

**T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**



**PARAMETRİK OLMAYAN REGRESYON ANALİZİNDE
KULLANILAN YÖNTEMLERİN İNCELENMESİ**

Yücel GÜL

Yüksek Lisans Tezi

İSTATİSTİK ANABİLİM DALI

TEMMUZ 2020

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İstatistik Anabilim Dalı

Yüksek Lisans Tezi

PARAMETRİK OLMAYAN REGRESYON ANALİZİNDE
KULLANILAN YÖNTEMLERİN İNCELENMESİ

Tez Yazarı
Yücel GÜL

Danışman
Prof. Dr. Mehmet GÜRCAN

TEMMUZ 2020
ELAZIĞ

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İstatistik Anabilim Dalı

Yüksek Lisans Tezi

Başlığı:	Parametrik Olmayan Regresyon Analizinde Kullanılan Yöntemlerin İncelenmesi
Yazarı:	Yücel GÜL
İlk Teslim Tarihi:	.2020
Savunma Tarihi:	.2020

TEZ ONAYI

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına göre hazırlanan bu tez aşağıda imzaları bulunan jüri üyeleri tarafından değerlendirilmiş ve akademik dinleyicilere açık yapılan savunma sonucunda OYBİRLİĞİ ile kabul edilmiştir.

Danışman:	Prof. Dr. Mehmet GÜRCAN Fırat Üniversitesi, Fen Fakültesi	<i>İmza</i> Onayladım
Başkan:	Unvan Adı SOYADI ... Üniversitesi, ... Fakültesi	Onayladım
Üye:	Unvan Adı SOYADI ... Üniversitesi, ... Fakültesi	Onayladım
Üye:	Unvan Adı SOYADI ... Üniversitesi, ... Fakültesi	Onayladım

Bu tez, Enstitü Yönetim Kurulunun/...../20..... tarihli toplantısında tescillenmiştir.

İmza
Prof. Dr. Soner ÖZGEN
Enstitü Müdürü

BEYAN

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım “Parametrik Olmayan Regresyon Analizinde Kullanılan Yöntemlerin İncelenmesi” Başlıklı Yüksek Lisans Tezimin içindeki bütün bilgilerin doğru olduğunu, bilgilerin üretilmesi ve sunulmasında bilimsel etik kurallarına uygun davrandığımı, kullandığım bütün kaynakları atıf yaparak belirttiğimi, maddi ve manevi desteği olan tüm kurum/kuruluş ve kişileri belirttiğimi, burada sunduğum veri ve bilgileri unvan almak amacıyla daha önce hiçbir şekilde kullanmadığımı beyan ederim.

.../.../2020

Yücel GÜL



ÖNSÖZ

Regresyon analizi iki veya daha fazla değişken arasında ki ilişkiyi inceleyen bir istatistiksel yöntemdir. İstatistiksel veriler dağılımlarına göre sınıflandırılır. Normal dağılıma sahip verilerde parametrik yöntemler kullanılırken normal dağılıma sahip olmayan verilerin analizinde ise parametrik olmayan istatistiksel analizler kullanılır. Bu çalışmamızda parametrik olmayan regresyon analizinde kullanılan yöntemleri inceledik. Bu yöntemlerden Nadarya Watson tahminleyicisi ve Bernstein operatörünün kullanıldığı analizde MATLAB programı kullanılarak veriler analiz edilmiştir. Uygulama olarak covid vakalarının kullanıldığı çalışma sonuç ve öneriler ile sonlandırılmıştır.

Tez çalışmam boyunca benden yardım ve desteklerini esirgemeyen, akademik kişiliği ve titiz çalışmalarıyla bana her zaman yol gösterici olan, anlayışlı yaklaşımları, sabır ve güven veren tavırlarıyla beni cesaretlendiren değerli danışman hocam Prof. Dr. Mehmet GÜRCAN 'a, bilgi ve fikirlerinden yararlandığım başta İstatistik bölümümüzün başkanı Prof. Dr. Sinan ÇALIK hocam olmak üzere bölümümüzün değerli hocalarına, her daim desteğini hissettiğim kardeşim Ali Tunç'a ve değerli aileme teşekkürlerimi ve şükranlarımı sunarım.

Yücel GÜL

ELAZIĞ, 2020

İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	iv
İÇİNDEKİLER	v
ÖZET	vi
ABSTRACT	vii
ŞEKİLLER LİSTESİ	viii
TABLolar LİSTESİ.....	ix
SİMGELER VE KISALTMALAR	x
1. GİRİŞ	1
2. MATERYAL VE METOT	2
2.1. Regresyon Analizine Genel Bakış	2
2.2. Parametrik Regresyon Analizi	4
2.3. Basit Doğrusal Regresyon	4
2.4. En Küçük Kareler Yöntemi	5
2.5. En Küçük Medyan Kareler Yöntemi	7
2.6. Parametrik Olmayan Regresyon Analizinde Kullanılan Metotlar	10
2.7. Bootstrap Yöntemi.....	13
2.8. Regresyon Analizinde Bootstrap Yönteminin Kullanımı:	14
2.9. Bernstein Çekirdeği:	14
2.10. Kernel Fonksiyonlarının Özellikleri	15
2.11. Nadaraya-Watson (NW) Yöntemi	16
2.12. Klasik ve Uygulanabilir Nadaraya-Watson Çekirdek Kestiricileri.....	16
2.13. Değişen Bant Genişliğini Kullanan Uyarlanabilir Nadaraya-Watson Çekirdek Kestiricileri:	18
3. BULGULAR VE TARTIŞMA	20
3.1. COVID-19 İle İlgili Genel Bilgi.....	20
3.2. COVID-19 Türkiye Verilerinin Bernstein Polinomu Yardımıyla İncelenmesi	21
4. SONUÇLAR.....	35
ÖNERİLER	37
KAYNAKÇA.....	38
ÖZGEÇMİŞ	

ÖZET

Parametrik Olmayan Regresyon Analizinde Kullanılan Yöntemlerin İncelenmesi

Yücel GÜL

Yüksek Lisans Tezi

FIRAT ÜNİVERSİTESİ

Fen Bilimleri Enstitüsü

Risk Analizi Anabilim Dalı

Ocak 2020, Sayfa: x + 40

Regresyon analizi iki veya daha fazla bağımsız değişkenin bağımlı değişken üzerindeki etkisini açıklamak için kullanılan istatistiksel bir yöntemdir. Bu çalışmada varsayımların gerçekleşmesi koşuluna göre kullanılacak regresyon metodu üzerinde durulmuştur. En küçük kareler yöntemi, yarı parametrik ve parametrik olmayan regresyon analizi metotları kullanılmıştır. Bu metotlarda kullanılan operatörler açıklanmıştır. Nadarya-Watson tahmincisi ve Bernstein polinomu üzerinde durulmuştur. Bernstein polinomu parametrik olmayan regresyon yöntemleri içerisinde özel bir regresyon modelidir. Son olarak çalışmada Türkiye de gözlenen COVID-19 verileri Bernstein polinomu kullanılarak incelenmiştir.

Anahtar Kelimeler: Regresyon Analizi, Nadarya-Watson Tahmincisi, En Küçük Kareler Yöntemi, Parametrik Olmayan Regresyon Analizi, Bernstein Polinomu.

ABSTRACT

Examination of Methods Used in Non - Parametric Regression Analysis

Yücel GÜL

Master's Thesis

FIRAT UNIVERSITY
Graduate School of Natural and Applied Sciences

January 2020, Pages: x + 40

Regression analysis is a statistical method used to explain the effect of two or more independent variables on the dependent variable. In this study, the regression method that will be used according to the condition of assumptions is emphasized. The least squares method, semi-parametric and non-parametric regression analysis methods are used. The operators used in these methods are explained. Nadarya-Watson estimator and Bernstein polynomial are studied. Bernstein polynomial is a special regression model among nonparametric regression methods. Finally observed in the study Covidien Turkey-19 data were analyzed using Bernstein polynomial.

Keywords: Regression Analysis, Nadarya-Watson Estimator, Least Squares Method, Nonparametric Regression Analysis, Bernstein Polynomial.

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 3.1. Yeni Vaka Sayısının Günlere Göre Grafiği.....	24
Şekil 3.2. Yeni Vaka Sayısının Tahmin Değerlerinin Grafiği.....	24
Şekil 3.3. Hata Değerleri Grafiği.....	25
Şekil 3.4. Tahmin Eğrisinin 1. Türevinin Grafiği.....	25
Şekil 3.5. Yeni Vaka / Test Sayısı Oranının Günlere Göre Grafiği.....	27
Şekil 3.6. Yeni Vaka / Test Sayısı Oranının Tahmin Değerleri Grafiği	28
Şekil 3.7. Yeni Vaka / Test Sayısı Oranının Hata Değerleri Grafiği	29
Şekil 3.8. Tahmin Eğrisinin 1. Türevinin Grafiği.....	30
Şekil 3.9. Ölüm Sayısının Günlere Göre Grafiği.....	31
Şekil 3.10. Ölüm Sayısının Tahmin Değerlerinin Grafiği	32
Şekil 3.11. Ölüm Sayısının Hata Değerleri Grafiği	33
Şekil 3.12. Tahmin Eğrisinin 1. Türevinin Grafiği.....	34

TABLÖLAR LİSTESİ

	Sayfa
Tablo 2.1. Uygun Regresyon modelinin belirlenmesi.....	3
Tablo 3.1. 11 Mart 2020 ile 4 Haziran 2020 tarihleri arasında gözlenen yeni vaka sayısı.....	22
Tablo 3.2. Vaka sayılarının Bernstein polinomu ile elde edilen tahmin değerleri	22
Tablo 3.3. Hata değerleri.....	23
Tablo 3.4. 11 Mart 2020 ile 4 Haziran 2020 tarihleri arasında gözlenen yeni vaka sayısının Test sayısına oranı	26
Tablo 3.5. Yeni vaka sayısının test sayısına oranının tahmin değerleri	27
Tablo 3.6. Yeni vaka sayısının test sayısına oranının tahminlerinin hata değerleri	28
Tablo 3.7. Yeni vaka sayısının test sayısına oranının tahminlerinin hata değerleri türevleri.....	29
Tablo 3.8. 11 Mart-4 Haziran 2020 tarihleri arasında gözlenen ölüm sayıları.....	30
Tablo 3.9. 11 Mart-4 Haziran 2020 tarihleri arasında gözlenen ölüm sayılarının tahmin değerleri gösteren tablo	31
Tablo 3.10. 11 Mart-4 Haziran 2020 tarihleri arasında gözlenen ölüm tahmin değerlerinin hata oranları verilmiştir	32
Tablo 3.11. 11 Mart -4 Haziran 2020 tarihleri arasında ölüm tahmin değerlerinin türev değerleri.	33

SİMGELER VE KISALTMALAR

Simgeler

E	: Beklenen değer.
p	: Değişken sayısı.
X	: Bağımsız değişken.
Y	: Bağımlı değişken.
β	: Regresyon katsayısı.
Cov	: Kovaryans.

Kısaltmalar

N-W	: Nadarya- Watson tahmincisi.
B	: Bernstein çekirdeği.
K	: Kernel tahminleyeni.
EKMK	: En küçük medyan kareler yöntemi.
YV	: Yeni vaka.
TS	: Test sayısı.

1. GİRİŞ

Regresyon analizi, neden sonuç bağlantısı olan iki ya da daha fazla değişken arasındaki etkileşimi incelemek ve çeşitli tahminlerde bulunabilmek için matematiksel modelleri kullanarak yapılan istatistiksel bir analiz yöntemidir.

Regresyon incelemesinde ilk olarak matematiksel modelin yapısına karar verilmelidir. Model yapısına karar verildikten sonra model parametrelerinin hesaplanabilmesi için farklı istatistiksel yöntemler kullanılmaktadır. Regresyon analizinin temelleri 19. yüzyıla kadar uzanmaktadır. Bu konu ile ilgili ilk araştırmacıların başında Francis Galton gelmektedir. Francis'in yapmış olduğu çalışmada baba ve annelerin boyları ile çocuklarının boyları arasında bir ilişkinin olup olmadığı incelenmeye çalışılmıştır. Kısa boylu baba ve annelerin çocuklarının boy uzunluklarının kısa, uzun boylu baba ve annelerin çocuklarının boy uzunluklarının ise uzun olduğu çalışmada gözlenmiştir. Buna ilaveten çocuk boy uzunluklarının ana kitle boy ortalamasına doğru yaklaşma eğilimi saptanmıştır [1]. Francis' in yapmış olduğu bu çalışma istatistik biliminde değişkenler arasındaki etkileşimi ortaya koyan regresyon analizinin temeli olarak kabul edilmektedir. Günümüzde ise regresyon analizi pek çok problemde sıklıkla kullanılan bir inceleme alanı haline gelmiştir.

Regresyon analizi temelde birkaç ana başlığa veya birkaç ana yönteme ayrılabilir. İncelenmek üzere gözlenen değişkenler toplu olarak gözlenebileceği gibi zamana bağlı olarak ta gözlenebilir. Bu durumların her birisi için farklı regresyon tipleri veya modelleri karşımıza çıkabilmektedir.

Bu tez çalışmasında amaç, parametrik olmayan regresyon analizinde kullanılan yöntemlerin incelenmesidir. Çalışmada regresyon analizi ile ilgili temel kavramlara yönelik bilgiler verilerek bazı önemli parametrik olmayan regresyon yöntemlerinden bahsedilmiş, son yaşanan pandemi süreci üzerinde durularak vakaların Türkiye de ki eğilimi üzerinde yorumlamalarda bulunulmuştur.

2. MATERYAL VE METOT

Parametrik olmayan regresyon analizinde kullanılan metotların incelendiği çalışmada regresyon analizinin genel yapısı üzerinde durulmuştur. Regresyon analizini genel bir tanımlamayla iki veya daha fazla değişken arasındaki ilişkiyi inceleyen istatistiksel bir yöntem olarak ifade edebiliriz [2]. Çalışmada parametrik olmayan metot üzerinde durulurken istatistiksel olarak non parametrik oluşum koşulları açıkça ifade edilmiştir. Parametrik yöntemlerle parametrik olmayan yöntemler karşılaştırılmış aralarındaki farklar ortaya konulmuş ve kullanım alanları anlatılmıştır. Çalışmada Nadarya Watson tahmincisi ve yöntem olarak Bernstein çekirdeği kullanılmıştır. Korona virüs verilerinin kullanıldığı örnek uygulamayla çalışma tamamlanmıştır.

2.1. Regresyon Analizine Genel Bakış

Regresyon yöntemleri, modele alınacak değişken sayısına, matematiksel yapıya göre farklı şekillerde sınıflandırılmaktadır. Aşağıda değişken sayısına ve matematiksel modele göre bazı regresyon türleri verilmiştir [3].

Basit doğrusal regresyon modeli:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (2.1)$$

Çoklu doğrusal regresyon modeli:

$$= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (2.2)$$

Polinomsal regresyon modeli:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_r X^r + \varepsilon \quad (2.3)$$

Lineer olmayan regresyon modeli:

$$Y = f(X, \beta) + \varepsilon \quad (2.4)$$

Uygun modelin belirlenmesi kesin kurallara tabi olmamasına rağmen, Tablo2.1’de uygun modelin belirlenmesindeki yapısal özellikler ile ilgili bilgiler yer almaktadır [4].

Tablo 2.1. Uygun Regresyon modelinin belirlenmesi.

