (Şahinler, 2000).

Tahmin edilen $\hat{a}$ ve $\hat{b}$ parametrelerinin önem testlerinin yapılabilmesi için bu parametrelere ait varyans değerlerinin hesaplanması gerekir. Bu değerler sırası ile (3.10), (3.11), (3.12), (3.13) ve (3.14) eşitliklerinde verilmiştir.

$$Var(\hat{a}) = \frac{\sum X_i^2}{n\sum (X_i - \bar{X})^2} \tag{3.10}$$

$$S(\hat{a}) = \sqrt{Var(\hat{a})} \tag{3.11}$$

$$Var(\hat{b}) = \frac{s^2}{\sqrt{\sum (X_i - \bar{X})^2}} \tag{3.12}$$

$$S(\hat{b}) = \sqrt{Var(\hat{b})} \tag{3.13}$$

$$s = \sqrt{\sum \frac{e^2}{n-2}} \tag{3.14}$$

Tahmin edilen parametrelerin önem testini yapmak için kullandığımız $\hat{a}$ ve $\hat{b}$ katsayılarının ortalama ve beklenen değerleri a ve b katsayılarına eşit olsa da populasyonun parametresine kesin eşitliği söylenemez. Tahmin değerlerinin güvenirliliğine standart hata ve varyansının küçüklüğüne bakarak populasyona yakınlığı görülür.

$\hat{b}$ için güven aralığı:

$$\hat{b} - Z_{1-a/2}.S(\hat{b}) \leq b \leq \hat{b} + Z_{1-a/2}S(\hat{b}) \tag{3.15}$$

$\hat{a}$ için güven aralığı:

$$\hat{a} - Z_{1-a/2}.S(\hat{a}) \leq a \leq \hat{a} + Z_{1-a/2}.S(\hat{a}) \qquad (3.16)$$

dir. Eşitsizliğin ilk kısmı güven aralığının alt sınırıdır. İkinci kısım ise üst sınır değeridir. Güven aralığı, bakılacak a ve b katsayıları için, hipotez testi kurularak, kurulan hipotezdeki değerin güven aralığını içinde olup, olmadığı ile bakılarak önemlendirilir.

Hipotez Testi: Basit doğrusal regresyon analizinde parametrelerin sıfıra eşitliği test edilir. İki değişkenli bir regresyon analiz modelinin parametrelerinin test edilebilmesi için düzenlenecek hipotez testleri aşağıdaki gibidir.

α parametresi için:

$$H_0: \alpha = 0$$
$$H_1: \alpha \neq 0 \qquad (3.17)$$

β parametresi için:

$$H_0: \beta = 0$$
$$H_1: \beta \neq 0 \qquad (3.18)$$

dır. α ve β katsayılarına ilişkin hipotez testlerinde kullanılacak test istatistikleri sırasıyla:

$$Z = \frac{\hat{a} - a}{S(\hat{a})} \sim N_z\,(0,1) \qquad (3.19)$$

$$Z = \frac{\hat{b} - b}{S(\hat{b})} \sim N_z(0,1) \qquad (3.20)$$

3.2.2. Medyana Göre Parametre Tahmininde Regresyon Doğrusunun Tahmini

n gözlem çiftinden oluşan örneklem

$(X_1,Y_1),\ldots,(X_n,Y_n)$ için ,

$$Y_i = \hat{a} + \hat{b}X_i \qquad (3.21)$$

3.2.2.1. Mood-Brown Yöntemi

Eşitlik (3.21) ile verilen regresyon doğrusu için a ve b 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. a ve b ' nin istenen değerleri, iki grubun herbirinde regresyon doğrusundan sapmaların medyanının sıfır olduğu tahmindir. a ve b 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 (Daniel,1990).

7. Sonuçta elde edilen doğrunun Y ile kesişimi a katsayısını verirken, b katsayısı eşitlik (3.23) ile hesaplanır.

$$b_{ij} = (Y_2 - Y_1) / (X_2 - X_1) \tag{3.22}$$

(X_1,Y_1) ve (X_2,Y_2) doğru üzerindeki herhangi iki noktanın koordinatlarıdır.

Medyana Göre Parametre Tahmininde α ve β‘ ya İlişkin Hipotezlerin Testleri: Kurulan regresyon eşitliği için ilgilendiği a ve β katsayılarının birisine ya da her ikisine ilişkin hipotezlerin testidir. $a = a_0$ ve $\beta = \beta_0$ hipotezlerinin eşanlı testleri ele alınmıştır

Mood ve Brown tarafından önerilen bu testlerin işleyişi şöyledir (Daniel,1990).

Varsayımlar:

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

Hipotezler:

$$H_0 : a = a_0 , b = b_0$$

$$H_1 : a \neq a_0 \text{ ve/veya } b \neq b_0 \tag{3.23}$$

Test Ölçütü:

Test ölçütünü hesaplarken şu adımlar izlenmektedir:

1. Verilerin serpilme diyagramı hazırlanır.
2. Serpilme diyagramında $Y = a_0 + b_0 X$ doğrusu çizilir.
3. X değerlerinin medyanından geçecek şekilde dikey bir doğru çizilir.
4. n_1, belirtilen regresyon doğrusunun üstündeki ve X’ lerin medyanından 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 ölçütünü eşitlik (3.25) daki gibi izah etmiştir. Tate ve Clelland, (1957) n değerinin yaklaşık olarak 10 veya daha büyük olması halinde bu yaklaşımın iyi sonuçlar verdiğini söylemişlerdir.

$$c^2 = \frac{8}{n}((n_1 - \frac{n}{4})^2 + (n_2 - \frac{n}{4})^2) \tag{3.24}$$

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

Karar Kuralı: Eğer (3.25) eşitliği ile bulunan hesap değeri, ilgili c^2 tablosu' dan iki serbestlik derecesi için bulunan tablo değerini aşarsa H_0 hipotezi red edilir.

$\beta = \beta_0$ Testi: Regresyon analizinde genellikle regresyon doğrusunun eğimi için $H_0 : b = b_0$ testi ile ilgilenilir. Bu hipotezi $H_1 : b \neq b_0$ 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. a, Y_i-β_0X sapma değerlerinin medyanı ve β_0 hipotezde ileri sürülen değer olmak üzere, verilere Y= a + β_0X doğrusu uydurulur. Genellikle bu doğru en kolay şekilde Y = β_0X doğrusunu çizerek ve bu doğruya paralel bir doğruyu verileri iki eşit gruba bölecek şekilde çizerek belirlenir.
4. Y = a + β_0X doğrusunun üzerindeki ve X' lerin medyanının solundaki veri noktaları sayılır ve n_1 ile gösterilir. Buna göre test ölçütü eşitlik (3.26) deki gibi izah edilir(Mood, 1950 ve Mood-Brown, 1951).

$$c_{bi}^2 = \frac{16}{n}(n_1 - \frac{n}{4})^2 \tag{3.25}$$

bulunan bu değer tablo'dan bir serbestlik dereceli ki- kare tablo değeri ile karşılaştırılır. Tablo değerini aşması durumunda H_0 hipotezi red edilir. Tate ve Clelland, (1957) n değerinin 20 veya daha büyük olması halinde bu yaklaşımın iyi sonuçlar verdiğini söylemişlerdir.

3.2.2.2. Theil Yöntemi

1950 yılında Theil tarafından ileri sürülen yöntem araştırmacıların en çok başvurdukları eğim bulma yöntemlerindendir. 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 ve Sprent, 1983). Sahip olunan

$(x_1, y_1),...,(x_n, y_n)$

n tane gözlem çiftinin;

$$Y_i = a + bX_i + e_i \;, \qquad i=1,2,\dots,n \qquad (3.26)$$

Burada a ve b_1 bilinmeyen regresyon parametreleridir. X_i değerleri birbirlerinden farklı bilinen sabitler olup $x_1 < x_2 <,...,< x_n$ şeklinde sıralanmaktadır (Yıldız, ve Topal, 2001). Bu modelde e_i'ler, varyansı s_e^2 ve medyanı sıfır olan simetrik bir sürekli dağılışa sahip bağımsız ve aynı dağılımdan meydana gelen şansa bağlı hatalardan oluşurlar (Rao ve Gore, 1982). Theil yönteminde a ve b öyle tahmin edilmeli ki e_i hata terimlerinin medyanı sıfır olmalıdır (Maritz, 1979). b'nın tahmini $\hat{b}$, i<j $(x_i \neq x_j)$ olmak üzere elde edilen N= $\binom{n}{2}$ tane $b_{ij} = (y_j - y_i) / (x_j - x_i)$ eğimlerinin tümünün bir ağırlıklı medyanı olur (Daniel, 1995; Wang ve Yu, 2004). Yani,

$$\hat{b} = \text{medyan}\{b_{ij}\} \qquad (3.27)$$

dir ve $\hat{a}$;

$$\hat{a} = medyan(Y) - (\hat{b})medyan(X_i) \qquad (3.28)$$

değerlerinin medyanı olduğu belirtilmiştir (Hussain ve Sprent, 1983).