Regresyon Analizi Ayrımı	Regresyon Modeli	Bilgi Notları
Bağımlı değişken sayısına göre	Tek Değişkenli Regresyon	Tek 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 bağımsız değişken içermelidir.
	Çoklu Regresyon	Bir bağımlı iki veya daha fazla bağımsız değişken içermelidir.
Matematiksel biçimlere göre	Doğrusal Regresyon	Tüm parametreler doğrusal olmalıdır. Dönüştürme yöntemiyle parametrik olan değişkenlerde bu kısımda yer alır.
	Doğrusal olmayan Regresyon	Bazı parametrelerin doğrusal olmadığı durumlarda, değişken dönüştürme yöntemleri ile de doğrusal hale getirilemeyen değişkenler için kullanılır.
Bağımsız gölge değişkeni	Varyans Analizi modeli	Tüm bağımsız değişkenler nitel olmalıdır.
	Kovaryans Analizi modeli	Bazı bağımsız değişkenler nitel bazıları ise nicel olması durumunda kullanılır.
Varsayımların sağlanması durumuna göre	Parametrik Regresyon	Verinin normal dağılıma uyması, eşit varyanslı olması oto korelasyonlu olmaması gibi durumları içerir.
	Parametrik olmayan Regresyon	Parametrik regresyonun uygulanması için gerekli varsayımların sağlanmadığı durumlarda kullanılır.

2.2. Parametrik Regresyon Analizi

Matematiksel modelin istatistiksel olarak anlamlı hale gelebilmesi için $y = f(x)$ denkleminde x ve y 'nin tesadüfi değişken olarak kullanılması gerekmektedir. Bu sayede matematiksel eşitlik anlamlı hale gelebilir. X bağımsız değişken Y ise bağımlı değişken olarak adlandırıldığında $Y = f(X) + \varepsilon$ bağıntısı $\varepsilon \sim N(0, \sigma^2)$ şartı altında araştırmacıya daha fazla kullanım imkanı vermektedir. Bu noktada kullanılan tesadüfi değişkenlerin bir dağılıma sahip olduğunu kabul ettiğimizde artık kullandığımız yöntem parametrik yöntemdir. Regresyona da parametrik regresyon denilmesinin nedeni budur [5].

Bu aşamada artık araştırmacının elinde regresyon denklemi olarak $y = f(x)$ eşitliği değil $E\{Y|X = x\}$ şartlı ortalaması vardır. Bu değer bize bağımsız değişkenin $X = x$ sabit bir değerine karşılık bağımlı değişken Y 'nin alabileceği ortalama değeri göstermektedir[6].

Modeldeki hata terimi $\varepsilon = Y - f(x)$ tahmin değerleri karşısında ne kadar küçük kalabilirse regresyon modelinin gözlenen verilere uyumu o kadar iyi olur. Bu olgudan yola çıktığımızda regresyon katsayılarının tahminini yapabilmek için aşağıdaki hata kareler toplamını minimum yapmamız yeterli olacaktır [7].

$$S = \sum_j (Y_j - \hat{Y}_j)^2 \quad (2.5)$$

2.3. Basit Doğrusal Regresyon

Basit doğrusal regresyonda regresyon denklemi aşağıdaki şekildedir,

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (2.6)$$

Regresyon ile Korelasyon kavramları birbirleri ile iç içe geçmiş istatistiksel kavramlardır. X ve Y değişkenleri arasında doğrusal bir ilişkinin olduğunu varsaymak ve bu iki değişken arasındaki bağıntının regresyon denklemini belirlemek için aşağıdaki koşulların geçerli olması gerekir [8].

- 1) Değişkenler normal dağılıma sahip olmalıdır,
- 2) Bağımsız değişken ölçme hatası taşınamalıdır,
- 3) n birimden elde edilen X ve Y değişkenlerinin serpm diyagramı çizildiğinde ilişki grafiğinde yer alan $Q(X_i, Y_i)$ noktaları bir elips içinde toplanmalıdır,
- 4) Hata terimi normal dağılıma sahip olmalı, $\varepsilon \sim N(0, \sigma^2)$ ve hata terimleri aralarında ilişkisiz olmalıdır,
- 5) Modelde gözlem sayısı parametre sayısından az olmamalıdır.

2.4. En Küçük Kareler Yöntemi

En küçük kareler regresyon yöntemi hata kareler toplamını minimuma düşürmeyi amaçlayan bir optimizasyon yöntemidir. Yani başka bir ifade ile En Küçük Kareler tekniği kullanılarak, örneklemden elde edilen verilerle artık kareler toplamını en küçük yapan regresyon doğru denkleminin tahmin edilmeye çalışılmasıdır. Bu yöntem, verinin normal dağılıma uygunluk, sabit varyanslık, aşırı uç değer içermeme gibi bazı varsayımların sağlanması durumunda regresyon katsayılarının doğru bir şekilde tahmin edilmesini sağlamaktadır [9].

En küçük kareler yöntemi hatayı minimum yapan yegâne yöntemdir. Ancak bazı durumlarda hatanın minimum olması araştırmada yetmemektedir. Regresyon analizindeki en önemli amaç hatayı minimum yaparken hatanın varyansının da minimum kalmasıdır. İstenen bu iki durum bazı verilerde beraber sağlanmamaktadır. Bu durumda araştırmacı iki seçenektan birisini tercih etmelidir. Ya hata minimum olmalı ya da varyans minimum olmalıdır. Bu iki seçenekle karşı karşıya kalındığında araştırmacılar daima varyansın minimum olmasını tercih etmelidirler. Bu durumda tahmin denklemi ara değer problemini sağlıklı bir şekilde cevaplayabilmektedir [10].

Regresyon analizinde varsayımların sağlanmadığı yapılardan birisi de veride sapan değerlerin var olduğu durumlardır. Sapan değer, veri kümesinde gözlemlerin çoğunun sahip olduğu dağılıma veya modele uymayan gözlemler olarak ifade edilir [11]. Sapan değer içeren veri setinde varsayımların sağlanmamasından ötürü kurulan regresyon modelinden alınan sonuçlarda yanıltıcı olacağından bu metodun pekte sağlıklı olduğu söylenemez.

Bu nedenle regresyon analizinde verilerin analizinin oldukça önemli olduğunu belirtmek gerekir. Sapan değerlerin veri setinden çıkartılması ve veri setinin normal dağılıma çekilmesi regresyon denklemini kısmen veya tamamen değiştirebildiğini söyleyebiliriz. Bu yöntemlerle regresyon metodunun daha sağlıklı çalışıp daha doğru sonuçlar vereceği de unutulmamalıdır. Bundan ötürü büyük artık değerleri var olan gözlemler regresyon metodunun çözümlenmesinde oldukça etkilidir. Bu tür durumlarda sapan değerlerin belirlenmesi ve sonuçların güvenilirliği için güçlü regresyon yöntemlerinin tercih ediliyor olması bizleri daha sağlıklı sonuçlara götürüyor olacaktır. Bu güçlü metotlardan birisi de en küçük medyan kareler (EKMK) metodudur [12].

Daha öncede belirtildiği gibi Regresyon analizi aralarında neden-sonuç ilişkisi olan iki veya daha çok değişken arasındaki ilişkiyi ortaya koyarak o konu hakkında tahminlerde bulunabilmek ya da kestirimler yapabilmek amacıyla kullanılan istatistiksel bir metottur. Bu çözümlemenin yapılabilmesi için ise matematiksel materyallere ihtiyaç duyulur ve bu materyaller kullanılarak bir model kurulur ve bu modele regresyon modeli denilir. İstatistiksel olarak model kurduktan sonra kurulan modelin gerçekliliğini ortaya koyabilmek regresyon analizinde çözümlemenin önemli bir noktasıdır. Tahmin edilen modelin esas modele ne kadar yaklaştığını belirlemek için ise EKK yöntemi kullanılır [13].

Kitleden alınan n birimlik örneklem için basit doğrusal regresyon (BDR) denklemi aşağıdaki şekilde yazılır,

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.7)$$

BDR çözümlemesinde kullanılacak olan regresyon denklemlerinin tahmin amacı ile kullanılabilmesi için; hata terimlerinin ($\varepsilon_i = Y_i - \hat{Y}_i$) tesadüfi olup normal dağılım göstermesi, hataların beklenen değerlerinin sıfır olması ve varyanslarının sabit olması, hatalarının birbirinden bağımsız olması, hata terimleri ile bağımsız değişken arasında korelasyonun bulunmaması gibi bazı varsayımların sağlanması gerekmektedir [14]. Bu varsayımlardan en az birinin sağlanmaması durumunda EKK metodu ile gözlemler ve ön kestiriciler üzerindeki kararlı ve küçük varyanslı olma gibi faktörleri kaybederek tutarsız veya etkisiz olacağı söylene bilinir.

Ana Kitle regresyon denkleminde yer alan β_0 ve β_1 parametrelerinin örneklemde elde edilen kestirimleri $\widehat{\beta}_0$ ve $\widehat{\beta}_1$ olarak ifade edilecek olunursa, tek değişkenli regresyon modelinin tahmin denklemi aşağıdaki gibi gösterilebilir [15].

$$\hat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i, \quad i = 1, 2, \dots, n \quad (2.8)$$

Denklemden kullanılan $\widehat{\beta}_0$ ve $\widehat{\beta}_1$ terimlerinin değerleri tespiti için EKK metodunun temelini, toplam sapmaların karelerinin toplamını minimum yapacak değerlerin bulunması oluşturmaktadır. Hata terimleri ise gözlemlenen Y_i değeri ile beklenen $\hat{Y}_i$ değerleri arasında ki farklar alınarak bulunmaktadır [16].

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i \quad (2.9)$$

Üstteki eşitlikte verilen ifade ile hesaplanan hata terimleri pozitif, negatif veya sıfır değerine sahip olabiliyor iken bu farkların toplamı ise;

$$\sum_{i=1}^n \hat{\varepsilon}_i = \sum_{i=1}^n (Y_i - \hat{Y}_i) = 0 \quad (2.10)$$

olur. EKK metodu, β_0 ve β_1 parametrelerinin kestirimi olan $\widehat{\beta}_0$ ve $\widehat{\beta}_1$ 'nin farkını minimum yapacak şekilde aşağıdaki gibi belirler;

$$S = \text{minimum} \sum_{i=1}^n \hat{\varepsilon}_i^2 = \text{minimum} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2.11)$$

Burada regresyon katsayılarının EKK metodunun tahminlerini elde edebilmek amacı ile üstteki eşitlikte $\hat{\beta}_0$ ve $\hat{\beta}_1$ 'ya göre kısmi türevi alınıp sıfıra eşitlendiğinde normal denklemler elde edilir [17]. Bu eşitlikler kullanılarak gerekli yöntemlerle çözümlemeler yapıldığında β_0 ve β_1

parametrelerinin tahminleri olan $\hat{\beta}_0$ ve $\hat{\beta}_1$ 'değerlerinin bulunabileceği eşitlikler 2.12 ve 2.13' de ki gibi elde edilir;

$$S = \sum_{i=1}^n (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{li}))^2 \quad (2.12)$$

$$\sum_{i=1}^n Y_i = \hat{\beta}_0 n + \hat{\beta}_1 \sum_{i=1}^n X_i \quad (2.13)$$

$$\sum_{i=1}^n X_{li} Y_i = \hat{\beta}_0 \sum_{i=1}^n X_{li} + \hat{\beta}_1 \sum_{i=1}^n X_{li}^2 \quad (2.14)$$

$\hat{\beta}_0$ ve $\hat{\beta}_1$ regresyon belirtme katsayılarının hesaplanması ise aşağıdaki gibidir.

$$\hat{\beta}_1 = \frac{n[\sum_{i=1}^n X_{li} Y_i] - (\sum_{i=1}^n X_{li})(\sum_{i=1}^n Y_i)}{n(\sum_{i=1}^n X_{li}^2) - (\sum_{i=1}^n X_{li})^2} = \frac{\sum_{i=1}^n (X_{li} - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_{li} - \bar{X})^2} \quad (2.15)$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n Y_i + \hat{\beta}_1 \sum_{i=1}^n X_{li}}{n} = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (2.16)$$

$$\sum_{i=1}^n X_{li} Y_i = \hat{\beta}_0 \sum_{i=1}^n X_{li} + \hat{\beta}_1 \sum_{i=1}^n X_{li}^2 \quad (2.17)$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{y})^2}{\sum_{i=1}^n (Y_i - \bar{y})^2} \quad (11) \quad (2.18)$$

2.5. En Küçük Medyan Kareler Yöntemi

Gerekli koşulların sağlanmadığı durumlarda da regresyon tahmini metoduna alternatif bir metot olarak EKMK yöntemi ele alınabilir. Güçlü tahmin yöntemlerinden olan regresyon metodu içerisinde EKMK yöntemi en fazla kullanılan yöntemlerinden birisi olarak karşımıza çıkmaktadır [18].

Veri kümesinde sapan tek bir değer bulunması halinde bile bu değer, diğer bütün verilerden elde edilecek olan bilgilerin önüne geçecek ve ilgili istatistiğin güvenilirlik seviyesini aşağıya çekecektir. Ortalama standart sapma ve aşırı student sapma ya bağlı olarak Extreme Studentized Deviate testi, veri setinde tek bir tane sapan değer olduğu saptanınca kullanılır [19]. Ancak veri setinin birden fazla sapan değer değer içerdiği koşullarda bu değerler bazen birbirlerini örtbilmekte ve hatta bu değerler klasik tahmin yöntemleri ile belirlenen güvenilir verileri bile sapan değer olarak görülmesine neden olabilmektedir. EKMK metodu veri setinin %50'ye kadar sapan değer içerdiği koşullarda da iyi tahmin değerleri ortaya koyan güçlü bir regresyon modeli

çözümlemesi olarak ele alınmaktadır. Ancak EKMK metodu artıkların medyan değerini minimum yapmayı hedeflerken arta kalan $(n-1)$ tane gözlem ise dikkate alınmaz. Bu sebeplerden ötürü örneklem sayısı arttıkça (örnekler çoğaldıkça) regresyon katsayısının kestirilmesinde EKMK metodu EKK metodu kadar etkiyi ortaya koyamamaya başlar [20].

WALD metodu, iki değişkenli bir örneklem kullanarak tesadüfi değişkenin medyanını temel alarak, bu gözlem değerlerinin serpmeye diyagram grafiğinde veri setini sol ve sağ bölge olmak üzere basit bir metodu önermiştir. Ayrıca ayrılan her bölgenin bağımlı ve bağımsız değişkenlerin ortalaması $((\bar{x}_{sağ}, \bar{y}_{sağ}), (\bar{x}_{sol}, \bar{y}_{sol}))$ şeklinde ifade edilirken bu ortalamaların hesaplanmasında yalnızca o bölgeye ait olan bağımlı ve bağımsız değişkenlerin gözlem değerleri kullanılır.

NAIR ve SHRIVASTAVA tarafınca önerilen metotta ise ilk olarak iki değişkenli veri kümesindeki X ve Y değişkenleri kendi aralarında sıralanır. Daha sonra sıralanan bu değişkenlerin $(x_1 \leq x_2 \leq \dots \leq x_n)$, birbirine yakın olan değerleri aynı parçada olacak şekilde üç ana bölgeye ayrılır. Son olarak WALD yönteminde de bahsettiği gibi ayrılan her bölgenin X ve Y gözlenen değerlerinin ortalaması $((\bar{x}_{sağ}, \bar{y}_{sağ}), (\bar{x}_{sol}, \bar{y}_{sol}))$, yalnızca o bölgeye ait X ve Y gözlenen değerleri kullanılarak elde edilir.

$$\begin{aligned}\bar{x}_{sol} &= \frac{x_1 + x_2 + \dots + x_{n_{sol}}}{n_{sol}} \\ \bar{y}_{sol} &= \frac{y_1 + y_2 + \dots + y_{n_{sol}}}{n_{sol}}\end{aligned}\tag{2.19}$$

$$\begin{aligned}\bar{x}_{sağ} &= \frac{x_{n-n_{sağ}+1} + x_2 + \dots + x_n}{n_{sağ}} \\ \bar{y}_{sağ} &= \frac{y_{n-n_{sağ}+1} + y_2 + \dots + y_n}{n_{sağ}}\end{aligned}\tag{2.20}$$

Eşitliklerde n_{sol} : birinci gruba ait gözlenen değişken sayılarını, $n_{sağ}$: ikinci gruba ait gözlem sayılarını göstermektedir. Geriye kalan $(n - n_{sağ} - n_{sol})$ tane gözlem veri kümesinden çıkartılır. Buradan birinci ve ikinci gruba ait gözlem sayıları $n/3$ değerine varacak şekilde bir tamsayı değeri olup her iki gruba ait gözlem sayıları da birbirine eşittir. Bu eşitlik $n_{sağ} = n_{sol}$ olacak biçimde ifade edilir. EKMK metodunun uygulandığı basit doğrusal regresyona ait parametre kestirimi aşağıdaki şekilde hesaplanır.

$$\hat{\beta}_0 = \bar{y}_{sol} - \hat{\beta}_1 \bar{x}_{sol} = \bar{y}_{sağ} - \hat{\beta}_1 \bar{x}_{sağ}\tag{2.21}$$

$$\hat{\beta}_1 = \frac{\bar{y}_{sağ} - \bar{y}_{sol}}{\bar{x}_{sağ} - \bar{x}_{sol}} \quad (2.22)$$

ROUSSEEUW' in EKK tahminini 2.5. eşitlikte verilen amaç fonksiyonunda “ Σ ” yerine “medyan” koymak olarak tanımlar.

$$en\ küçük\ medyan\ [Y_i - \hat{Y}_i]^2 \quad (2.23)$$

Fakat 2.16-ıncı eşitlikte ortaya konulan amaç fonksiyonundan analitik bir çözümleme ile uygun sonuca varmak oldukça zor olduğundan, β_i parametre tahminlerinin değerleri bilgisayar iterasyonları ile ancak buluna bilinir.