Medyana Göre Parametre Tahmininde α ve β‘ ya İlişkin Hipotezlerin Testleri: Kurulan regresyon eşitliği için ilgilenilen a ve β katsayılarının birisine ya da her ikisine ilişkin hipotezlerin testidir. $a = a_0$ ve $\beta = \beta_0$ hipotezlerinin eşanlı testleri ele alınmıştır

Theil tarafından eğimin testi için önerilen bu yöntem Kendall’ ın Tau istatistiğine dayanmaktadır (Daniel,1990).

Varsayımlar:

A. Veriler $(x_1,y_1),\ldots,(x_n,y_n)$ şeklindeki n tane gözlem çiftinden oluşmakta olup uygun denklem:

$$Y_i = a + \beta X_i + e_i, \qquad i = 1,\ldots,n \tag{3.29}$$

dir. Burada X_i’ ler bilinen sabitler, a ve β ise bilinmeyen parametrelerdir.

B. Herbir X_i değeri için Y değerlerinin bir alt populasyonu vardır.

C. Y_i, X_i değerindeki sürekli Y raslantı değişkeninin gözlenen değeridir.

D. Tüm X_i değerleri farklıdır ve $x_1 < x_2 < \ldots < x_n$ sırasındadır.

E. e_i’ ler karşılıklı bağımsız olup aynı sürekli populasyondan gelirler.

Hipotezler:

A. (Çift yönlü) $H_0 : b = b_0$

$$H_1 : b \neq b_0 \tag{3.30}$$

B. (Tek Yönlü) $H_0 : b \leq b_0$

$$H_1 : b > b_0 \tag{3.31}$$

C. (Tek Yönlü) $H_0 : b \geq b_0$

$$H_1 : b < b_0 \tag{3.32}$$

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 sutunda doğal sırada düzenlenmiştir.
2. Her bir Y_i-$\beta_0 X_i$ değeri, altındaki herbir 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 (3.34) eşitliği ile izah edilmiştir.

$$t = \frac{S}{n.(n-1)/2} \qquad (3.33)$$

Hipotezlere Göre Karar Kuralı:

A. Hesaplanan t değeri pozitif olup, Kendall' ın Tau tablosun' dan n ve α/2 için bulunacak değerden büyükse ya da negatif bulunan t değeri aynı tablo değerinden küçük ise H_0 hipotezi red edilir.

B. Hesaplanan t değeri pozitif olup, Kendall' ın Tau tablosun' dan n ve α için bulunacak kritik değerden büyük ise H_0 hipotezi red edilir.

C. Hesaplanan negatif t 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 ½ uyumlu ve ½ uyumsuzdur. Bu durum t katsayısı hesabında bir değişiklik yaratmaz. Ancak tekrar durumunda t ' nin hesabında değişiklik yapar.

Kendall'ın t 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 t katsayısı, (3.24) eşitliği ile hesaplanmıştır.

$$t = \frac{P-Q}{P+Q} \tag{3.34}$$

Bu yaklaşımın avantajı tekrar durumuna rağmen ilişki miktarını 1 veya -1 elde etme şansının olmasıdır. İlk kez Goodman ve Kruskal, (1963). 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) tarafından incelenmiştir (Daniel, 1990).

Kendall'ın Tau tabloları n' in yalnızca 40' dan küçük veya eşit değerleri için yazılmıştır. Örnek genişliği 40' dan fazla olduğunda test istatistiği;

$$Z = \frac{3t\sqrt{n(n-1)}}{\sqrt{2(2n+5)}} \qquad \sim N(0,1) \tag{3.35}$$

eşitliği ile hesaplanabilmektedir (Aytaç, 1991). Bu test istatistiği veri seti içerisinde, eşit gözlem değerleri var olduğu zaman da başarı ile kullanılabilmektedir.

Medyana Göre Parametre Tahmininde Eğim Katsayısı İçin Güven Aralığı:

Çoğu durumlarda 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). Bu teknik, %(1-α) güven aralığının, a yanılma düzeyinde red edilemeyecek β_0 değerlerinden oluştuğu bilgisine dayanmaktadır. Bu yaklaşıma alternatif bir yaklaşım yine Mood tarafından önerilmiştir(Yıldız ve ark., 2004).

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

b için bir güven aralığı çift yönlü hipotez $H_0 : b = b_0$, $H_1 : b \neq b_0$ testiyle elde edilir. Güven aralığı i<j olmak üzere düzenlenmiş b_{ij} eğimler setinde uygun olarak seçilmiş sınırlı noktalara sahiptir (Sievers, 1978).

$x_j - x_i \neq 0 \quad (1 \leq i < j \leq n)$ farklarının sayısı N ise, önerilen nokta tahmincisi $x_i \neq x_j$ için N tane b_{ij} eğiminin medyanı olur (Sen, 1968). i<j olmak üzere toplam $N = \binom{n}{2}$ adet b_{ij} değeri küçükten büyüğe doğru sıralanır. b ' nın güven aralığının alt sınırı $\hat{b}_a$ küçükten büyüğe doğru sıralanan k'ncı b_{ij} değeridir. b ' nın güven aralığının üst sınırı $\hat{b}_ü$ ise büyükten küçüğe doğru sıralanan k'ncı b_{ij} değeridir (Daniel, 1990). Burada k değeri, (3.37) deki eşitlikte izah edildiği gibidir.

$$k = \frac{N - S_{a/2} - 2}{2} \tag{3.36}$$

$S_{a/2}$ değerini belirlemek için, n ve $a/2$ ' ye göre Kendall' ın Tau test istatistik değerleri tablosuna bakılır. Bulunan değer eşitlik (3.37) da yerine konarak k değeri tespit edilir.

Eğer n değeri çok büyük (n≥ 30) ise, dağılış normal dağılışa yaklaşır. Buna göre standart normal dağılım tablosu kullanılarak k değeri eşitlik (3.38) deki gibi hesaplanır (Griffin, 1962).

$$k = \frac{N - Z_{a/2}\sqrt{\dfrac{n(n-1)(2n+5)}{18}}}{2} \tag{3.37}$$

3.2.2.3. k-En Yakın Komşuluk Tahmin Yöntemi

Parametrik olmayan regresyon fonksiyon analiz yöntemlerinden en yakın komşu(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 (i=1,…,n)

$$Y_i = f(X_i) + e_i \tag{3.38}$$

düzgün (smooth) fonksiyonu ile ifade edilmektedir. Burada hata terimi e_i’ lerin bağımsız, sıfır ortalamalı olmaları dışında hiçbir kısıtlayıcı varsayım bulunmamaktadır.

Parametrik olmayan regresyon fonksiyon analizinde 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 fonksiyonunun tahmin yöntemlerinden parametre tahmin hesabı yapan k-NN tahmin yöntemi anlatılmıştır.