Parametrelerden kestirimlerin bulunmasını sağlayan iterasyonlar için farklı algoritmalar oluşturulması gerektiğinin üzerinde durmuştur. Geniş şekilde kullanılan algoritmasında, n elemanlı bir veri kümesinin tüm mümkün p elemanlı alt kümelerine $\binom{n}{p}$ EKK metodu uygulanır ve her biri için medyan artık değeri bulunur [21,22, 23, 24]. Bu değerler içerisinde minimum medyan değerine sahip olan alt kümenin EKK metodu ile tahminleri EKK yöntemi tahmini olarak kabul edilir. Küçük veri setleri için EKK yöntemi ile bulunan tahminlerin kesin değerlerini bulmak bu şekilde mümkün olsa da, büyük verilerden oluşan kümelerde mümkün olan tüm altkümelerin elde edilmesi ve EKK metodunun uygulanması işlem yükü açısından çokta mümkün görünmemektedir. Bu durumda veriler arasından bazı altkümelerin rasgele olarak çekilmesi ve amaç fonksiyonunun bu altkümelerde üzerinde uygulanması daha doğru bir tercih olacaktır. Belirli kısıtlar altında verilen rasgele bir biçimde çekilecek olan en az bir altkümenin amaçlanan sonucu verme olasılığının 1'e yakın olduğunu ispatlamıştır. Buna bakılacak olunursa bir veri setinden p elemanlı k adet altküme seçtiğimizi düşünürsek p tane aşırı olmayan değere sahip en az bir altküme rastlama olasılığını (n/p) 'nin maksimum değerleri için aşağıdaki ifadeye eşit olacağını göstermiştir [25-26].

$$P_{sapandeğer\ içermeyen} = 1 - [1 - (1 - \varepsilon)^p]^k \quad (2.24)$$

2.24 numaralı eşitlikte veri setinin kirlilik oranı ε ile gösterilmektedir. Bu ifade bakılarak kirlenme oranının ε olduğu bir veriden p birimlik k tane alt küme çekildiğinde bunlardan en az birinin sapan değer içermeyen gözlemlerden oluşuyor olma olasılığı bulunur. Kirlenme oranının $\varepsilon = \%50$ olduğu bir veriden 15 birimlik alt kümeler çekildiğinde bunlardan en az bir tanesinin sapan değer içermeyen gözlemlerden oluşma olasılığını aşağıdaki gibidir.

$$1 - [1 - (1 - \varepsilon)^p]^k = 0.95$$

olması için çekmemiz gereken 15 birimlik alt kümelerin sayısı 0.95'dir.

EKMK metodu için standart sapma değeri ve regresyon modelinin açıklayıcı katsayısı değeri aşağıdaki gibi gösterilebilir. s^0 , toplam gözlem sayısı ve açıklayıcı değişken sayısına (n: gözlem sayısı, m: açıklayıcı değişken sayısı) bağlı bir düzeltme çarpanıyla çarpılmasıyla bulunabilir.

$$s^0 = 1,4826 \left(1 + \frac{5}{n-m}\right) \sqrt{med\sigma_i^2} \quad (i = 1, \dots, n) \quad (2.25)$$

s^0 Tahmin değeriyle standartlaşmış r_i/s^0 artıkları bulunur ve aşağıdaki gibi gözlemin w_i aralığının gösteriminde kullanılır.

$$w_i = \begin{cases} 1, & |r_i/s^0| \leq 2.5 \\ 0, & d.d \end{cases} \quad (2.26)$$

EKMK yöntemi ile regresyon için standart sapma aşağıdaki eşitlikte verilen ifade ile bulunur.

$$\sigma^* = \sqrt{\frac{\sum_{i=1}^n w_i \sigma_i^2}{\sum_{i=1}^n w_i - m}} \quad (2.27)$$

İfadede σ^* , %50 kirlilik oranına sahip olan bir data kümesi için standart sapmanın elde edilmesini gösterir. EKMK yöntemi ile regresyon için bağımlı değişkendeki değişimin ne oranda model tarafından açıklana bilindiğine bakılır. Tanımlayan belirleyicilik katsayısı ise regresyon modelinde sabit terim olup olmama durumuna göre aşağıdaki şekilde elde edilir.

Sabit Terimli Regresyon Modeli Sabit Terimi Olmayan Regresyon Modeli;

$$R^2 = 1 - \left(\frac{med|r_i|}{mad(y_i)}\right)^2 \quad R^2 = 1 - \left(\frac{med|r_i|}{med(y_i)}\right)^2 \quad (2.28)$$

İfadede mad=medyanın mutlak sapması (median absolute deviation) kısaltması olup $mad(y_i) = |y_i - med_i(y_i)|$ ile bulunur [27].

2.6. Parametrik Olmayan Regresyon Analizinde Kullanılan Metotlar

Parametrik olmayan regresyon modellerinin en önemli özelliği, parametrik olan yöntemlerin aksine modele ait bir fonksiyonel denklemi dikkate almadan bağımsız değişken ile bağımlı değişkenler arasındaki ilişkiyi inceliyor olmasıdır. Parametrik olmama durumu, regresyon eğrisinin (doğrusunun) esnek fonksiyonel formunu kapsamamasıdır. Ana kitle regresyon fonksiyonun genel gösterimi aynı zamanda parametrik olmayan regresyonun da genel gösterimi olarak kullanılmaktadır. Parametrik modellerin oluşturmuş olduğu dezavantajlar sonucunda birçok da parametrik olmayan model ortaya konulmuş ve parametrik olmayan tahmincilerin düzgün tahminleri üretmek amacıyla geliştirilmiştir [28].

Parametrik olmayan yöntemlerle regresyon ilişkilerinin belirlenmesinde bazı avantajlar vardır bunlar;

1. Her hangi bir fonksiyonel ifadeye ihtiyaç duyan parametrik modeli temel almaksızın gözlemlere ilişkin tahmin de bulunabilmesi.
2. Bağımlı ve bağımsız değişkenler arasındaki ilişkiyi ortaya koyarken çok yönlü ve esnek metotlarla bunu sağlaması.
3. Uç değerlerin analizde yer alması sebebiyle “sahte” gözlemleri bulmada bir araç olması.
4. Veri seti içerisinde kayıp gözlemlerin giderilmesi için esnek bir metot olması.

Parametrik regresyon modelinde uç değerler tam olarak dikkate alınmaz iken bu değerler, grafiklerin ölçeklerini etkiler, böylece verinin ana yapısı anlaşılabilir bir hal alır. Parametrik modeller için uç değerlerin varlığını ortaya koymaya çalışan birçok yöntem geliştirilmiş olsa da bu yöntemler de uç değerlerden etkilenmekte ve parametrik modeller uç değerlerin teşhisinde güvenilir değerler elde etmemektedir. Bu gerçeklik doğrultusunda uç değerlerin varlığında parametrik yöntemlerle elde edilen sonuçların güvenilirliği de tartışmaya açık olacaktır. Parametrik olmayan bir regresyon metodunu aşağıdaki gibi gösterebiliriz;

$$Y = m(X) + \varepsilon \quad (2.29)$$

Bu regresyon modelinde parametrik olmayan yaklaşım olarak adlandırılan yapı, $m(X)$ ' in herhangi bir fonksiyonel model varsayımından tahmin edilmesidir. Parametrik olmayan tahmin bu yapıyla ilgilenmektedir. Parametrik olmayan modellere özgü “ smoothing ” (düzgünleştirme) teknikleri ile $m(X)$ ana kitle regresyonuna ait en uygun düzgün tahmini üretmeye çalışmaktadır [52].

$\{X_i, Y_i\}_{i=1}^n$ veri kümesinin düzgünleştirilmesi, regresyon ilişkisindeki m ' in ortalama tepki eğrisinin yaklaşımını kapsamaktadır. Burada m ' in fonksiyonel formu regresyon eğrisinin kendisi, türevleri ya da uç değerleri gibi türevlerinin fonksiyonları olarak belirlenebilmektedir.

Analizde kullanılan verilere göre $m(X)$ ' in tahmini farklılık göstermektedir. Bağımsız değişken X ' in gözlenen değeri içerisinde tekrarlanan veya tekrarlanmayan değerlere göre ortaya çıkan farklılık sebebiyle tahmin değeri veri setine göre farklılık gösterecektir [29].

Tekrarlanan değer bağımsız değişkene ait ise bu durumda regresyon eğrisinin tahmini, bağımlı değişkenin uygun değerinin ortalaması yardımıyla elde edilebilmektedir.

Bağımsız değişkene ait gözlenen değerler X_i ($i = 1, 2, \dots, n$) ise bağımsız değişkene ait tüm gözlem değerleri içerisinde x gibi bir sabit bir gözlem belirlendiği durumda regresyon eğrisinin tutarlı tahmincisi aşağıdaki gibi elde edilir [30].

$$\hat{m}(X_i) = \frac{1}{n} \sum_{i=1}^n 1(X_i = x) \quad (2.30)$$

Burada $1(x_i = x)$, gösterge fonksiyonudur. Söz konusu gösterge fonksiyonu $x_i = x$ olduğu durumda “ 1 ” değerini, bu eşitlik sağlanmadığı durumlarda ise “ 0 ” değerini almaktadır. Bu gösterge fonksiyonundan hareketle $m(X_i)$ eğrisinin tahmini, X_i bağımsız değişkeninin sabit değeri ile kesit verilerinin kullanıldığı regresyon analizinde bu şekilde gerçekleştirilmektedir. Bu yaklaşım ise “Yerel Histogram Yaklaşımı” olarak adlandırılmaktadır, bir diğer adı ise “Yerel Ortalama Yaklaşımı”dır [31].

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 değeri etrafındaki ortalaması alınarak tahmin edilmelidir. Bu sebeple yerel histogram yaklaşımı ile yapılan parametrik olmayan tahminlerde kesit verileri yerine sürekli veriler ile analizler yapmak daha gerçekçi analizlerdir. Sürekli verilerle yapılan analizlerde bu yaklaşımla elde edilen regresyon eğri tahmininin hesaplanması da farklılık göstermektedir.

Burada X_i değerlerinin beklenen x değeri etrafında ortalaması ile kastedilen, $h/2$ gibi bir değere denk gelen aralık etrafındaki ortalamadır. h düzgünleştirme parametresi ya da bant aralığı olarak adlandırılmaktadır. Bu bilgiler ışığında yerel ortalama yaklaşımına göre elde edilen eğri tahmincisi aşağıdaki gibidir [32].

$$\hat{m}(X) = \frac{1}{nh} \sum_{i=1}^n I(-1/2 \leq \frac{x_i - x}{h} \leq \frac{1}{2}) \quad (2.31)$$

Burada $\hat{m}(X)$, orta noktası x olan $(x - \frac{h}{2}, x + \frac{h}{2})$ 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 değişkenlerine bağlı olarak değişmektedir [33].

Yerel ortalama yaklaşımında, $w_{ni}(x)$ aralığının pozitif değerler aldığı ve bağımsız değişkene ait tüm değerlerin toplamının 1 olduğu varsayılır,

$$\frac{1}{n} \sum_{i=1}^n w_{ni}(x) = 1 \quad (2.32)$$

Böylece $\hat{m}(X)$ ’in x noktasına göre en küçük kareler tahminine ait minimizasyonu şu şekildedir,

$$\frac{1}{n} \sum_{i=1}^n w_{ni}(x)(y_i - \hat{m}(x))^2 \quad (2.33)$$

Regresyona ait hata terimleri ikinci dereceden, karesel, ağırlıkla elde edilmektedir. Bu nedenle yerel ortalama yaklaşımı, yerel ağırlıklı en küçük kareler yöntemi ile benzerlikler göstermektedir [34].

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. Bu sayede bağımlı ve bağımsız değişkenler arasında nedensellik ilişkisi kurulmaya çalışılmaktadır. Ayrıca yerel ağırlıklı ortalama yaklaşımı genellikle doğrusal olmayan düzgünleştirme tekniklerine uygulanmaktadır [35,36].

Parametrik olmayan regresyon modellerinde tahmin yöntemleri:

Parametrik olmayan regresyon yöntemlerinde $m(X)$ ana kütle regresyon eğrisinin tahmin edilmesiyle bağımlı ve bağımsız değişkenler arasında ilişkinin yönü ve boyutunu belirlemek için en küçük kareler tahmincileriyle benzerlik gösteren ancak parametrik olmayan yöntemlere dayanan tahminciler literatürde kullanılmaktadır [37].

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

$$\hat{m}(x) = \frac{1}{\frac{1}{nf} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)} \int y \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{y_i - y}{h}\right) K\left(\frac{y_i - x}{h}\right) dy \quad (2.34)$$

2.7. Bootstrap Yöntemi

Bu yöntem, diğer yöntemlerin kullanılmasının uygun olmayacağı veya parametrik varsayımların sağlanmadığı durumlarda kullanılabilecek yöntemlere den birisidir. Bunun yanı sıra, kitleden alınan örnek veri kümesine bağlı bir tahminin güvenilirliğini de ölçmek amacıyla bu metod kullanılabilir. Bu yöntemi aşağıdaki gibi açıklayabiliriz [38].

n sayıda gözlemden oluşan, $X = (x_1, \dots, x_n)$ orijinal veriler kullanılarak iadeli ve rasgele yapılan örnekleme ile yeni bir veri setine $x^* = (x_1^*, \dots, x_n^*)$ ulaşılmaktadır. Orijinal veri setinden örnekleme alınarak oluşturulan x^* veri setinin içinde, bazı x_1 verilerinin birden daha fazla ya da hiç görülmemesi olasılığı bulunmaktadır. Yeniden örnekleme işlemi istenildiği kadar tekrarlanarak birbirinden farklı B adet Bootstrap gözlem seti oluşturulabilir ve $b = 1, \dots, B$ için oluşan bootstrap data seti x^{*b} ile gösterilebilir. İlgilenilen İstatistik oluşan bu yeni veri seti kullanılarak hesaplanır [39].

2.8. Regresyon Analizinde Bootstrap Yönteminin Kullanımı:

Değişkenler arasında olan ilişkiyi araştırarak ortaya çıkaran en önemli istatistiksel metotlardan birisi de regresyon analiz metodudur. Bu metotla hata terimleri ve bağımsız değişkenlerle ilgili yapılan varsayımların sağlanmaması durumunda bir düzeltme işlemi amacıyla da kullanılan bootstrap metodu, tahmin hatalarını en aza indirerek, güven aralıklarının oluşturulmasını sağlamak amacıyla geliştirilmiştir bir yöntemdir [40,41].

Doğrusal regresyon modeli $Y = X\beta + \varepsilon$ şeklinde olsun. Var olan örnekleme Y ve X değişkenleri ile oluşturulan regresyon denklem $Y = X\beta + \varepsilon$ tahmin edilir. Burada $\varepsilon = Y - \hat{Y}$ kullanılarak hata vektörü tahmin edilir [42].