Bu yönteme kısaca yakın komşuluk (nearest – neighbor) denilmesinin nedeni, yoğunluk tahmini belli bir noktanın en yakın komşularından çok yakın komşu noktalarına dayanır(Silverman, 1986).

k-NN regresyon fonksiyon tahmini $\hat{m}$ aşağıdaki gibi tarif edilir (Györfi, 1981; Devroye, 1978).

$$\hat{m}_k(x) = n^{-1}\sum_{i=1}^{n} W_{ki}(x) Y_i \quad ; \tag{3.39}$$

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

J_x = {i: X_i x' e en yakın gözlemlerin bir tanesi }

Komşu gözlemlerin indeksleri kullanılarak k-NN ağırlık kümesinin tertip edilişi şöyledir.

$$W_{ki}(x) = \begin{Bmatrix} n/k, & i \in J_x \ \ ise \\ 0, & (i \notin J_x) \end{Bmatrix} \tag{3.40}$$

Eşitlikte,

n: gözlem sayısını

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

(Stone, 1977 ; Devroye, 1978)'e göre, k-NN ağırlık fonksiyonları;

Üniform k-NN ağırlık fonksiyonu;

$$W_{ki}(x) = \begin{Bmatrix} 1/k, & 1 \le i \le k \\ 0, & i > k \end{Bmatrix} \text{ için,} \tag{3.41}$$

Üçgen (triangular) k-NN ağırlık fonksiyonu;

$$W_{ki}(x) = \begin{Bmatrix} (k-i+1)/b_k, & 1 \le i \le k \\ 0, & i > k \end{Bmatrix} \text{ için,} \tag{3.42}$$

burada b_k=k(k+1) / 2 dir.

Quadratik k-NN ağırlık fonksiyonu;

$$W_{ki}(x) = \begin{Bmatrix} (k^2 - (i-1)^2)/b_k, & 1 \le i \le k \\ 0, & i > k \end{Bmatrix} \text{ için,} \tag{3.43}$$

burada b_k=k(k+1).(4k-1) / 6 dır.

$W_{ki}(x)$ üniform ağırlıklar ile tayin edilen neighborhood vasıtasıyla e.k.k. doğrusunun a ve b_i parametreleri aşağıdaki şekilde hesaplanmıştır(Green et al.,1985 ve Hardle, 1997).

$$a_{xi} = \hat{m}_k(x_i) - b_i \overline{m}_{xi}, \tag{3.44}$$

$$b_{xi} = \frac{C_{xi} - \overline{m}_{xi}.m_k(X)}{V_{xi} - \overline{m}_{xi}^2}, \tag{3.45}$$

$$\overline{m}_x = k^{-1} \sum_{i \in j_x} X, \tag{3.46}$$

$$C_x = \sum_{i \in j_x} X_i Y_i, \tag{3.47}$$

$$V_x = \sum_{i \in j_x} X^2, \tag{3.48}$$

(3.48) eşitliğinde;

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

$b_{xi} = W_{ki}(x)$ 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 ve Narendra, 1975) tarafından önerilen ‘’branch ve bound ‘’ algoritmasıdır (Hand, 1986).

Parametrik Basit Doğrusal Regresyonun Varyans Analizi: parametrik basit doğrusal regresyon analiz yöntemlerinin varyans analiz tablosu Çizelge 3.2.’de belirtildiği gibi yapılmıştır.

Çizelge 3.2. Varyans Analiz Tablosu

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

Çizelge 3.3. Varyans Analiz Tablosu (Medyan Testine Göre)

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

Çizelge (3.2) ve Çizelge (3.3)’ deki eşitlikte,
n:gözlem sayısını
k: bağımsız değişken 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 3.3.’de belirtildiği gibi yapılmıştır.

4. BULGULAR VE TARTIŞMA

4.1. En Küçük Kareler Yöntemine Ait Bulgular

Çukurova Üniversitesi, Ziraat Fakültesi, Araştırma Uygulama Çiftliğinin Koyunculuk İşletmesinden, elde edilen veri değerleri kullanılarak, koyunların yaşının, koyunların doğum ağırlığı ile ilişkisi parametrik ve parametrik olmayan basit doğrusal regresyon analiz yöntemleri ile incelenmiştir. Uygulamaları ise; 0.05 önem seviyesinde, Çizelge 3.1.' deki değerler kullanılarak yapılmıştır.

E.k.k. yöntemindeki (3.6) ve (3.7) eşitliği kullanıldığında,

$\hat{a}$ = 4.17 elde edilir.

$\hat{b}$ = 0.082 elde edilir.

Parametrik basit doğrusal regresyon modelinin tahmini,

$$\hat{Y}_i = 4.17 + 0.082 X_i$$

dır. Buna göre;

Yeni doğmuş bir kuzunun doğum ağırlığının 4.17 kg olması beklenir. Yine benzer şekilde koyunun yaşı 1 yaş arttığı zaman doğum ağırlığının da 0.082 kg artması beklenir.

4.2. Parametrik Basit Doğrusal Regresyon Analizinin Belirtme Katsayısının Hesaplanması İle İlgili Bulgular

Parametrik basit doğrusal regresyon analizindeki gözlem değerlerinin modele uyumu, belirtme katsayısı için eşitlik (3.8) kullanıldığında,

$R^2 = 0.006$ elde edilir.

4.3. Parametrik Basit Doğrusal Regresyon Analizinin Güven Aralığının Bulunması ve Hipotez Testinin Hesaplanması İle İlgili Bulgular

E.k.k. yöntemindeki eşitlik (3.10) kullanıldığında,

$Var(\hat{a}) = 0.116$ elde edilir.

Eşitlik (3.11) kullanıldığında ise,

$S(\hat{a}) = 0.341$ elde edilir.

Eşitlik (3.12) kullanıldığında

$Var(\hat{b}) = 0.019$ elde edilir.

Eşitlik (3.13) kullanıldığında ise,

$S(\hat{b}) = 0.138$ elde edilir.

Eşitlik (3.14) kullanıldığında

$S^2 = 0.807$ elde edilir.

$\hat{a}$ için güven aralığı:

$$4.17-(1.96)(0.341) \leq a \leq 4.17+(1.96)(0.341)$$
$$3.503 \leq a \leq 4.836$$

olarak bulunur. Buna göre % 95 olasılıkla yeni doğan bir kuzunun doğum ağırlığının 3.503 ile 4.836 arasındaolması beklenir.

$\hat{b}$ için güven aralığı:

$$0.082- (1.96)(0.138) \leq b \leq 0.082+ (1.96)(0.138)$$

$$-0.189 \leq b \leq 0.353$$

olarak bulunur. Buna göre % 95 olasılıkla kuzunun yaşı 1 yaş arttığı zaman doğum ağırlığında -0.189 kg ile 0.353 kg arasında artması beklenir.

$\hat{a}$ için hipotez testi:

$$Z_{hes} = \frac{4.17 - 0}{0.341}$$

$$= 12.22$$

Z_{hes}= 12.22 > Z_{tab}= 1.96 olduğundan H_0 hipotezi red edilir. Buna göre tahmin edilen 4.17 değeri anlamlı bulunmuştur.

$\hat{b}$ için hipotez testi:

$$Z_{hes} = \frac{0.082 - 0}{0.138}$$

$$= 0.59$$

Z_{hes}= 0.59 < Z_{tab}= 1.96 olduğundan H_0 hipotezi kabul edilir. Buna göre yaştaki değişmelerin doğum ağırlığı üzerine bir etkisi yoktur.

4.4. En Küçük Kareler Varyans Analizi

Çizelge 4.1. Yaş (X) ve Doğum Ağırlığı (Y) Değişkenleri İçin; E.K.K. Varyans Analiz Tablosu

Varyans Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Kareler Ortalaması	F
Regresyon	1	0.284	0.284	0.284/0.807=0.35
Hata	55	44.383	0.807	
Genel	56	44.668		

4.5. Parametrik Olmayan Mood-Brown Yöntemi İle İlgili Bulgular

X_1 değişkeninin medyan değeri 2.16 olup, gözlemlerin birinde 29, birinde de 28 gözlem bulunur. Medyanın altındaki gözlemler 29 tane, medyanın yukarısındaki gözlemler ise, 28 tanedir.

Çizelge 4.1' e bakıldığı zamantahmin edilen regresyon modeli istatistiki olarak önemsizdir.

X_1 ve Y nin medyan değerleri, medyanın altında X=1.6 ve Y=4 iken, medyanın yukarısında X=3.135 ve Y= 4.525 dir. b_1 için ilk yaklaşım:

$$b_1 = \frac{(4.525-4)}{(3.135-1.6)} = \frac{(0.525)}{(1.535)} = 0.342$$

dir. Mood-Brown regresyon doğrusuna ilk yaklaşım;

$$\hat{Y}_i = 3.8 + 0.342X$$

dır. Gözlemlerin medyanları her iki grup için de sıfır olmadığından doğrunun görsel bir ayarlaması yapılır.

$$b_1^{'} = \frac{(4.6-3.9)}{(3.2-2.2)} = \frac{(0.7)}{(1)} = 0.7$$

Düzeltilmiş $b_1^{'}$=0.7 ve a=2.98 dir. Düzeltilmiş Mood-Brown regresyon doğrusu ise;

$$\hat{Y}_i = 2.98 + 0.7X$$

dir. Buna göre;

Yeni doğmuş bir kuzunun doğum ağırlığının 2.98 kg olması beklenir. Yine benzer şekilde koyunun yaşı 1 yaş arttığı zaman doğum ağırlığının da 0.7 kg artması beklenir.

4.6. Parametrik Olmayan Theil Yöntemi İle İlgili Bulgular

Yaş ve doğum ağırlığı arasındaki ilişkiyi tanımlamak için, b eğim katsayısı bulunmalıdır. Theil yöntemi kullanıldığında $\binom{57}{2} = 1596$ tane b_1 değeri bulunur. Bu sıralı b_1 lerin medyan değeri 0.0897 olup, b eğim katsayısı tahmini 0.0897 dir. Buradan medyan X_1=2.16 ve medyan Y= 4.5 olduğundan $\hat{a}$:

$$\hat{a} = 4.5\text{-}(0.089)(2.16)$$
$$= 4.30 \text{ olarak bulunur.}$$

Theil için tahmini basit regresyon eşitliği:

$\hat{Y}_i = 4.30 + 0.089 X_i$ olarak bulunur.

Buna göre;

Yeni doğmuş bir kuzunun doğum ağırlığının 4.30 kg olması beklenir. Yine benzer şekilde koyunun yaşı 1 yaş arttığı zaman doğum ağırlığının da 0.089 kg artması beklenir.

4.7. Parametrik Olmayan Mood-Brown Hipotez Testi İle İlgili Bulgular

Yaş ve doğum ağırlığı arasında kurulacak regresyon modeli için a ve β katsayılarının her ikisine ilişkin hipotezlerin testi yapılmıştır.

$$H_0 : a = 2.98, b = 0.7$$
$$H_0 : a = 2.98, b \neq 0.7$$

şeklindeki hipotezde, $n_1=14$ ve $n_2=17$ bulunmuştur. Test ölçütü ise:

$$c^2 = \frac{8}{57}\left[\left(14 - \frac{57}{4}\right)^2 + \left(17 - \frac{57}{4}\right)^2\right]$$

= (0.140)[(0.06)+(7.56)]

= 1.067 olarak bulunur..

Karar: $c^2 = 1.067$ değeri, tablo’ dan bulunan $c^2_{(2;0,05)} = 5.99$ değerinden küçük olduğundan H_0 hipotezi %5 önem düzeyinde kabul edilmiştir. Tahmin edilen edilen regresyon modelinin istatistiki olarak önemsiz olduğu söylenebilir.

4.8. Parametrik Olmayan $b_i = b_0$ Testi İle İlgili Bulgular

$$H_0 : b = 0$$
$$H_1 : b > 0$$

şeklindeki hipotezde, n_1=3 bulunmuştur. Test ölçütü ise:

$$c^2_{b1} = \frac{16}{57}\left(3 - \frac{57}{4}\right)^2 = 35.437 \text{ olarak bulunur.}$$

Karar: Bu değer, tablodaki kritik değer olan $c^2_{(0,05;1)} = 3.841$ değerini aştığı için H_0 hipotezi %5 önem düzeyinde red edilmiştir. $\hat{b}$ parametresinin anlamlı olduğu söylenebilmektedir şeklinde yorumlamak mümkündür.

4.9. Parametrik Olmayan Theil Yöntemi Hipotez Testi İle İlgili Bulgular

Yaş ve doğum ağırlığı ölçümleri verilmiştir. Bu iki ölçüm arasında kurulan regresyon modeli için $b_0 = 0$ hipotezini test edelim.

$$H_0 : b = 0$$
$$H_1 : b \neq 0$$

Test Ölçütü: $b_0 = 0$ olduğu için $Y_i - b_0 X_i = Y_i$ olur. (X_i, Y_i) gözlem çifti için Kendall'ın Tau ölçütü hesaplanmıştır. Bunun için gerekli işlemler aşağıda verilmiştir. X değişkeni için, P ve Q değerleri:

S.No:	P	Q
1	13	14
2	8	22
3	4	22
4	23	2
5	22,5	3,5
6	16,5	12,5
.	.	.
.	.	.

.	.	.
54	2	1
55	2	0
56	1	0
57	0	0

$$\sum_{i=1}^{57} P = 434 \qquad \sum_{i=1}^{57} Q = 479$$

Bu tablodan:

$$t = \frac{P-Q}{P+Q} = \frac{434-476}{434+479} = \frac{-45}{913} = -0.05 \text{ olarak bulunur.}$$

n = 57 olduğundan,

$$Z = \frac{3t\sqrt{n(n-1)}}{\sqrt{2(2n+5)}}$$

$$= \frac{3(-0.05)\sqrt{57(57-1)}}{\sqrt{2(2*57+5)}}$$

= -0.55 olur ve $Z_{a/2} = Z_{0.025} = 1.96$ olarak bulunur.

Karar: Z= -0.55 değeri tablo değerinin negatifinden daha büyük olduğu için H_0 hipotezi %5 önem düzeyinde kabul edilmiştir. Değişkenlerdeki veri değerlerine göre, koyunların yaşı ve doğum ağırlığı arasında anlamlı bir ilişki yoktur. Koyunlar büyük yaşta olursa, doğum ağırlığı azalmaktadır. Koyunlar küçük yaşta olursa, doğum ağırlığı ise artmaktadır şeklinde yorumlamak mümkündür.

4.10. Parametrik Olmayan Basit Doğrusal Regresyon Analizinde Eğim Katsayısı İçin Güven Aralığı İle İlgili Bulgular

X verileri için b ' nın %95 güven aralığı:

$$k = \frac{1596 - (1.96)(145.267)}{2} = 655.6 \cong 656 \text{ dır.}$$

Sıralı b_{ij} değerlerinin alttan 656 ncı değeri $b_a = -0.202$ ve üstten 656 ncı değeri $b_ü = 0.433$ tür. b_1 için %95 güven aralığı:

$(-0.202 < b_1 < 0.433)$

dır. Buna göre % 95 olasılıkla b parametresinin -0.202 ile 0.433 arasında bulunabileceği söylenebilir.

4.11. Medyana Göre Parametre Tahminin Mood-Brown Varyans Analizi

Çizelge 4.2. Yaş (X) ve Doğum Ağırlığı (Y) Değişkenleri İçin; Mood-Brown Varyans Analiz Tablosu

Varyans Kaynağı	**Serbestlik Derecesi**	**Kareler Toplamı**	**Kareler Ortalaması**	**F**
Regresyon	1	21.069	21.069	21.069/0.448=47.029*
Hata	55	24.653	0.448	
Genel	56	45.722		

*P<0,05

Buna göre regresyon modeli (tahmin modeli) istatistiki olarak önemli bulunmuştur.

Yaş ve doğum ağırlığı verileri kullanılarak elde edilen analiz sonuçları çizelge (4.2)' deki tablo sonucu göstermiştir ki koyunların büyük veya küçük yaşta olması ile doğum ağırlığı arasında doğrusal bir ilişki bulunmaktadır.