Burada hata terimlerinin bootstrap örnekleri elde edilir ve yeniden örneklenen hata terimleri denkleme ilave edilerek bootstrap Y değeri hesaplanır [43,44].

$$Y_b^* = \hat{Y} + \hat{\varepsilon}_b^* \quad (2.35)$$

Bootstrap yöntemi uygulanmış Y_b^* bağımlı ve X bağımsız değişkenlerinden hareketle her bir üretilmiş bootstrap örnekleme için $\hat{\beta}_b$ regresyon katsayısının bootstrap tahmini elde edilir [45].

$$Y_b^* = X\hat{\beta}_b^* + \hat{\varepsilon}_b^* \quad (2.36)$$

Bu noktada parametre tahmini yapılırken hangi tahmin edicinin özellikleri ile ilgileniyorsa o yöntem kullanılarak parametrenin tahmini yapılır. Bu işlemin “ B ” sayıda tekrarlanması ile tahmin edicinin örneklem dağılımı ve özellikleri elde edilir. Bunlara dayanan güven aralığı oluşturulur [46,47].

2.9. Bernstein Çekirdeği:

Sergei Natonovich tarafından bir polinom olarak kurulmuştur. Bernstein çekirdeği şu şekildedir [48].

$$B_{i,n}(x) = \binom{n}{i} x^i (1-x)^{n-i}, \quad i = 0, 1, 2, \dots, n \quad (2.37)$$

İkinci dereceden bir Bernstein çekirdeği aşağıdaki gibidir;

$$B_{0,2}(x) = (1-x)^2, \quad (2.38)$$

$$B_{1,2}(x) = 2x(1-x), \quad (2.39)$$

$$B_{2,2}(x) = x^2 \quad (2.40)$$

Üçüncü dereceden bir Bernstein çekirdeği ise aşağıdaki şekildedir;

$$B_{0,3}(x) = (1 - x)^3, \quad (2.41)$$

$$B_{1,3}(x) = 3x(1 - x)^2, \quad (2.42)$$

$$B_{2,3}(x) = 3x^2(1 - x), \quad (2.43)$$

$$B_{3,3}(x) = x^3 \quad (2.44)$$

Bernstein Çekirdeğinin Bazı Özellikleri: Simetrik özelliği $B_{i,n}(x) = B_{n-i,n}(1 - x)$, Pozitif tanımlılık özelliği $B_{i,n}(x) \geq 0$, Olasılık özelliği $0 \leq x \leq 1$ için $\sum_{i=0}^n B_{i,n}(x) = 1$ [49].

$f: [0,1] \rightarrow \mathbb{R}$ sürekli bir fonksiyon ve n pozitif bir tam sayı olmak üzere Bernstein polinomu aşağıdaki şekilde tanımlanır [50].

$$B_n(f; x) = \sum_{i=0}^n f(i/n) \binom{n}{i} x^i (1 - x)^{n-i} = \sum_{i=0}^n f(i/n) B_{i,n}(x) \quad (2.45)$$

Yukarıda tanımlanan Bernstein polinomunun en önemli özelliği $[0,1]$ aralığında f sürekli fonksiyonuna düzgün yakınsamasıdır.

Bernstein polinomu binom dağılımı ile yakından ilişkilidir. $S_n \sim \text{Binom}(n, x)$ olmak üzere Bernstein polinomu aşağıdaki beklenen değer şeklinde ifade edilir [51].

$$B_n(f; x) = E f(S_n/n) \quad (2.46)$$

Bernstein polinomu Korovkin tipli yakınsama kriterlerini de sağlar [52].

$$B_n(1; x) = 1 \quad (2.47)$$

$$B_n(x; x) = x, n \rightarrow \infty \quad (2.48)$$

$$B_n(x^2; x) = x^2, \quad n \rightarrow \infty \quad (2.49)$$

2.10. Kernel Fonksiyonlarının Özellikleri

Kernel fonksiyonu bir ağırlık fonksiyonudur ve aşağıdaki şartları sağlaması gerekmektedir.

$$1. K(u) \geq 0 \quad (2.50)$$

$$2. \int_{-\infty}^{\infty} K(u) du = 1 \quad (2.51)$$

$$3. \int_{-\infty}^{\infty} K^2(u) du < \infty \quad (2.52)$$

$$4. K(u) = K(-u) \quad (2.53)$$

Parzen, Kernel fonksiyonları üzerinde önemli çalışmalar gerçekleştirmiştir. Çalışmasında $f(x)$ fonksiyonunu tahminleyenlerinden oluşan bir sınıfın tutarlılık ve yansızlık özelliklerini araştırmıştır. Kernel regresyon tahmininin en önemlisi ise Nadarya- Watson yöntemidir [53].

2.11. Nadaraya-Watson (NW) Yöntemi

NW tahmincisinin özelliği rasgele bir u değerinin komşuluğunda bağımsız değişken olan x_i değerine karşılık gelen bağımlı değişken y_i değerlerinin ağırlıklı ortalaması olarak söylenebilir. Tahminci bu ağırlıklarla x değerine daha uzak olan x_i değeri için küçük, u değerine daha yakın olan u_i değeri için büyük olmalıdır. NW tahmincisi ile elde edilen regresyon eğrisinin tahmini aşağıdaki şekildedir`;

$$\hat{m}(x) = \sum_{i=1}^n w_{ni}(x)y_i \quad [54]. \quad (2.54)$$

Parametrik regresyonda fonksiyon form $m(x)$ 'in şartlı ortalaması olarak belirlenmektedir. Bu form regresyon eğri $m(x, \beta)$ 'nın belirli sayıdaki bilinmeyen β parametresine bağıllığını anlatmaktadır. İlgili duruma ait EKK tahmini aşağıdaki gibi gösterilebilir.[55,56]

$$\sum_{i=1}^n (y_i - m(x_i - \hat{\beta}))^2 \quad (2.55)$$

NW tahmincisine ait EKK tahmini ise aşağıdaki şekilde ifade edilebilir;

$$\sum_{i=1}^n w_{ni}[y_i - m(x)]^2 \quad (2.56)$$

Tahmincinin performansını etkileyen bant genişliğinin seçimi için farklı metotlar vardır. Bunlardan bazıları çapraz-geçerlilik, ceza fonksiyonları, bootstrap, genelleştirilmiş çapraz-geçerlilik metotlarıdır. Bu yöntemlerden hangisinin daha etkili olduğu konusunda ise fikir ayrılıkları mevcuttur. İçlerinden daha fazla kullanılabilecek olan ise genelleştirilmiş çapraz-geçerlilik metodudur [57].

2.12. Klasik ve Uygulanabilir Nadaraya-Watson Çekirdek Kestiricileri

Tahminlerin daha kolay ve uygulanabilir olmasından dolayı bu tahmin ediciler içerisinde en sık kullanılan çekirdek fonksiyonları NW tahmin edicileridir. Bu metot da tüm tahminler için aynı bant genişliği kullanılmaktadır. Bu metot bazı durumlarda olumlu sonuçlar vermeye bilir. Veriler sık olduğu vakit düzleştirme yapan bu yöntem, veriler seyrek olduğunda ise yeterli düzleştirmeyi yapamamaktadır [58,59]. Ayrıca bant genişliğinin fazla olduğu durumlarda fazla düzleştirme

yapmakta ve dağılımın önemli noktalarını yok edebilmektedir, bant aralığı normalden küçük olduğunda ise gerekli düzleştirmeleri yapamamakta ve yine başarısız olmaktadır [60].

Bu problemler göz önüne alınarak klasik yönteme karşılık alternatif yöntemler ortaya çıkmıştır.

1. Sabit bant genişliği kullanılan klasik Nadaraya-Watson çekirdek kestiricisi;

X ve Y , $f(x, y)$ bileşik olasılık yoğunluk fonksiyonuna sahip iki rasgele değişken olsun. Y 'nin verilen $X = x$ e göre koşullu beklenen değeri aşağıdaki şekildedir;

$$E(Y|X = x) = \int y f(y|x) dy = \int y \frac{f(x, y)}{f_x(x)} dy = m(x) \quad (2.57)$$

Burada $f(y|x)$, Y 'nin, verilen $X = x$ e göre koşullu olasılık yoğunluk fonksiyonu, $f_x(x)$ ise X 'in marjinal olasılık yoğunluk fonksiyonudur.

Veri kümesindeki her bir x_i için koşullu beklenen değer $m(x_i)$ elde edilir ve toplam n adet ($i = 1, \dots, n$) değerden oluşan koşullu beklenen değerler kümesi oluşturulur. Böylece Y ve X in ortalama olarak nasıl bir etkileşim içerisinde oldukları ortaya koyulmuş olunur. Bu nedenle regresyonda ilgilenilen temel nokta aşağıda verilen $m(x)$ ' in tahminidir [61].

$$m(x) = E(Y|X = x) = \int y \frac{f(x, y)}{f_x(x)} dy = \frac{\int y f(x, y) dy}{f_x(x)} \quad (2.58)$$

$\{X_i, Y_i\}, (i = 1, \dots, n)$ şeklinde verilen gözlemler için, üsteki fonksiyonda $f(x, y)$ ve $f_x(x)$ bilinmemektedir. Bir değişkenin olasılık yoğunluk fonksiyonu olmasından ötürü $f_x(x)$ ' in tahmini kolaydır. İki değişken X ve Y 'nin bileşik olasılık yoğunluk fonksiyonu olan $f(x, y)$ ' nin tahmin değeri için “ çekirdekler ile çekirdek yoğunluk fonksiyonu” özelliği kullanılmaktadır. h ve g sırasıyla X ve Y değişkenlerinin yoğunluklarının tahmininde kullanılan bant genişlikleri olmak üzere, bu özellikler göz önüne alınarak elde edilen yoğunluk fonksiyonu tahmini aşağıdaki biçimde gösterilebilir [62].

$$\hat{f}_{h,g}(x, y) = \frac{1}{n} \sum_{i=1}^n K_h\left(\frac{x - X_i}{h}\right) K_g\left(\frac{y - Y_i}{g}\right) \quad (2.59)$$

Çekirdek fonksiyonların integralinin 1' i verdiği, 0 (orjin) etrafında simetrik olduğu bilgisinden faydalanılarak veri kümesindeki herhangi bir x e karşılık gelen Y değerinin NW tahmini aşağıdaki gibi yazılabilir;

$$\hat{m}_h(x) = \frac{n^{-1} \sum_{i=1}^n K_h(x - X_i) Y_i}{n^{-1} \sum_{i=1}^n K_h(x - X_i)} \quad (2.60)$$

Burada bütün tahminler için bant genişliği aynı olarak h kullanıldığından dolayı metot klasik NW tahmin metodu olarak ifade edilmektedir.

2.13. Değişen Bant Genişliğini Kullanan Uyarlanabilir Nadaraya–Watson Çekirdek Kestiricileri:

Daha evvelden de belirtildiği gibi parametrik olmayan regresyon fonksiyonlarının tahmini genellikle koşullu ortalama fonksiyonu üzerine kuruludur. NW pay ve paydadaki olasılık yoğunluk fonksiyonlarının tahminde çekirdek kestirimlerinin kullanılmasını önermiştir. İki değişken durumunda Epanechnikov tarafından bileşik olasılık yoğunluk fonksiyonunun çekirdek kestirimi;

$$\hat{f}(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_x h_y} \check{K} \left(\frac{x - X_i}{h_x}, \frac{y - Y_i}{h_y} \right) \quad (2.61)$$

Biçiminde verilmiştir[63]. Burada $\check{K}(\dots)$ iki değişkenli çekirdek fonksiyonunu göstermekte çarpımsal çekirdek fonksiyonu kullanılarak aşağıdaki gibi elde edilir;

$$\check{K} \left(\frac{x - X_i}{h_x}, \frac{y - Y_i}{h_y} \right) = K_1 \left(\frac{x - X_i}{h_x} \right) K_2 \left(\frac{y - Y_i}{h_y} \right) \quad (2.62)$$

$K_1 = K_2 = K$ olması durumunda iki değişkenli olasılık yoğunluk fonksiyonunun çekirdek kestiricisi aşağıdaki gibi elde edilir;

$$\hat{f}(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_x h_y} K \left(\frac{x - X_i}{h_x} \right) K \left(\frac{y - Y_i}{h_y} \right) \quad (2.63)$$

Bu fonksiyonel denklem NW çekirdek kestiricisinin elde edilmesinde kullanılmaktadır.

n gözlemleri $T_1, \dots, T_q$ (q boyutlu durumda) değişkenleri için Sain tarafından, sabit bant genişliği yerine değişen bant genişliği kullanılarak çok değişkenli olasılık yoğunluk fonksiyonunun çarpımsal çekirdek kestiricisi aşağıdaki gibidir;

$$\hat{f}_u(t_1, \dots, t_q) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h(T_{1i}) \dots h(T_{qi})} K \left(\frac{t_1 - T_{1i}}{h(T_{1i})} \right) \dots K \left(\frac{t_q - T_{qi}}{h(T_{qi})} \right) \quad (2.64)$$

İki değişken x ve y için (2.64) eşitliği aşağıdaki şekle dönüşür;

$$\hat{f}_u(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h(X_i) h(Y_i)} K \left(\frac{x - X_i}{h(X_i)} \right) K \left(\frac{y - Y_i}{h(Y_i)} \right) \quad (2.65)$$

Eşitlik (2.63)' de paydadaki kestirici yerine;

$$\hat{f}_u(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h(X_i)} K \left(\frac{x - X_i}{h(X_i)} \right) \quad (2.66)$$

Biçiminde tanımlanan yoğunluk fonksiyonunun uyarlanabilir çekirdek kestirici ve paydadaki kestirici yerine, (2.64) eşitliğinde yerine koyulduğunda verilen regresyon fonksiyonunun çekirdek kestiricisi elde edilmektedir.



3. BULGULAR VE TARTIŞMA

Covid 19 virüsü 2019'un son çeyreğinde Çin'in Vuhan eyaletinde ilk olarak görülmüş kısa sürede hızla yayılarak bütün dünyayı etkisi altına almıştır. Türkiye de de etkisi görülen virüs bir çok ölüme sebep olmuştur. Bizde bu çalışmamızda Türkiye de meydana gelen vakaları İstatistiksel olarak inceledik. Yapılan analizlerde Parametrik olmayan Regresyon analizi yöntemlerinden faydalanılmış, çıkan sonuçlar değerlendirilerek önerilerde bulunulmuştur.

3.1. COVID-19 İle İlgili Genel Bilgi

Yeni korona virüs hastalığı (COVID-19), ilk olarak Çin'in Vuhan eyaletinde aralık ayının sonlarında, ateş, öksürük, nefes darlığı gibi solunum yolu belirtileri gelişen bir grup hastada yapılan araştırmalar sonucunda 13 Ocak 2020'de tanımlanan bir virüstür. Hastalığın ilk olarak bu bölgedeki deniz ürünleri ve hayvan pazarında bulunan insanlarda görülmesi hastalığın pazarda satılan bu tip hayvanlardan insanlara bulaştığı kanısını ortaya koymuştur. Ancak daha sonra insandan insana bulaşarak Vuhan başta olmak üzere Hubei eyaletindeki diğer şehirlere ve Çin Halk Cumhuriyeti'nin diğer eyaletlerine ve diğer dünya ülkelerine yayılmıştır [64].

Korona virüsler, hayvanlarda veya insanlarda hastalığa neden olabilecek büyük bir virüs ailesidir. İnsanlarda, birkaç korona virüsün soğuk algınlığından Orta Doğu Solunum Sendromu (MERS) ve Şiddetli Akut Solunum Sendromu (SARS) gibi daha şiddetli hastalıklara kadar solunum yolu rahatsızlıklarına neden olduğu bilinmektedir. Yeni korona virüs hastalığına SAR-CoV-2 virüsü neden olmaktadır.

Bazı insanlarda belirtisiz olgular olabileceği bildirilmekle birlikte, bunların oranı tam olarak bilinmemektedir. En çok karşılaşılan belirtiler ateş, öksürük ve nefes darlığıdır. Şiddetli olgularda zatürre, ağır solunum yetmezliği, böbrek yetmezliği ve ölüm gelişebilmektedir.

Hasta bireylerin öksürmeleri ile ortama saçılan damlacıkların solunması ile bulaşmaktadır. Hastaların solunum parçacıkları ile kirlenmiş yüzeylere dokunulduktan sonra ellerin yıkanmadan yüz, göz, burun veya ağıza götürülmesi ile de virüs alınabilmektedir. Kirli ellerle göz, burun veya ağıza temas etmek oldukça risklidir.

COVID-19 rahatsızlığı ile ilgili şimdiye kadar edinilen bilgiler, bazı insanların daha fazla hastalanma ve ciddi belirtiler gösterme riski altında olduğunu göstermiştir. Yapılan istatistiklerde vakaların % 80'i hastalığı hafif geçirmekte olup %20'si hastane koşullarında tedavi edilmektedir. Hastalık, genellikle 60 yaş ve üzerindeki kişileri daha fazla etkilemektedir.

Hastalıktan en çok etkilenen kişiler, 60 yaş üstü olanlar ile kalp rahatsızlığı, yüksek tansiyon, diyabet, kronik solunum yolu rahatsızlığı ve kanser gibi ciddi kronik tıbbi hastalıkları olan insanlardır.

Çocuklarda hastalık nadiren ve oldukça hafif görünmektedir. COVID-19 rahatsızlığı gelişen gebe kadınlarda hastalığın ciddiyeti konusunda sınırlı bilimsel bilgi vardır.

Bununla birlikte mevcut kanıtlar COVID-19 rahatsızlığı sonrası hamile kadınlar arasındaki hastalık şiddetinin, hamile olmayan yetişkin COVID-19 vakalarına benzer olduğunu ve hamilelik sırasında COVID-19 ile rahatsızlığı fetüste olumsuz bir etkisi olduğunu gösteren hiçbir veri olmadığını göstermektedir. Şu ana kadar COVID-19'un hamilelik sırasında anneden bebeğe bulaştığına dair de herhangi bir kanıt bulunmamaktadır.

Yeni Korona virüs tanısı için gerekli moleküler testler ülkemizde mevcuttur. Tanı testi sadece Halk Sağlığı Genel Müdürlüğü Ulusal Viroloji Referans Laboratuvarında ve belirlenmiş Halk Sağlığı Laboratuvarlarında yapılmaktadır.

Mümkün olduğu kadar yurtdışına yolculuk yapılmaması hastalığın yayılmasının önlenmesi için şiddetle önerilmektedir. Yurtdışına çıkışın zorunlu olduğu durumlarda ise bazı kurallara dikkat edilmesi hastalığın bulaşmaması için önerilmektedir. Akut solunum yolu rahatsızlıklarının genel bulaşma riskini azaltmak için önerilen temel ilkeler Yeni Korona virüs Hastalığı (COVID-19) için de geçerlidir.

Öncelikle el temizliğine azami derecede dikkat edilmelidir. Eller en az 20 saniye boyunca sabun ve suyla yıkanmalı, sabun ve suyun olmadığı durumlarda alkol temelli el antiseptiği kullanılmalıdır. Antiseptik veya anti-bakteriyel içeren sabun kullanmaya gerek yoktur, normal sabun yeterli olmaktadır. Eller yıkanmadan ağız, burun ve gözlerle temas edilmemeli, hasta insanlarla temastan kaçınılmalıdır. Özellikle hasta insanlarla veya çevreleriyle doğrudan temas ettikten sonra eller sık sık yıkanmalıdır.

Öksürme veya hapşıрма sırasında burun ve ağız tek kullanımlık kâğıt mendil ile örtülmeli, kâğıt mendilin bulunmadığı durumlarda ise dirsek içi kullanılmalı, mümkünse kalabalık yerlere girilmemeli, eğer girmek zorunda kalınıyorsa ağız ve burun kapatılmalı, tıbbi maske kullanılmalıdır.

3.2. COVID-19 Türkiye Verilerinin Bernstein Polinomu Yardımıyla İncelenmesi

Türkiye de 11 Mart 2020 ile 04 Haziran 2020 tarihleri arasında gözlenen 86 günlük yeni vaka sayısı incelenmiştir. Virüse bağlı olarak gelişen vakalarda meydana gelen ölüm oranları bu oranların tahmin değerleri ve hata değerleri incelenmiştir. Tablo3.1’de günlere bağlı olarak gelişen yeni vaka sayıları verilmiştir.

Tablo 3.1. 11 Mart 2020 ile 4 Haziran 2020 tarihleri arasında gözlenen yeni vaka sayısı

Sıra No:	Yeni vaka sayısı:	Sıra No:	Yeni vaka sayısı:	Sıra No:	Yeni vaka sayısı:	Sıra No:	Yeni vaka sayısı:	Sıra No:	Yeni vaka sayısı:	Sıra No:	Yeni vaka sayısı:
1	1	17	2.069	33	4.789	49	2.392	65	1.635	81	983
2	0	18	1.704	34	4.093	50	2.936	66	1.708	82	839
3	4	19	1.815	35	4.062	51	2.615	67	1.610	83	827
4	1	20	1.610	36	4.281	52	2.188	68	1.368	84	786
5	12	21	2.704	37	4.801	53	1.983	69	1.158	85	867
6	29	22	2.148	38	4.353	54	1.670	70	1.022	86	988
7	51	23	2.456	39	3.783	55	1.614	71	972		
8	93	24	2.786	40	3.977	56	1.832	72	961		
9	168	25	3.013	41	4.674	57	2.253	73	952		
10	311	26	3.135	42	4.611	58	1.977	74	1.186		
11	277	27	3.148	43	3.083	59	1.848	75	1.141		
12	289	28	3.892	44	3.116	60	1.546	76	987		
13	293	29	4.117	45	3.122	61	1.542	77	948		
14	343	30	4.056	46	2.861	62	1.114	78	1.035		
15	561	31	4.747	47	2.357	63	1.704	79	1.182		
16	1.196	32	5.138	48	2.131	64	1.639	80	1.141		

Üçüncü bölümde parametrik olmayan regresyon modellerinden bahsedilirken 2.3 alt bölümünde Denklem 2.45’de verilen Bernstein operatörü kullanılarak,300 iterasyon, sonunda aşağıdaki tabloda bulunan tahmin değerleri elde edilmiştir.

Vaka sayılarının tahmin yöntemleri ile elde edilen değerleri Tablo 3.2’de verilmiştir.

Tablo 3.2. Vaka sayılarının Bernstein polinomu ile elde edilen tahmin değerleri

Sıra No:	Yeni vaka sayısının tahmini	Sıra No:	Yeni vaka sayısının tahmini	Sıra No:	Yeni vaka sayısının tahmini	Sıra No:	Yeni vaka sayısının tahmini	Sıra No:	Yeni vaka sayısının tahmini	Sıra No:	Yeni vaka sayısının tahmini
1	1	17	2069,03	33	4789,016	49	2391,973	65	1634,858	81	982,651
2	-0,003	18	1704,03	34	4093,015	50	2935,992	66	1707,932	82	838,702
3	4,003	19	1815,01	35	4062,008	51	2615,019	67	1610,074	83	827,601
4	1,021	20	1609,97	36	4280,998	52	2188,041	68	1368,183	84	785,622
5	11,94	21	2703,96	37	4800,989	53	1983,046	69	1158,155	85	867,103
6	29,01	22	2147,97	38	4352,984	54	1670,025	70	1021,980	86	988
7	51,05	23	2456,00	39	3782,985	55	1613,986	71	971,7909		
8	92,99	24	2786,02	40	3976,991	56	1831,946	72	960,7742		
9	167,9	25	3013,027	41	4674,001	57	2252,929	73	951,9921		
10	310,9	26	3135,015	42	4611,012	58	1976,952	74	1186,257		
11	277,0	27	3147,997	43	3083,020	59	1848,009	75	1141,262		
12	289,0	28	3891,983	44	3116,019	60	1546,072	76	986,9345		
13	292,9	29	4116,979	45	3122,010	61	1542,102	77	947,6420		
14	342,9	30	4055,985	46	2860,994	62	1114,068	78	1034,835		
15	560,9	31	4746,997	47	2356,977	63	1703,980	79	1182,329		
16	1195	32	5138,009	48	2130,968	64	1638,887	80	1141,305		

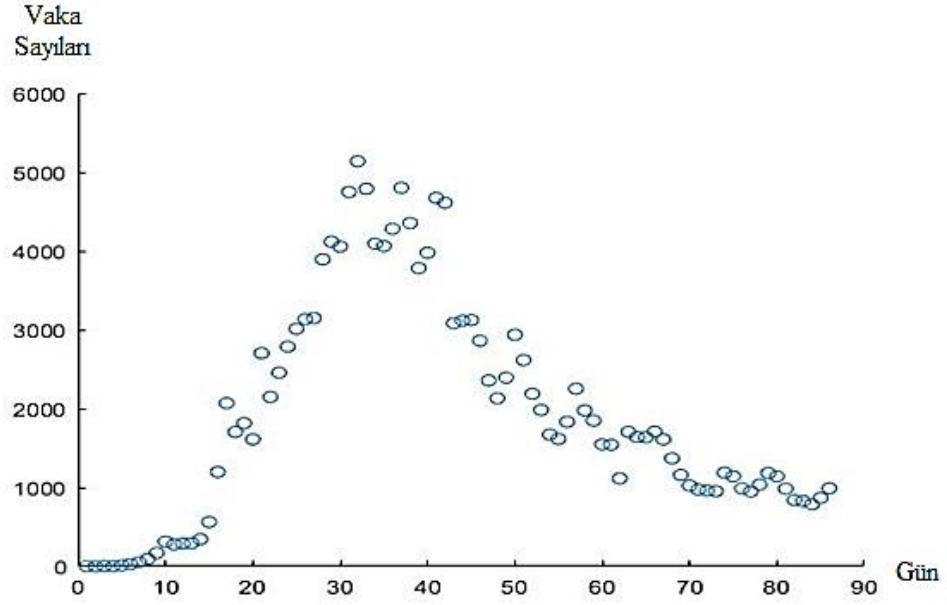
Tablo 3.2’de vaka sayıları göz önünde bulundurularak yeni oluşa bilecek vakaların tahmin değerlerine yer verilmiştir. Kullanılan analizler yardımıyla verilere bakıldığında ilk vakanın görüldüğü tarih itibarıyla tahmini en yüksek vakaya 32. Gün itibarı ile ulaşılacağı düşünülmektedir. Bernstein operatörü kullanılarak elde edilen tahmin değerlerinin hataları Tablo 3.3’de verilmiştir.

Tablo 3.3. Hata değerleri

Sıra No:	Yeni vaka sayısı:	Sıra No:	Yeni vaka sayısı:	Sıra No:	Yeni vaka sayısı:	Sıra No:	Yeni vaka sayısı:	Sıra No:	Yeni vaka sayısı:	Sıra No:	Yeni vaka sayısı:
1	0	17	0,0367	33	0,0161	49	-0,0263	65	-0,1411	81	-0,34
2	-0,003	18	0,0397	34	0,0153	50	-0,0071	66	-0,0675	82	-0,29
3	0,003	19	0,0117	35	0,0084	51	0,01975	67	0,074919	83	0,601661
4	0,021	20	-0,021	36	-0,001	52	0,04193	68	0,183964	84	-0,37761
5	-0,050	21	-0,035	37	-0,010	53	0,04606	69	0,155452	85	0,103996
6	0,0121	22	-0,023	38	-0,015	54	0,02549	70	-0,01923	86	0
7	0,0535	23	0,0014	39	-0,014	55	-0,0139	71	-0,20903		
8	-0,007	24	0,0222	40	-0,008	56	-0,0538	72	-0,22577		
9	-0,056	25	0,0270	41	0,0019	57	-0,0703	73	-0,00785		
10	-0,020	26	0,0159	42	0,0128	58	-0,0478	74	0,257542		
11	0,042	27	-0,002	43	0,0200	59	0,00934	75	0,262818		
12	0,049	28	-0,016	44	0,0196	60	0,07280	76	-0,06543		
13	-3,76E-05	29	-0,020	45	0,0102	61	0,10211	77	-0,35797		
14	-0,045	30	-0,014	46	-0,005	62	0,06878	78	-0,16437		
15	-0,041	31	-0,002	47	-0,022	63	-0,0194	79	0,329075		
16	-0,0006	32	0,0097	48	-0,031	64	-0,1122	80	0,305025		

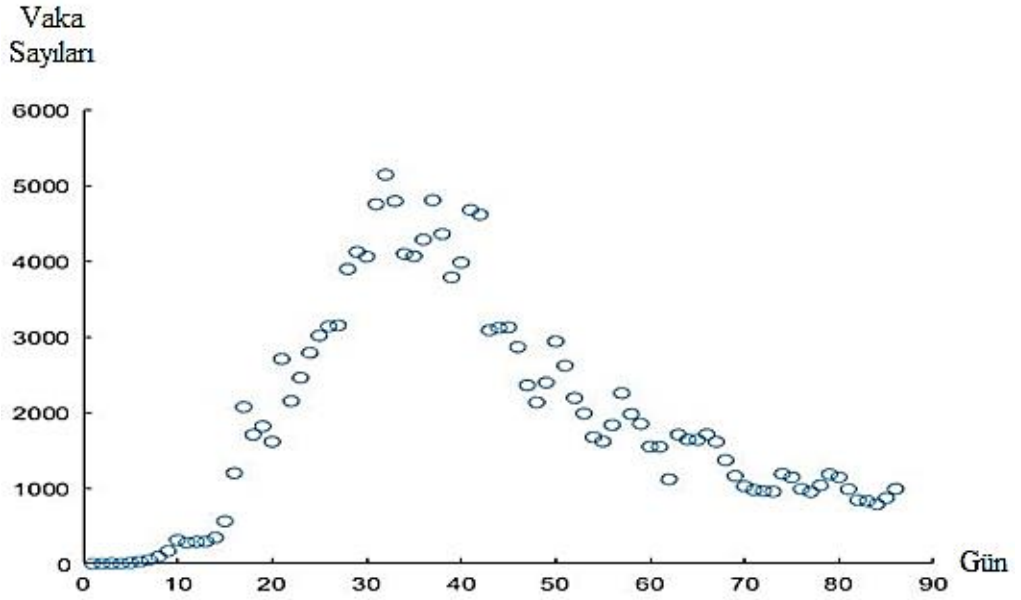
Tablo 3.3’de ise Bernstein tahminleyicisi kullanılmış ve daha önceki analizlerde elde edilmiş olunan yeni vaka tahmin değerleri üzerindeki hata payları ölçülerek tabloda gösterilmiştir.