4.12. Medyana Göre Parametre Tahmininin Theil Varyans Analizi

Çizelge 4.3. Yaş (X) ve Doğum Ağırlığı (Y) Değişkenleri İçin; Theil Varyans Analiz Tablosu

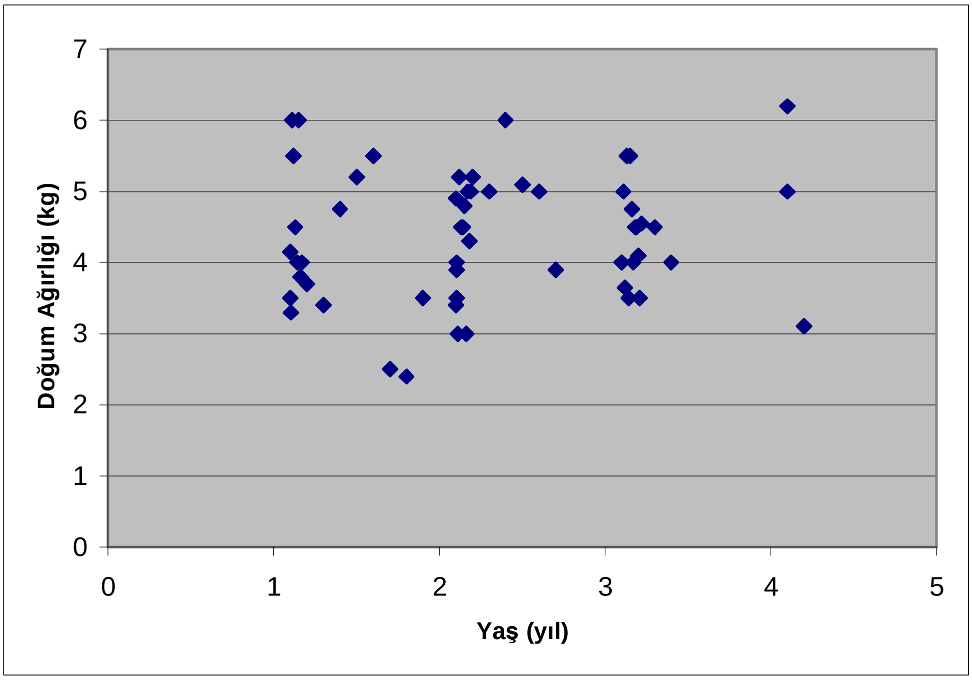
Varyans Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Kareler Ortalaması	F
Regresyon	1	0.334	0.334	0.334/0.825=0.404^{-}
Hata	55	45.388	0.825	
Genel	56	45.722		

H_0 hipotezi %5 önem düzeyinde kabul edilmiştir(P>0.05). Theil yöntemine dayalı olarak bu değişkenlere ait veri değerlerinin kullanılmasıyla hesaplamış olduğumuz regresyon doğrusu, bağımlı ve bağımsız değişkenler arasında doğrusal bir ilişkiyi göstermemektedir.

Yaş ve doğum ağırlığı verileri kullanılarak elde edilen analiz sonuçları çizelge (4.3)' deki tablo sonucu göstermiştir ki koyunların büyük veya küçük yaşta olması ile doğum ağırlığı arasında doğrusal bir ilişki bulunmamaktadır şeklinde yorumlamak mümkündür.

4.13. Parametrik Olmayan k-En Yakın Komşuluk Tahmin Yöntemi (k-NN) İle İlgili Bulgular

k-NN tahmin yöntemi için, yaş ve doğum ağırlığı değişkenlerinin regresyon parametreleri tahmin edilmiştir.



Şekil 4.1. X_1 ve Y Değişkenleri İçin Serpilme Diyagramı

n=57 tane veri ve k=20, x=2.16 için $\hat{m}_k(x)$:

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

$J_x = J_{2,16} =$ {(1.1),(1.13),(1.14),(1.16),…,(3.22),(3.3),(3.4)}

$$W_{ki}(x) = \begin{cases} n/k, & i \in j_x; \\ 0, & i \notin j_x; \end{cases}$$

buna göre W_{ki} (x)' ler;

W_{k1}(2.16)=57/20, W_{k2}(2.16)= 0, W_{k3} (2.16)= 0, W_{k4} (2.16)=0,…, W_{k55} (2.16)= 0, W_{k56} (2.16)= 0, W_{k57} (2.16)= 0 dır.

k-NN tahmin ise;

$\hat{m}_{20}^{(2,16)} = 1/57\{(57/20*4.15)+(0*3.5)+(0*3.3)+(0*6)+\ldots +(0*6.2)+(0*5) +(0*3.1)\}$

$\hat{m}_{20}^{(2,16)} = 1/57[11.82+8.01+7.12+6.76+\ldots+8.10+8.01+7.12]$

= 237.5/57

= 4.16 dır.

b_1 katsayısı ise;

$$b_{x1} = \frac{C_{x1} - \overline{m}_{x1}.\hat{m}_k(X_1)}{V_{x1} - \overline{m}^2{}_{x1}}$$ formülünden

$$\overline{m}_x = k^{-1}\sum_{i\in j_x} X_1$$

= 1/20 *(46.416)=2.3

$$C_x = \sum_{i\in j_x} X_i Y_i$$

= 195.306

$$V_x = \sum_{i\in j_x} X_i^2$$

= 123.272

$$b_{x1} = \frac{195.306 - (2.3)(4.16)}{123.272 - (2.3)^2}$$

$$= \frac{185.738}{120.972}$$

= 1.57 dur.

a katsayısı ise:

$$a_x = \hat{m}_k(X) - b_1\overline{m}_{x1}$$
$$= 4{,}16\text{-}(1.57)*(2.3)$$
$$= 4.16\text{-}(3.61)$$
$$= 0.54$$

dır. k-NN yönteminin basit doğrusal regresyon eşitliği:

$$\hat{Y}_i = 0.54 + 1.57X_i$$

dir. Buna göre;

Yeni doğmuş bir kuzunun doğum ağırlığının 0.54 kg olması beklenir. Yine benzer şekilde koyunun yaşı 1 yaş arttığı zaman doğum ağırlığının da 1.57 kg artması beklenir.

4.14. Regresyon Fonksiyonunun k-Nearest Neighbor Varyans Analizi

Çizelge 4.4. Yaş (X), ve Doğum Ağırlığı (Y) Değişkenleri İçin; k-NN Varyans Analiz Tablosu

Varyans Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Kareler Ortalaması	F
Regresyon	1	39	39	39/0.101=386.13
Hata	55	5.6	0.101	
Genel	56	44.668		

Bu yöntemde regresyon tahmin modeli istatistiki olarak önemli bulunmuştur. Bu durum çizelge 4.4' de görülmektedir.

Yaş ve doğum ağırlığı verileri kullanılarak elde edilen analiz sonuçları çizelge (4.4)' deki tablo sonucu göstermiştir ki koyunların büyük veya küçük yaşta olması ile doğum ağırlığı arasında doğrusal bir ilişki bulunmaktadır şeklinde yorumlamak mümkündür.

k-NN ve Mood-Brown yöntemlerine göre elde edilen regresyon doğru eşitliğinin gerçek verilere uyumunun diğer yöntemlerden daha iyi olmasının sebebi; k-NN yönteminde, belirlenen herhangi bir X_i değerinden uzakta olan bazı gözlemlerin işleme tabi tutulmaması ile elde edilen regresyon doğrusundan gerçek verilerin sapmalarının küçük olmasından kaynaklanmıştır. Mood-Brown yönteminde ise, (X_i, Y_i) nokta değerlerinin regresyon doğrusuna dikey sapmalarının medyanı sıfır olana kadar regresyon doğrusunun konumunun değiştirilmesidir.

Parametrik olmayan yöntemler uygulamada oluşabilecek bazı olumsuzlukları da giderme imkanı sağlar. Parametrik olmayan yöntemler uygulanırken örnek genişliğinin de önemi azalmaktadır. Özellikle küçük örnek genişliğinden oluşan uygulamalarda araştırmacı, parametre ve değişkenler hakkında bilgi sahibi olmazsa bile araştırmaya devam etme imkanı elde etmektedir. Parametrik doğrusal modeller ile karşılaştırılabilir diğer önemli bir konu da varsayımlardan hareket etme şartıdır. Parametrik doğrusal yöntemlerde var olan varsayımlardan hareket etme şartında; parametrik olmayan yöntemler uygulanırken, bazı durumlarda yerine gelmemekte ve araştırmacı, varsayımlar yerine gelmezse bile araştırmasını sürdürmektedir.