Şekil 3.1’de vaka sayılarının günlere göre değişim grafiği verilmiştir;



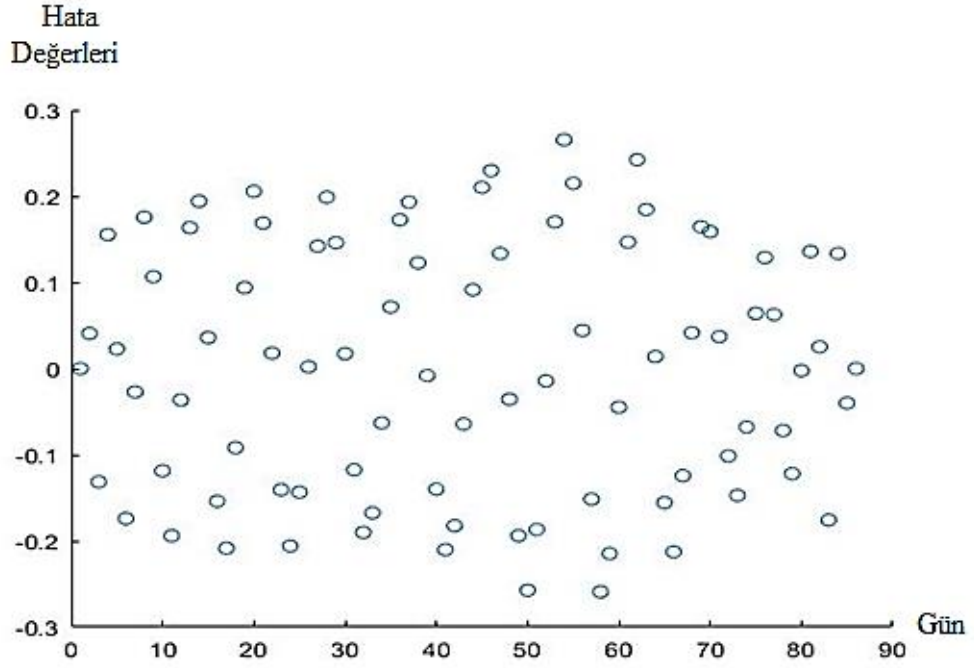
Şekil 3.1. Yeni vaka sayısının günlere göre grafiği

Şekil 3.2’de vaka sayılarının tahmini dağılımı grafik şeklinde verilmiştir;



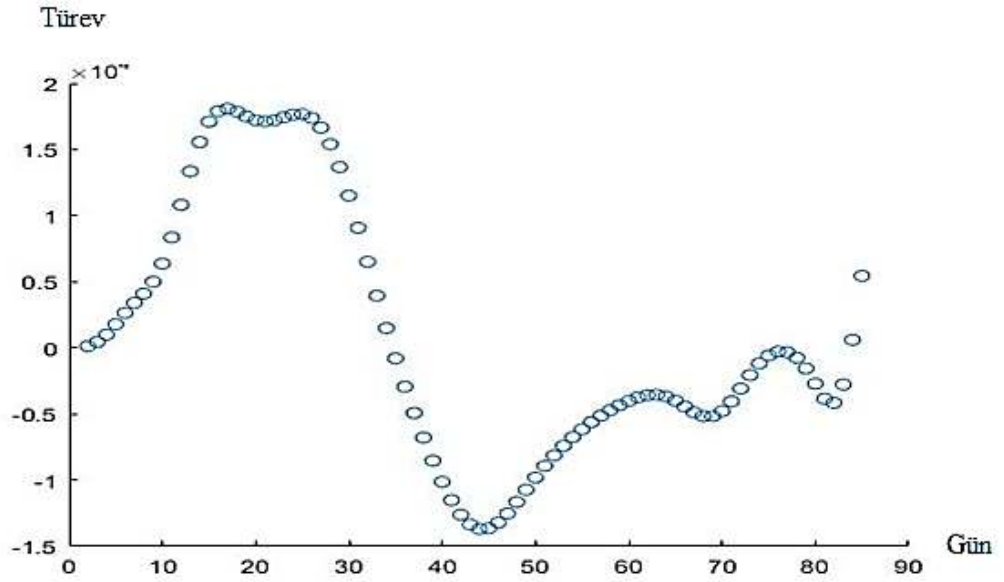
Şekil 3.2. Yeni vaka sayısının tahmin değerlerinin grafiği

Yeni vaka tahmin değerlerinin günlere göre değişimi Şekil 3.3’ deki gibidir;



Şekil 3.3. Hata değerleri grafiği

Tahmin eğrisinin 1.türevinin zamana bağlı değişimini gösteren grafik Şekil3.4’ deki gibidir;



Şekil 3.4. Tahmin eğrisinin 1. türevinin grafiği

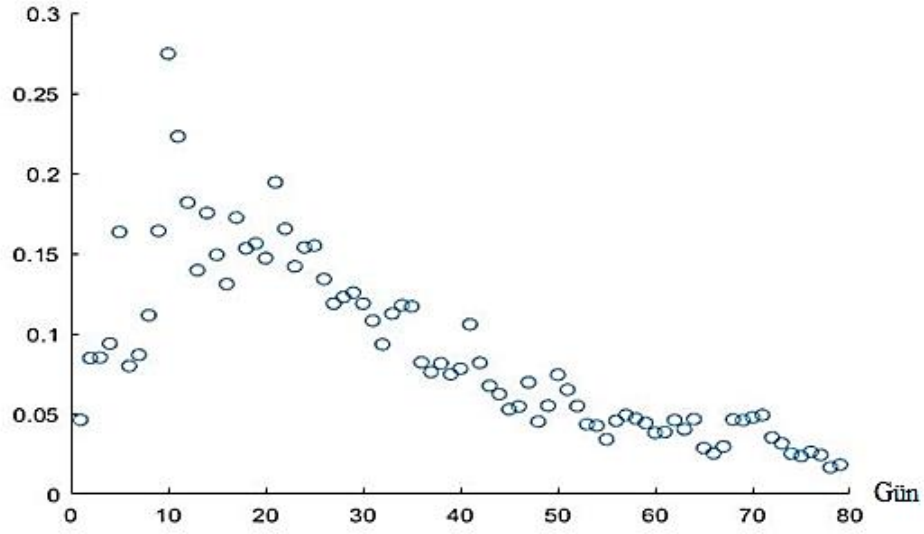
Vaka sayılarının tespiti korona virüs rahatsızlığından dolayı bireyin sağlık kurumuna müracaatı ve hastalık riski bulunan kişilere korona virüs testi uygulamakla yapılmaktadır. Bundan dolayı vaka sayısında gözlenen değişiklik büyük ölçüde yapılan test sayısına bağlı olarak değişiklik göstermektedir. Bu durum göz önünde bulundurulduğunda vaka sayılarının değişim eğrisinin yanında yapılan test sayıları da dikkate alınarak (yeni vaka sayısı / yapılan test sayısı) oranının değişimini gösteren eğrinin de grafiğinin tahmin edilmesi oldukça önemlidir. Bu bakımdan Tablo 3.4’de yeni vaka sayısının test sayısına oranları verilmiştir.

Tablo 3.4. 11 Mart 2020 ile 4 Haziran 2020 tarihleri arasında gözlenen yeni vaka sayısının test sayısına oranı

Sıra No:	YK/TS	Sıra No:	YK/TS	Sıra No:	YK/TS	Sıra No:	YK/TS	Sıra No:	YK/TS	Sıra No:	YK/TS
1	0	17	0,130938	33	0,093361	49	0,04512	65	0,046646	81	*
2	0,046154	18	0,172401	34	0,112523	50	0,055043	66	0,028573	82	*
3	0,084806	19	0,153224	35	0,117724	51	0,074349	67	0,025382	83	*
4	0,085066	20	0,156242	36	0,116944	52	0,065044	68	0,029519	84	*
5	0,093803	21	0,147103	37	0,082137	53	0,054858	69	0,046403	85	*
6	0,163462	22	0,194376	38	0,076071	54	0,043421	70	0,045924	86	*
7	0,079793	23	0,165341	39	0,081406	55	0,042612	71	0,047751		
8	0,086791	24	0,141927	40	0,074684	56	0,034044	72	0,049185		
9	0,11142	25	0,153804	41	0,078106	57	0,045621	73	0,035222		
10	0,16415	26	0,154899	42	0,105794	58	0,049172	74	0,031559		
11	0,274658	27	0,134071	43	0,081834	59	0,046954	75	0,025057		
12	0,223007	28	0,118789	44	0,067497	60	0,044289	76	0,023567		
13	0,181827	29	0,12283	45	0,062256	61	0,038119	77	0,026233		
14	0,139575	30	0,125579	46	0,052811	62	0,038678	78	0,024316		
15	0,175334	31	0,118757	47	0,054601	63	0,04606	79	0,016576		
16	0,149198	32	0,108095	48	0,06958	64	0,040265	80	0,018217		

Yeni vaka sayısının test sayısına olan oranının günlere göre değişim grafiği Şekil3.5’de verilmiştir;

Yeni Vaka / Test Sayısı



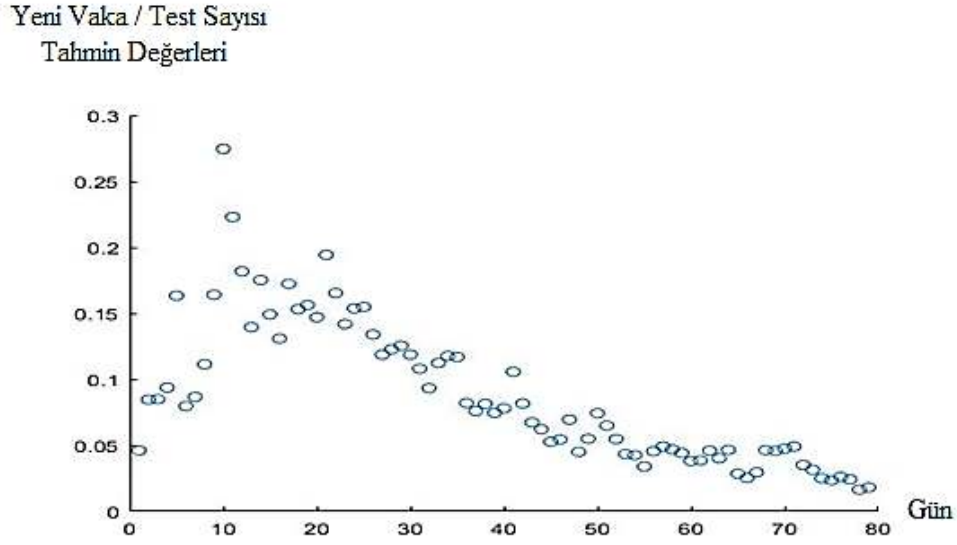
Şekil 3.5. Yeni Vaka / Test Sayısı oranının günlere göre grafiği

Yeni vaka sayısının test sayısına oranı Tablo3.5’de verilmiştir. Bu bağlamda yeni vakanın test sayısına göre değişimi görüle bilinir.

Tablo 3.5. Yeni vaka sayısının test sayısına oranının tahmin değerleri

Sıra No:	YV/TS tahmin	Sıra No:	YV/TS tahmin	Sıra No:	YV/TS tahmin	Sıra No:	YV/TS tahmin	Sıra No:	YV/TS tahmin	Sıra No:	YV/TS tahmin
1	Tahmin	17	0,130928	33	0,093359	49	0,045119	65	0,046639	81	*
2	0,046154	18	0,172398	34	0,112518	50	0,055036	66	0,028569	82	*
3	0,084812	19	0,153229	35	0,117719	51	0,074341	67	0,025385	83	*
4	0,085041	20	0,156249	36	0,116942	52	0,065039	68	0,029525	84	*
5	0,093842	21	0,147107	37	0,082139	53	0,05486	69	0,046406	85	*
6	0,163441	22	0,194376	38	0,076077	54	0,043428	70	0,045921	86	*
7	0,079774	23	0,165338	39	0,081413	55	0,04262	71	0,047746		
8	0,08681	24	0,141924	40	0,074688	56	0,034048	72	0,049185		
9	0,111438	25	0,153802	41	0,078104	57	0,045618	73	0,035226		
10	0,164142	26	0,154898	42	0,105787	58	0,049164	74	0,03156		
11	0,27464	27	0,13407	43	0,081826	59	0,046948	75	0,025053		
12	0,223003	28	0,11879	44	0,067493	60	0,044288	76	0,023567		
13	0,181839	29	0,122832	45	0,062258	61	0,038125	77	0,026237		
14	0,139588	30	0,125582	46	0,052818	62	0,038685	78	0,024313		
15	0,175335	31	0,11876	47	0,054609	63	0,046063	79	0,016577		
16	0,149189	32	0,108097	48	0,069585	64	0,040261	80	0,018217		

Şekil3.6’da ise yeni vaka tahmin değerinin test sayısı tahmin değerlerine oranının zamana bağlı değişiminin grafiği verilmiştir;



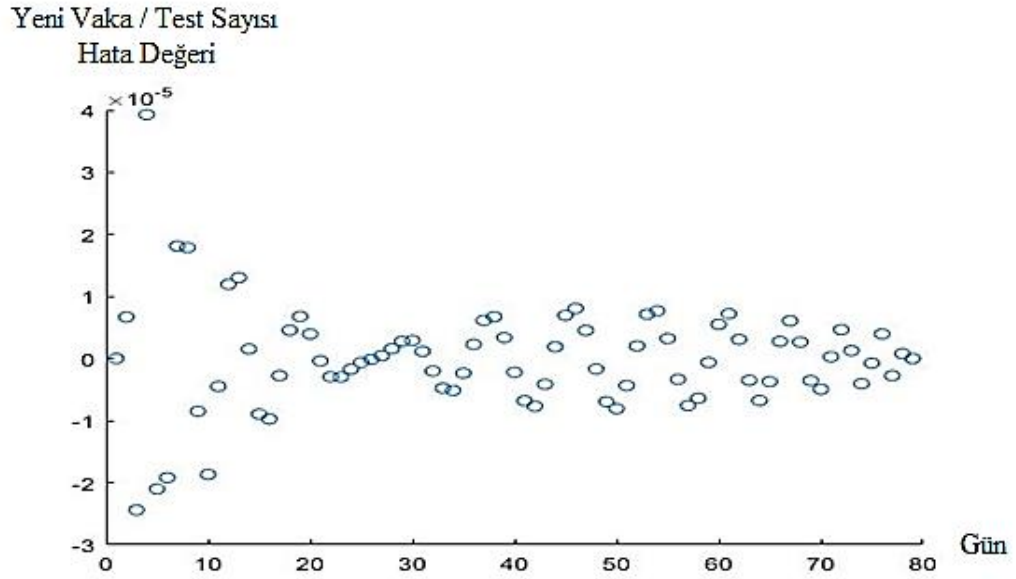
Şekil 3.6. Yeni Vaka / Test Sayısı oranının tahmin değerleri grafiği

Yeni vaka sayısının test sayısına oranın tahminlerinin hata değerleri Tablo3.6’da verilmiştir.

Tablo 3.6. Yeni vaka sayısının test sayısına oranının tahminlerinin hata değerleri

Sıra No:	Hata	Sıra No:	Hata	Sıra No:	Hata	Sıra No:	Hata	Sıra No:	Hata	Sıra No:	Hata
1	0	17	-9,74E-06	33	-2,01E-06	49	-1,69E-06	65	-6,77E-06	81	*
2	0	18	-2,75E-06	34	-4,78E-06	50	-6,94E-06	66	-3,70E-06	82	*
3	6,70E-06	19	4,55E-06	35	-5,14E-06	51	-8,08E-06	67	2,76E-06	83	*
4	-2,44E-05	20	6,78E-06	36	-2,36E-06	52	-4,35E-06	68	6,07E-06	84	*
5	3,93E-05	21	3,95E-06	37	2,29E-06	53	2,04E-06	69	2,66E-06	85	*
6	-2,10E-05	22	-3,80E-07	38	6,15E-06	54	7,14E-06	70	-3,55E-06	86	*
7	-1,92E-05	23	-2,94E-06	39	6,72E-06	55	7,68E-06	71	-4,95E-06		
8	1,81E-05	24	-2,97E-06	40	3,36E-06	56	3,23E-06	72	2,91E-07		
9	1,79E-05	25	-1,71E-06	41	-2,21E-06	57	-3,34E-06	73	4,66E-06		
10	-8,50E-06	26	-6,46E-07	42	-6,81E-06	58	-7,56E-06	74	1,27E-06		
11	-1,87E-05	27	-1,27E-07	43	-7,68E-06	59	-6,40E-06	75	-4,05E-06		
12	-4,48E-06	28	4,50E-07	44	-4,13E-06	60	-6,29E-07	76	-7,42E-07		
13	1,20E-05	29	1,57E-06	45	1,89E-06	61	5,50E-06	77	3,96E-06		
14	1,30E-05	30	2,79E-06	46	6,95E-06	62	7,22E-06	78	-2,76E-06		
15	1,54E-06	31	2,92E-06	47	8,08E-06	63	3,08E-06	79	7,84E-07		
16	-8,94E-06	32	1,15E-06	48	4,53E-06	64	-3,50E-06	80	0		

Şekil3.7’de ise hata paylarının zamana balı olarak değışiminin grafiđi verilmiřtir;



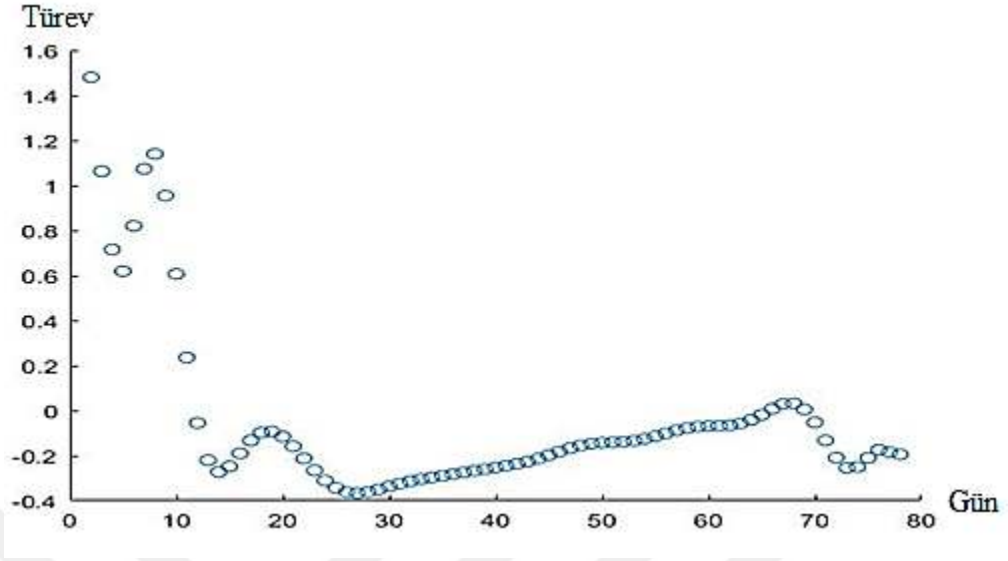
Şekil 3.7. Yeni Vaka / Test Sayısı oranının hata değeri grafiđi

Yeni vaka sayılarının test sayısına oranının hata terimlerinin türevleri Tablo3.7’de verilmiřtir.