Parametrik olmayan doğrusal regresyon yöntemlerin küçük örnekler için daha iyi sonuçlar verdiği ifade edilmesine karşın, n>100 gözlem değerleri için de kullanıldığı durumlar vardır. Örnek genişliği n>100 iken modelin kendine ait değişkenler normal dağılmazsa bile örnek ortalamaları normal dağılımı izlemektedir. Bundan dolayı büyük örnekler için parametrik yöntemler daha çok duyarlı ve daha büyük istatistik güce sahiptir (Şengün,1999).

5. SONUÇLAR VE ÖNERİLER

5.1. $\hat{b}$, R^2, H.K.O., U_1, U_2, Değerlerine Göre Parametrik ve Parametrik Olmayan Basit Doğrusal Regresyon Analiz Yöntemlerinin Karşılaştırmalı Olarak İncelenmesi

Theil, Mood-Brown, E.K.K., k-NN tahmin yöntemlerine göre elde edilen regresyon parametre değerleri ve regresyon doğrusunun eğimi olan b ' nın önemlilik durumu Çizelge (5.1.)' de özet olarak gösterilmiştir.

Çizelge 5.1. Çeşitli Yöntemlere Göre Elde Edilen Regresyon Parametre Değerleri Tablosu

Metodlar	$\hat{a}$	$\hat{b}$
En Küçük Kareler	4.17	0.082^{-}
Mood-Brown	2.98	0.7^{*}
Theil	4.30	0.089^{-}
k-Nearest Neighbor	0.54	1.57^{*}

$*$: b_i katsayısı ($p<0.05$) de önemlidir.

$-$: b_i katsayısı ($p>0.05$) de önemsizdir..

Theil' in U_1 ve U_2 katsayıları:

$$U_1 = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}}{\sqrt{\sum_{i=1}^{n}\hat{Y}_i^2} + \sqrt{\sum_{i=1}^{n}Y_i^2}}$$

ve

$$U_2 = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}Y_i^2}}$$

formülüne göre hesaplanmıştır(Kovalyova ve Cox, 2004). Yukardaki eşitliklerde

Y_i :gözlenen değerler

$\hat{Y}_i$: beklenen değerler

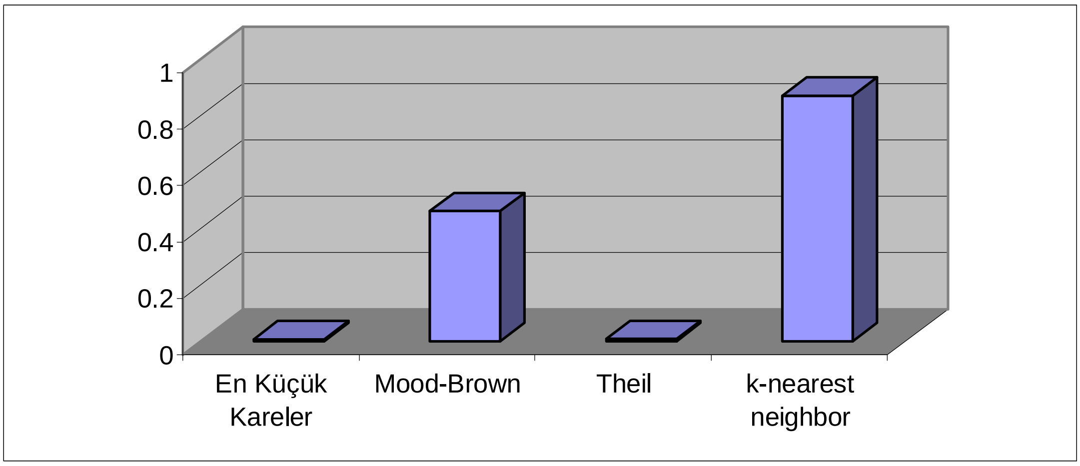
n : örnek büyüklüğü

U_1 katsayısı 0 ve 1 ($0<U_1<1$) arasında ve U_2 katsayısı 0 ve $+\infty$ ($0<U_2<+\infty$) arasında değişmektedir. Her iki katsayının da sıfıra yakın değerleri iyi uyumu gösterirken, büyük değerleri kötü uyumu göstermektedir.

Çizelge 5.2. Çeşitli Yöntemlere Göre Elde Edilen Belirtme Katsayısı (R^2), H.K.O., ve Theil' in U_1 ve U_2 Katsayı Değerleri Tablosu

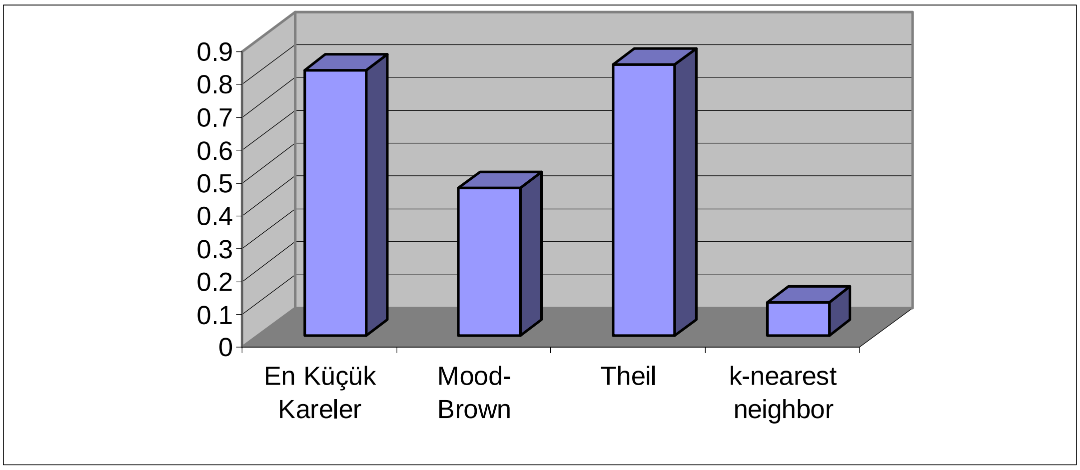
Metodlar	R^2	**H.K.O.**	U_1	U_2
En Küçük Kareler	0.006	0.807	0.013	0.198
Mood-Brown	0.460	0.448	0.009	0.147
Theil	0.007	0.825	0.013	0.200
k-nearest neighbor	0.87	0.101	0.004	0.07

R^2 değerlerine göre yöntemleri karşılaştırmada, yöntemlere göre elde edilen regresyon doğrularının gerçek verilere ne derece uyumlu olduğu tespit edilmeye çalışılır. Karşılaştırma yapılırken R^2 değeri büyük çıkan yönteme göre elde edilen modelin gerçek verilere daha iyi uyum sağladığı söylenebilir. Her bir yönteme göre elde edilen modelin gerçek verilere uyum durumuna göre yöntemler arasındaki karşılaştırma şekil (5.1)' deki gibi elde edilmiştir.



Şekil 5.1. Yöntemlerin Belirtme Katsayısına Göre Durumları

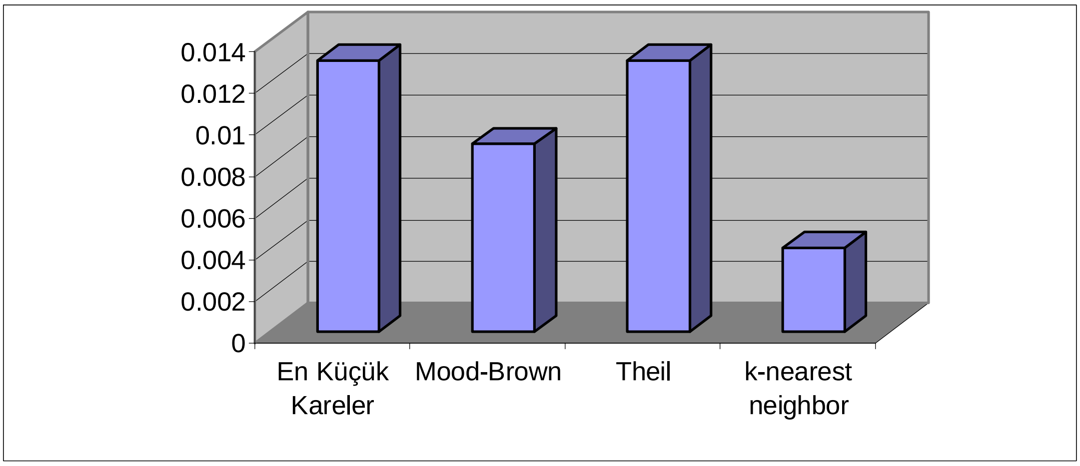
Yöntemler arasındaki karşılaştırma H.K.O. göre yapıldığında: H.K.O. en küçük olan modelin en uygun olduğu ifade edilir. Çünkü her bir yönteme göre elde edilen modelde H.K.O. en küçük modelde gözlenen değerler ile beklenen değerler arasındaki sapma da o derece küçük olur. Gözlenen değerler ile beklenen değerler arasındaki sapma küçüldükçe o yönteme göre uydurulan modelin gerçek verilere uyumluluğu da daha iyi olur. H.K.O.' na göre modellerin gerçek verilere uyumluluk durumuna göre yöntemler arasındaki karşılaştırma şekil (5.2)' deki gibi elde edilmiştir.



Şekil 5.2. Yöntemlerin Hata Kareler Ortalamasına Göre Durumları

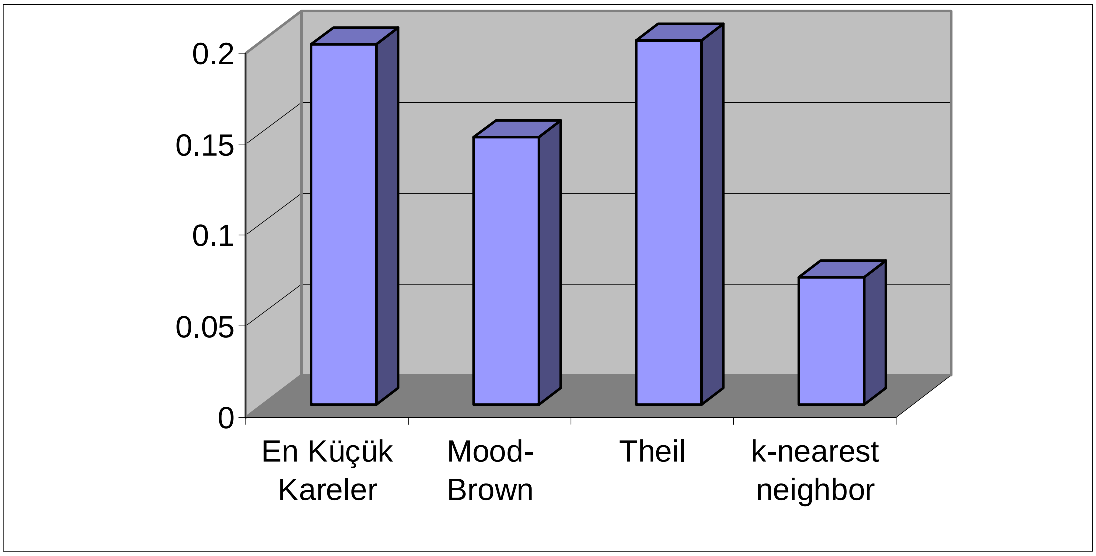
Yöntemler arasındaki karşılaştırma Theil'in U_1 ve U_2 katsayı değerlerine göre yapıldığında, yöntemlere göre elde edilen regresyon doğrularının gerçek verilere ne derece uyumlu olduğu tespit edilmeye çalışılmaktadır. Karşılaştırma yapılırken her iki katsayının da sıfıra yakın değerleri iyi uyumu gösterirken, büyük değerleri kötü uyumu göstermektedir. Herbir yönteme göre elde edilen modelin gerçek verilere uyum durumuna göre yöntemler arasındaki karşılaştırma şekil (5.3) ve şekil (5.4)'deki gibi elde edilmiştir.

Theil'in U_1 katsayısına göre:



Şekil 5.3. Yöntemlerin Theil' in U_1 Katsayısına Göre Durumları

Theil'in U_2 katsayısına göre:



Şekil 5.4. Yöntemlerin Theil' in U_2 Katsayısına Göre Durumları

Bağımlı ve bağımsız değişkenler arasındaki ilişkinin incelenmesi için e.k.k. yönteminde kullanılan çizelge (3.1)' deki veri değerleri parametrik basit doğrusal regresyon için uygun değildir. Çünkü kullanılan e.k.k. yöntemi, koyunların yaşı ile doğum ağırlığı arasında doğrusal bir ilişki oluşturmamıştır. Koyunların doğum ağırlığındaki artma veya azalma yaş'a bağımlı değildir. Koyunların doğum ağırlığındaki artma veya azalmanın, koyunların yaşından farklı özelliklerle ilişkili olabileceğini göstermiştir. Doğum ağırlığı fazla olan koyunların veya doğum ağırlığı az olan koyunların yaşlarıyla ilişkili olabileceği düşünülmemelidir. Doğum ağırlığı fazla olan koyunun yaşı küçük veya doğum ağırlığı az olan koyunun yaşı büyük olabilmektedir.

Bağımlı ve bağımsız değişkenler arasındaki ilişkinin incelenmesi için Mood-Brown yönteminde kullanılan çizelge (3.1)' deki veri değerleri parametrik olmayan basit doğrusal regresyon için uygundur. Çünkü kullanılan Mood-Brown yöntemi, koyunların yaşı ile doğum ağırlığı arasında doğrusal bir ilişki oluşturmuştur. Koyunların doğum ağırlığındaki artma veya azalma yaş'a bağımlıdır. Doğum ağırlığı fazla olan koyunların veya doğum ağırlığı az olan koyunların yaşlarıyla ilişkili olabileceği düşünülmelidir. Doğum ağırlığı fazla olan koyunun yaşı büyük veya doğum ağırlığı az olan koyunun yaşı küçük olabilmektedir.

Bağımlı ve bağımsız değişkenler arasındaki ilişkinin incelenmesi için Theil yönteminde kullanılan çizelge (3.1)' deki veri değerleri parametrik olmayan basit doğrusal regresyon için uygun değildir. Çünkü kullanılan Theil yöntemi, koyunların yaşı ile doğum ağırlığı arasında doğrusal bir ilişki oluşturmamıştır. Koyunların doğum ağırlığındaki artma veya azalma yaş'a bağımlı değildir. Koyunların doğum ağırlığındaki artma veya azalmanın, koyunların yaşından farklı özelliklerle ilişkili olabileceğini göstermiştir. Doğum ağırlığı fazla olan koyunların veya doğum ağırlığı az olan koyunların yaşlarıyla ilişkili olabileceği düşünülmemelidir. Doğum ağırlığı fazla olan koyunun yaşı küçük veya doğum ağırlığı az olan koyunun yaşı büyük olabilmektedir.

Bağımlı ve bağımsız değişkenler arasındaki ilişkinin incelenmesi için k-NN yönteminde kullanılan çizelge (3.1)' deki veri değerleri parametrik olmayan basit doğrusal regresyon için uygundur. Çünkü kullanılan k-NN yöntemi, koyunların yaşı ile doğum ağırlığı arasında doğrusal bir ilişki oluşturmuştur. Koyunların doğum ağırlığındaki artma veya azalma yaş'a bağımlıdır. Doğum ağırlığı fazla olan koyunların veya doğum ağırlığı az olan koyunların yaşlarıyla ilişkili olabileceği düşünülmelidir. Doğum ağırlığı fazla olan koyunun yaşı büyük veya doğum ağırlığı az olan koyunun yaşı küçük olabilmektedir.

Parametrik olmayan regresyon parametrelerini tahminde medyana göre tahmin yöntemlerinin, regresyon fonksiyonunun tahmin yönteminden daha kolay anlaşıldığı ve uygulamasının daha kolay olduğu gözlenmiştir. Parametrik olmayan doğrusal yöntemlerin, parametrik doğrusal tahmin yöntemlerine nazaran daha az sınırlayıcı olması ile bu yöntemlerde örnek populasyonun sadece devamlı tanımlanmış bir dağılımdan gelme varsayımı, parametrik olmayan yöntemlerin daha çok kullanılmasına ve yayılmasına neden olmuştur. Bu durumda, etkili olan diğer bir konu da parametrik olmayan yöntemlerin, parametrik yöntemlerin geçerli olduğu durumda kullanılmasıdır.