Tablo 3.7. Yeni vaka sayısının test sayısına oranının tahminlerinin hata değeri türevleri

Sıra No:	Türev	Sıra No:	Türev	Sıra No:	Türev	Sıra No:	Türev	Sıra No:	Türev	Sıra No:	Türev
1	0	17	-0,18978	33	-0,31276	49	-0,15467	65	-0,04151	81	*
2	0	18	-0,13277	34	-0,30393	50	-0,14674	66	-0,01835	82	*
3	1,480113	19	-0,09767	35	-0,29658	51	-0,14255	67	0,008277	83	*
4	1,063039	20	-0,09252	36	-0,28971	52	-0,14052	68	0,029195	84	*
5	0,715406	21	-0,11555	37	-0,28257	53	-0,13834	69	0,031944	85	*
6	0,619414	22	-0,15892	38	-0,27496	54	-0,13379	70	0,005517	86	*
7	0,820232	23	-0,21224	39	-0,26714	55	-0,12564	71	-0,05288		
8	1,07375	24	-0,26545	40	-0,25952	56	-0,11409	72	-0,13238		
9	1,139937	25	-0,31063	41	-0,25222	57	-0,10072	73	-0,20876		
10	0,954665	26	-0,34303	42	-0,2448	58	-0,08785	74	-0,25355		
11	0,606837	27	-0,36111	43	-0,23634	59	-0,07767	75	-0,25018		
12	0,23515	28	-0,36612	44	-0,22585	60	-0,07136	76	-0,20989		
13	-0,05486	29	-0,36114	45	-0,21281	61	-0,06858	77	-0,17414		
14	-0,2206	30	-0,35008	46	-0,19759	62	-0,0674	78	-0,18508		
15	-0,2723	31	-0,33668	47	-0,18154	63	-0,0647	79	-0,19431		
16	-0,24775	32	-0,32374	48	-0,16661	64	-0,05695	80	0		

Şekil3.8’de ise tahmin eğrisinin türevinin zaman bağlı değişimi verilmiştir;



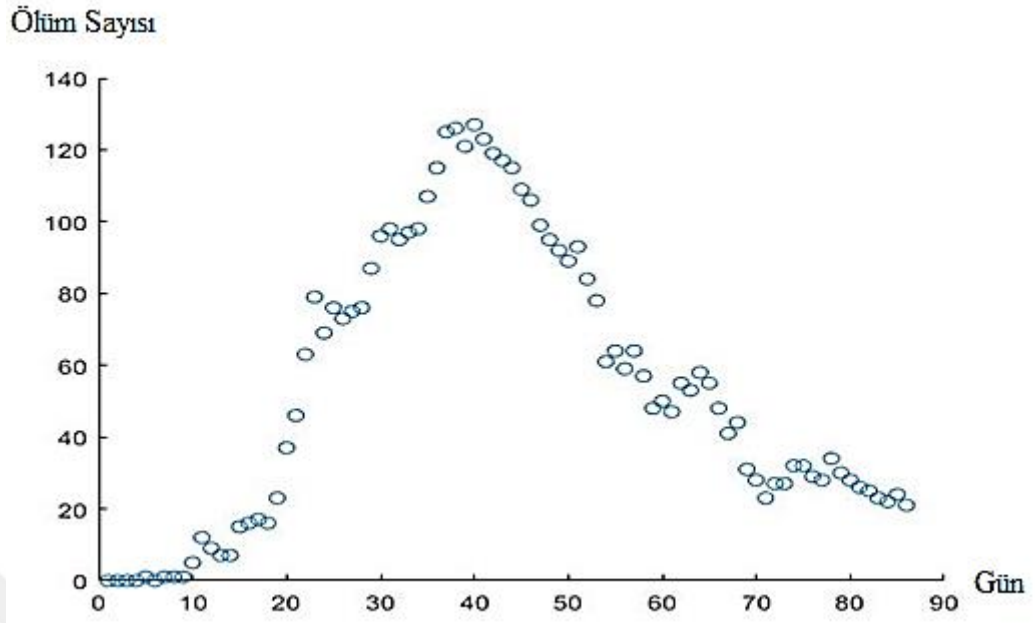
Şekil 3.8. Tahmin eğrisinin 1. türevinin grafiği

Korona virüs hastalığından gözlenen ölümlerin sayısı da hastalığın seyrinin incelenmesi için oldukça önemlidir. Bu nedenden dolayı gözlenen ölüm sayılarının da tahmin değerlerinin grafiğinin oluşturularak birinci türev incelemesi yapılmalıdır. Bu bakımdan gözlenen ölüm sayıları Tablo3.8’de ki gibidir.

Tablo 3.8. 11 Mart-4 Haziran 2020 tarihleri arasında gözlenen ölüm sayıları

Sıra No:	Ölüm sayısı	Sıra No:	Ölüm sayısı	Sıra No:	Ölüm sayısı:	Sıra No:	Ölüm sayısı	Sıra No:	Ölüm sayısı	Sıra No:	Ölüm sayısı
1	0	17	16	33	95	49	95	65	58	81	28
2	0	18	17	34	97	50	92	66	55	82	26
3	0	19	16	35	98	51	89	67	48	83	25
4	0	20	23	36	107	52	93	68	41	84	23
5	0	21	37	37	115	53	84	69	44	85	22
6	1	22	46	38	125	54	78	70	31	86	24
7	0	23	63	39	126	55	61	71	28		
8	1	24	79	40	121	56	64	72	23		
9	1	25	69	41	127	57	59	73	27		
10	1	26	76	42	123	58	64	74	27		
11	5	27	73	43	119	59	57	75	32		
12	12	28	75	44	117	60	48	76	32		
13	9	29	76	45	115	61	50	77	29		
14	7	30	87	46	109	62	47	78	28		
15	7	31	96	47	106	63	55	79	34		
16	15	32	98	48	99	64	53	80	30		

Vakalarda ölüm sayısının zamana bağlı değişiminin grafiği Şekil3.9’da verilmiştir;



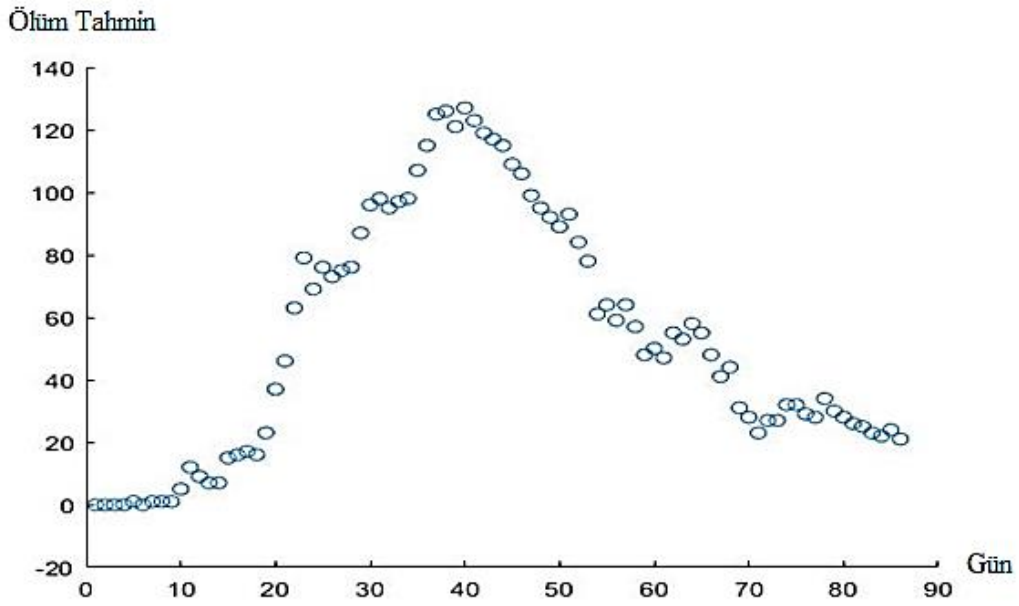
Şekil 3.9. Ölüm sayısının günlere göre grafiği

Ölüm sayılarının tahmin değerleri Tablo3.9’da verilmiştir.

Tablo 3.9. 11 Mart-4 Haziran 2020 tarihleri arasında gözlenen ölüm sayılarının tahmin değerleri gösteren tablo

Sıra No:	Tahmin	Sıra No:	Tahmin	Sıra No:	Tahmin	Sıra No:	Tahmin	Sıra No:	Tahmin	Sıra No:	Tahmin
1	0	17	15,99474	33	94,99898	49	94,9991	65	57,99951	81	28,00021
2	0	18	16,99771	34	97,00191	50	91,9992	66	54,99869	82	26,00208
3	0,000503	19	16,00264	35	98,00349	51	88,9995	67	47,99875	83	24,9998
4	-0,00201	20	23,00536	36	107,003	52	93,00001	68	40,99958	84	22,99778
5	0,003517	21	37,00383	37	115,0011	53	84,00063	69	44,0006	85	22,00194
6	0,998417	22	45,99943	38	124,9988	54	78,0011	70	31,00115	86	23,99938
7	-0,00316	23	62,99556	39	125,9975	55	61,00113	71	28,00094		
8	1,002139	24	78,99493	40	120,9976	56	64,00055	72	23,00013		
9	1,003844	25	68,99779	41	126,9989	57	58,99958	73	26,99925		
10	0,999153	26	76,00199	42	123,0004	58	63,99873	74	26,99886		
11	4,995415	27	73,00471	43	119,0014	59	56,99854	75	31,99934		
12	11,99778	28	75,00432	44	117,0016	60	47,9992	76	32,00051		
13	9,003048	29	76,00131	45	115,0011	61	50,00038	77	29,00139		
14	7,004963	30	86,99771	46	109,0003	62	47,00135	78	28,00083		
15	7,001559	31	95,99573	47	105,9996	63	55,00149	79	33,99905		
16	14,99651	32	97,99636	48	98,99922	64	53,00071	80	29,9983		

Şekil 3.10’da zamana bağlı olarak vakalardaki ölüm sayısının tahmin değerlerinin zamana bağlı değişimi verilmiştir;



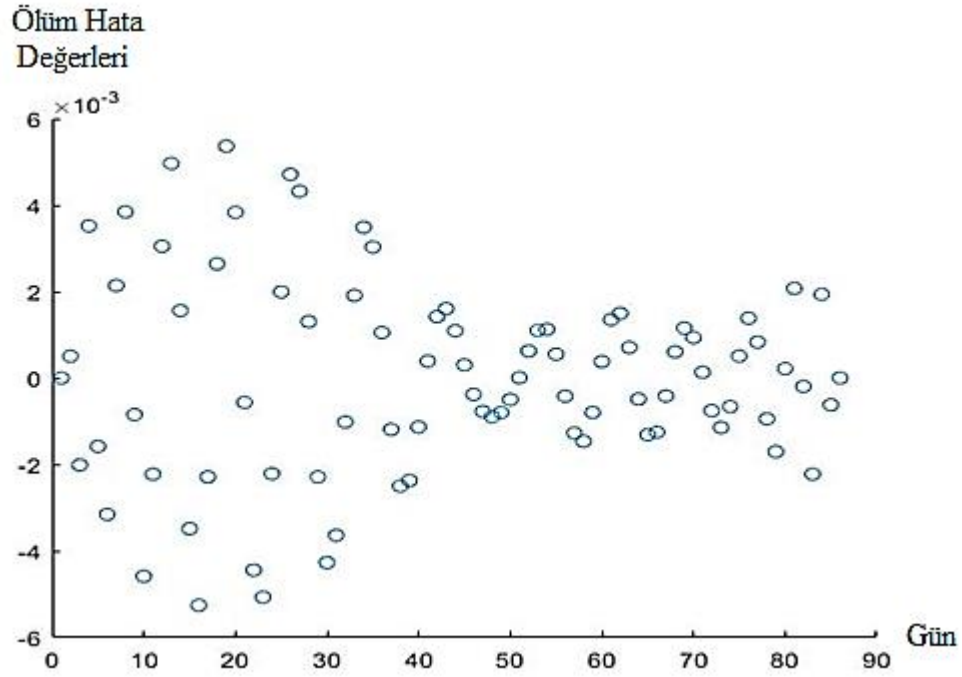
Şekil 3.10. Ölüm sayısının tahmin değerlerinin grafiği

Gözlenen ölüm oranlarının İstatistiksel olarak analizlerle tahmininin hata oranları Tablo3.10’da ki gibidir.