Parametrik olmayan doğrusal regresyon yöntemleri, parametrik doğrusal regresyon yöntemleri ne zaman geçerli olursa uygulanmaktadır. Bunun için herhangi bir sınırlama yoktur.

KAYNAKLAR

AKDENİZ, F., 2001. The Examination and Analysis of Residuals for Some Biased Estimators in Linear Regression. Communications in Statistics Theory and Methods, 30(6), 1171-1183.

AYTAÇ M., 1984. Uygulamalı Parametrik Olmayan İstatistik Testleri. Uludağ Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Bursa, 44-47.

—————, 1991. Uygulamalı Parametrik Olmayan İstatistik Testleri,Uludağ Üniversitesi Basımevi,Bursa.

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

BROWN, G.W. AND A.M. MOOD, 1951. On Median Tests for Linear Hypotheses. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, The University of California Press. p:159-166.

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

¾ ¾ ¾ ¾ , 1995. Biostatistics: A Foundation for Analysis in The Health Sciences. John Wiley & Sons, Inc Canada, 6nd edition. p:622.

DEVROYE, L.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.

ERGÜNEŞ. E., 2004, En Küçük Kareler Yöntemi İle Ridge Regresyon Yönteminin Karşılaştırılmalı Olarak İncelenmesi, Ç.Ü.F.B.E Zootekni Anabilim Dalı, YLT., Adana.

FUKUNAGA K. AND NARENDRA M., 1975. A Branch and Bound Algorithm for Computing k-Nearest Neighbors. IEEE Transactions On Computing, 24, 750-753.

GAMGAM, H. (1989), Parametrik Olmayan İstatistiksel Teknikler, Ankara. 242s.

GOODMAN, L.A., AND KRUSKAL, W.H., 1963. ‘’Measures of Association for Cross Classifications. III: Approximate Sample Theory’’. Journal of the American Statistical Association, 58, 310-364.

GREEN, P., JENNISON, C. AND SEHEULT, A., 1985. Analysis of Field Experiments by Least Squares Smoothing. Journal of the Royal Statistical Society, Series B, 47 (2), 299-315.

GRIFFIN, J. I., 1962. Statistics Methods and Application. Rinehart and Winston. 259-273.

GYÖRFI, L., 1981. The Rate of Convergence of k_n-NN Regression Estimates and Classification Rules. IEEE Transactions of Information Theory, 27, 500-509.

HAND, D.J., 1986. Discrimination and Classification, Wiley & Sons, New York.

HARDLE, W., 1997. Applied Nonparametric Regression. Cambridge University, UK, 6^{nd} edition. P: 14-85.

HUSSAIN, S.S., AND SPRENT, P., 1983. Non Parametric Regression. J.Roy Statist. Soc., Series A, 146, 182-191.

IŞIK, A. (2006), İstatistik II, Yazın Matbaacılık, Ankara. 194s.

İMİR, E., 1986. Çoklu Doğrusal Modellerde Ridge Regresyon Modeli ile Parametre Kestirimi, Anadolu Üniv. Yayınları No: 212, 76s. , Eskişehir.

KIROĞLU G., 2001. Uygulamalı Parametrik Olmayan İstatistiksel Yöntemler. Mimar Sinan Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, İstanbul.

KOVALYOVA, A.E. AND COX, A.M., 2004. Missouri Show Me Model and its Improvement:What Makes it Different From Other States Models. www.rupri.org/ get File asp? filelD=13107.

LEVINE, D.M., M.L. BERENSON, AND D. STEPHAN, 1997. Statistics for Managers Using Microsoft Excel. Prentice-Hall Press.

LESAFFRE, E., and MARX, B.D., 1993, Collinearity in Generalized Linear Models, Communications in Statistics Theory and Methods, 18(9), 3463-3472.

MACKINNON, M.J., and PUTERMAN, M.L., 1989, Collinetary in Generalized Linear Regression, Communications in Statistics Theory and Methods, 22(7), 1933-1952.

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

MONTGOMERY D.C., E.A. PECK, 1992. Introduction to Linear Regression Analysis. John Willey and Sons, Inc. Canada.

MOOD, A.M., 1950. Introduction to the Theory of Statistics, New York: Mcgraw-Hill.

MOOD, A.M. AND BROWN, G.W. 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.

NEWTON, H.J., J.H. CARROLL, N. WANG, D. WHITING**,** 1996. Statistics 30x Class Notes. University of Leeds.U.K.

OĞUZLAR A., 1999. Parametrik Olmayan Basit Doğrusal Regresyon Analizi. Uludağ Üniversitesi IIBF Dergisi, Cilt:17, Sayı:1-2, Mayıs.

PRIESTLEY, M.B AND CHAO, M.T., 1972. Nonparametric Function Fitting, Journal of the Royal Statistical Society, Series B, 34, 385-392.

RAO, K.S.M. AND GORE, A.P., 1982. Nonparametric Tests for Intercept in Linear Regression Problems. Australian J.of Statistics, 24(1), 42-50.

RICE, J.A., 1984. Bandwidth Choice for Nonparametric Regression, The Annals of Statistics, 12, No:4, 1215-1230.

SEN, P.K., 1968. Estimates of the Regression Coefficient Based on Kendall’s Tau. J. Amer.Statist.Ass., 63, 1379-1389.

SIEVERS, G.L., 1978. Weighted Rank Statistics For Simple Linear Regression. J.Amer.Statist.Ass., 73, 628-631.

SILVERMAN, B.W., 1986. Density Estimation for Statistics and Data Analysis, Chapman & Hall, London.

STONE, C.J., 1977. Consistent Nonparametric Regression, The Annals of Statistics, Vol:5, No: 4, 595-645.

ŞAHİNLER S., 2000. En Küçük Kareler Yöntemi İle Doğrusal Regresyon modeli Oluşturmanın Temel Prensipleri. Mustafa Kemal Üniversitesi, Ziraat fakültesi Dergisi, 5(1-2), 57-73.

ŞENGÜN M., 1999. Nonparametrik Regresyon Analizi Türk İmalat Sanayiinde Çalışanların Ücret Farklılaşmalarına Yönelik Bir Uygulama. Yüksek Lisans

Tezi. İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Bölümü, İstanbul.

TATE, MERLE W. AND RICHARD C. CLELLAND, 1957. Nonparametric and Shorulcut Statistics in the Social. Biological and Medical Sciences. Danvill, III.: Interstate Printers and Publishers.

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

WANG, X. AND YU, Q., 2004. Unbiasedness of the Theil-Sen Estimator. http://home.olemiss.edu/xuegin/papers/tsun.pdf.

WATSON, G.S., 1964. Smooth Regression Analysis. Sankhya, Series A, 26, 359-372.

YILDIZ N. Ve TOPAL M., 2001. Nonparametrik Regresyon Metodlarının İncelenmesi. Atatürk Üniversitesi, Ziraat Fakültesi Dergisi, 32(4), 429-435.

YILDIZ N., TOPAL M., VE BİLGİN Ö.C., 2004. Farklı Dağılış Gösteren Verilerde Parametrik ve Nonparametrik Regresyon Metodlarının İncelenmesi. 4 Ulusal Zootekni Bilim Kongresi Sözlü Bildiriler Programı, Atatürk Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Erzurum.

ÖZGEÇMİŞ

1975 yılında Hatay’ ın Samandağ ilçesinde doğdum. İlk öğrenimimi Samandağ Mehmet Akif Ersoy İlkokulunda, orta öğrenimimi Samandağ Ortaokulunda ve lise öğenimimi Samandağ Lisesinde tamamlayarak 1995 yılında Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi İstatistik Anabilim Dalında lisans öğrenimime başladım. 2001 yılında 7. Kolordu Merkez Komutanlığında askerliğimi kısa dönem olarak yaptım. 2002 yılında Çukurova Üniversitesi Ziraat Fakültesi Zootekni Anabilim Dalında lisansüstü öğrenimime başladım.