Tablo 3.10. 11 Mart-4 Haziran 2020 tarihleri arasında gözlenen ölüm tahmin değerlerinin hata oranları verilmiştir

Sıra No:	Hata	Sıra No:	Hata	Sıra No:	Hata	Sıra No:	Hata	Sıra No:	Hata	Sıra No:	Hata
1	0	17	-0,00526	33	-0,00102	49	-0,0009	65	-0,00049	81	0,000214
2	0	18	-0,00229	34	0,00191	50	-0,0008	66	-0,00131	82	0,002075
3	0,000503	19	0,002641	35	0,003489	51	-0,0005	67	-0,00125	83	-0,0002
4	-0,00201	20	0,005363	36	0,003029	52	7,92E-06	68	-0,00042	84	-0,00222
5	0,003517	21	0,003832	37	0,00105	53	0,000625	69	0,000605	85	0,001936
6	-0,00158	22	-0,00057	38	-0,00119	54	0,001101	70	0,001154	86	-0,00062
7	-0,00316	23	-0,00444	39	-0,0025	55	0,001126	71	0,000936		
8	0,002139	24	-0,00507	40	-0,00237	56	0,000553	72	0,000132		
9	0,003844	25	-0,00221	41	-0,00113	57	-0,00042	73	-0,00075		
10	-0,00085	26	0,001991	42	0,000391	58	-0,00127	74	-0,00114		
11	-0,00459	27	0,00471	43	0,00142	59	-0,00146	75	-0,00066		
12	-0,00222	28	0,004321	44	0,001605	60	-0,0008	76	0,000507		
13	0,003048	29	0,001305	45	0,001094	61	0,00038	77	0,001385		
14	0,004963	30	-0,00229	46	0,000304	62	0,001351	78	0,00083		
15	0,001559	31	-0,00427	47	-0,00038	63	0,001493	79	-0,00095		
16	-0,00349	32	-0,00364	48	-0,00078	64	0,000707	80	-0,0017		

Şekil3.11’de zamana bağlı olarak vakalardaki ölüm sayısının tahmin değerlerinin hata paylarının zamana bağlı değişimi verilmiştir;



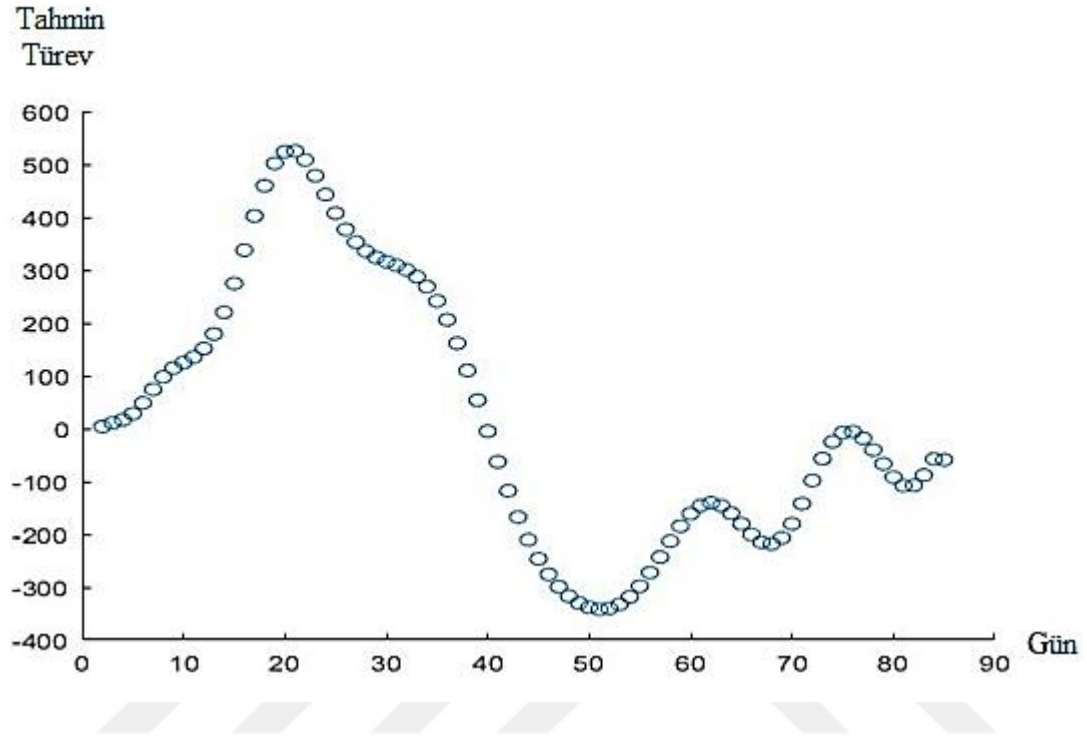
Şekil 3.11. Ölüm sayısının hata değerleri grafiği

Ölüm sayılarının tahmin değerlerinin türevleri ise Tablo3.11’de ki gibidir.

Tablo 3.11. 11 Mar - 4 Haziran 2020 tarihleri arasında ölüm tahmin değerlerinin türev değerleri.

Sıra No:	Türev	Sıra No:	Türev	Sıra No:	Türev	Sıra No:	Türev	Sıra No:	Türev	Sıra No:	Türev
1	0	17	337,7972	33	300,3653	49	-316,8	65	-160,171	81	91,9691
2	0	18	402,3464	34	287,6556	50	-329,834	66	-180,134	82	107,983
3	4,068674	19	459,7099	35	268,6024	51	-338,243	67	-200,359	83	107,493
4	11,04044	20	502,0228	36	241,576	52	-341,819	68	-214,819	84	87,8007
5	16,84573	21	524,2397	37	205,8038	53	-340,044	69	-218,049	85	57,6318
6	28,04204	22	525,3247	38	161,5473	54	-332,315	70	-206,586	86	58,8898
7	48,50513	23	508,1923	39	110,1196	55	-318,251	71	-180,107		
8	74,33127	24	478,5304	40	53,71382	56	-297,986	72	-141,856		
9	97,94423	25	443,0457	41	-4,92746	57	-272,389	73	-98,1115		
10	114,607	26	407,7885	42	-62,928	58	-243,16	74	-56,6906		
11	125,3333	27	377,0416	43	-117,687	59	-212,777	75	-24,8794		
12	135,4696	28	352,9381	44	-167,197	60	-184,281	76	-7,49436		
13	151,5982	29	335,6767	45	-210,223	61	-160,914	77	-5,85096		
14	179,0194	30	324,0442	46	-246,318	62	-145,61	78	-18,0337		
15	220,3246	31	315,9697	47	-275,685	63	-140,385	79	-40,0506		
16	274,7015	32	308,9547	48	-298,937	64	-145,733	80	-66,7838		

Şekil3.12’de ölüm oranlarının tahmin değerlerinin türevlerinin zamana bağlı değişimi gösterilmiştir;



Şekil 3.12. Tahmin eğrisinin 1. türevinin grafiği

4. SONUÇLAR

Korona virüs hastalığının epidemiolojisinin oluşturulmasında hastalık seyrinin takibi ve istatistiksel değerlendirmenin anlamlı olarak yapılabilmesi oldukça önemlidir. Özellikle dünyayı bu denli sarsan bir pandeminin hem ülkeler bazında hem de genel ölçüde değerlendirilmesi alınan kararların pandeminin seyrinin değiştirilmesindeki etkisi yüksek olabilir. Bu çalışmada korona virüs hastalığında Türkiye örneği ele alınarak geniş ölçüde incelenmeye çalışılmıştır. Çalışmada 86 günlük veri kullanılmıştır. İlk olarak yeni vaka sayıları incelenmiş ve Bernstein polinomu kullanılarak yeni vaka sayılarının tahmin eğrisi elde edilmiştir. Tahmin değerleri dikkate alınarak tahmin polinomunun hata kareler toplamı bulunmuştur. Elde edilen hata değerlerinin serpmme grafiği Şekil3.3 de verilmektedir. Şekil3.1, Şekil3.2 ve Şekil3.3 birlikte değerlendirildiğinde tahmin polinomunun yeni vaka sayılarını göstermedeki başarısı oldukça yüksektir. Bu bakımdan tahmin eğrisinin birinci türev grafiği hastalığının seyrini güçlü bir şekilde yansıtabilmektedir. Şekil3.4 de yer alan birinci türev grafiği incelendiğinde grafiğin 34 ve 35-inci günlerde pozitif değerden negatif değere geçtiği görülmektedir. Bu bize yeni vaka sayısının bu günlerde azalma eğilimine geçtiğini rahatlıkla belirtir. Bu hastalığın seyri için önemli bir istatistiksel bilgidir. Birinci türev grafiğine dikkat edildiğinde 17 ve 25-inci günlerde maksimum olduğu görülmektedir. İlk maksimum değeri olan 17-inci günden sonra yeni vaka sayısının artış hızı azalmaya başlamış fakat 25-inci güne doğru artış hızı tekrar artmaya başlayarak 25-inci günden sonra düşüşe geçmiştir. Hastalığın seyri grafiğin minimum noktaları olan 43, 68 ve 82-inci günlerde farklılık göstererek azalma eğilimini sürdürmüştür. Aralarda yeni vaka sayısının azalma hızı yavaşlarsa da yeni vaka sayısı azalmaya devam etmiştir. Türev grafiğine bakıldığında hastalığın ikiye katlama süresi 20 ile 25-inci günler arasında gözlemlendiğinden beş gün olarak alınması doğru olabilir.

Yeni tip korona virüs ile mücadelede test sayıları oldukça önemlidir. Virüsün Türkiye de yayılma diyagramına bakıldığında virüs ile mücadelenin ilk aşamalarında test sayılarının yetersiz düzeyde olduğu görülür iken zamanla artan test sayısı ile yeni vakaların belirlenmesi ve yayılımının hızının azaltıldığı görülmektedir. Vakalara bakıldığında da 33. Gün itibari ile yeni vaka sayısının test sayısına olan oranında ki azalma dikkat çekmektedir. Yani 33. Gün itibari ile iyice artırılmış olunan testlerle birlikte yeni vakaların hemen belirlendiği ve oluşabilecek yayılımın önüne geçmek için istatistiksel olarak anlamlı düzeyde verilerin bize bilgi verdiğini söyleye biliriz. Zaten yapılan tahmin analizine göre de zamana bağlı olarak test sayılarının artıp yayılımının önüne geçilip yeni vakada da süreç itibari ile azalış yaşanacağı açıkça görülmektedir. Alınan türev analizinde ise testin geliştirilerek uygulamasının artırılması süreci boyunca oranın yüksek olduğu lakin testin uygulamasının artırılması itibari ile oranın azalarak yayılımının önüne geçileceği de aşikar olarak ortaya konulmuştur.

Korona virüs vakalarına baęlı olarak Türkiye de 11 Mart 2020 ile 4 Haziran 2020 tarihleri arasında gözlenen ölüm sayılarına bakıldığında ise vakalardan kaynaklı ilk ölümün virüsün görüldüğü tarihten itibaren 6. Günde olduęu görülmektedir. yine alınan veriler itibariyle korona virüs sebepli ölümler vakanın görüldüğü ilk günden itibaren her ne kadar inişli çıkışlı bir çizgi izlemiş ise de en yüksek ölüm sayısına vakanın ilk görüldüğü tarihten itibaren 41. Günde görülmüş ve 127 kişi vefat etmiştir. En yüksek ölüm sayısına ulaşarak pik noktasını gören virüse baęlı ölümler 41. Gün sonunda ise azalarak sürmüştür ve 48.gün itibariyle çift hanelere gerilemiştir. Sürekli artırılan test sayıları ve virüsle mücadelede kullanılan tekniklerin bu azalmada payının olduęu söylenile bilinir



ÖNERİLER

Yirmi birinci yüzyılın ilk pandemisi olarak gözlenen yeni korona virüs, Covid-19, 2019'un sonunda Wuhan, Çin'de ortaya çıkmıştır. 24 Ocak 2020 itibarıyla, dünyada dokuz ülkede en az 830 vaka teşhis edilmiştir: Çin, Tayland, Japonya, Güney Kore, Singapur, Vietnam, Tayvan, Nepal ve Amerika Birleşik Devletleri. İlk gözlemlerde kronik rahatsızlığı olan hastalarda 26 ölüm vakası gözlenmiştir. 2002 yılında şiddetli akut solunum sendromu korona virüs (sars) salgını ve 2012'de Orta Doğu solunum sendromu korona virüs (mers) salgını göz önüne alındığında, Covid-19 son yirmi yılda insan popülasyonunda ortaya çıkan üçüncü korona virüştür.

Açıkça görülmektedir ki insandan insana etkili bulaşma, yeni korona virüsün salgına dönüşmesinde oldukça etkilidir. Aynı zamanda korona virüse yakalanan bireylerde hastalığın seyri, hastalığın teşhisi için oldukça önemlidir. Yeni korona virüs hastalığının bazı bireylerde hiçbir belirti göstermemesi hastalığın yayılmasında önemli bir faktördür.

Bu konuyla yapılan ilk çalışmalarda korona virüs vakalarının istatistik incelemesinden daha çok hastalığın klinik değerlendirmesi yapılmıştır. Hastalığın istatistiksel olarak incelenmesi üzerinde yapılan ilk çalışmalardan olan hastalığın modellenmesi ile ilgilidir [65]. Çalışmada SEIR modeli göz önünde bulundurulmuştur. Aron JL , Schwartz IB, Teheor J. tarafından geliştirilen bu model bir salgın hastalığın mevsimsellik ve vaka sayısının ikiye katlanması ile ilgili bir çalışmadır [66].

Uygulama kısmında yapılan analiz de Peng, L, Yang, W, Zhang, D. Zhuge,C.,Hong L. ve Aron, Teheor J. ve Schwartz IB, Teheor J. kaynaklarda yapılan çalışmalarla diferansiyel bakımdan benzerdir. Verinin yorumlanmasında polinomsal tahmin eğrilerinin birinci türevlerinden yararlanılmıştır. Bu yaklaşımın uygun olmasının sebebi bernstein polinomunun kullanılan veriyi minimum hata ile temsil etmesinden kaynaklanır.

KAYNAKLAR

- [1] Francis,G.(1886) *Regression Towards Mediocrity in Hereditary Stature*, The Journal of the Anthropological Institute of Great Britain and , C.15, 246-263, Ireland.
- [2] Alakuşı, K ve Erilli, N. (2015). *Non Parametric Regressions*, Erciyes Üniversitesi Fen Bilimleri Enstitüsü Dergisi C. 31, 70-82.
- [3] William Dale, B.(1993).*Understanding Regression Assumptions*, Little green book, Cilt.2, 70-110
- [4] Arık, Ö. (2014). *Yaşam Çözümlemesinde Yarı Parametrik Regresyon Yöntemleri ve Uygulaması*, Ankara Üni Övgü Gürel Yılmaz Dok Tezi.
- [5] Orhunbilge,N.(2000). *Uygulamalı Korelasyon ve Regresyon Analizi.*, Dergipark C.23 Sayı. 1, 23-30.
- [6] Özdamar,K.(2015). *Spss ile Biyoistatistik*, Nisan yayıncılık cilt. 1, 150-200, Ankara.
- [7] Ünver,Ö.,-Gangam,H.(2008).*Uygulamalı Temel İstatistik Yöntemleri*, Seçkin Yayın Evi C.1, 14, Ankara.
- [8] Çömlekçi,N.(1998). *Temel İstatistik ve Teknik Bilimleri*, Bilim Teknik Yayın Evi, 493.
- [9] Fox,J.(1997).*Applied Regression Analysis Linear Models*, A Research Tool, (Second Edit), 494-520, New York.
- [10] Gibbons,J. (1993).*Nonparametric Statistics book*,(Founding Edit), 86-110. Spain.
- [11] Watson,G.(1964). *Smooth Regression Analysis*, Indian Journal of Statistics, Series A, C. 26, 359-372. Sankhya.
- [12] Wolfgang,H.(1994). *Applied Nonparametric Regression, Econometric Society Monographs*, (Revised Edit), 45-48, German.
- [13] Tate,M.W Clelland,R.C. *Nonparametric and Statistics*, New York Heideerg (Berlin Edit), C.12, 55-58, Berlin.
- [14] Devroye, P.(1978) *The Uniform Convergence of Nearest Neighbor Regression Function Estimators*, Ieee Transactions on information theory C. 24 Sayı, 2, 142-151. Ney York.
- [15] Alpu, Ö., Şamkar, H.,Altan, E.(2010).*Sağlam Ride Regresyon Analizi*, Dokuz Eylül Üni.İktisadi idari bilimler dergisi, C. 25, Sayı. 2, 137-148.
- [16] Yadolah, D. (1999). *Altarnative Methods and Sons book*,(Wiley Edit), C. 240, 69. New York
- [17] Trevor,H.,Tibshirani,R.(1990). *Generalized Additive Models*, (An Introduction with R, Second Edit), 50-70, Toronto.
- [18] Hepsağ,A.(2007).*Parametrik Olmayan Regresyon ve Tahmin Yöntemleri*, Aycan Hesapdağ Yüksek Lisans Tezi, Sakarya Üni, 25-60.
- [19] Rice,J.(1984).*Nonparametric Regression The Annals of Statistics*, The Annals of Statistics, C. 2 Sayı: 4, 1215-1230.
- [20] Alakuşı, K ve Erilli, N. (2015). *Non Parametric Regressions*, Erciyes Üniversitesi Fen Bilimleri Enstitüsü Dergisi C. 31, 70-82.
- [21] Modd,A.M.(1950).*Introduction to the Theory of Statistics Parametrik Olmayan Regresyon analizi*, (Thirt Edit), 225-227,Tokyo.
- [22] Sen,P.K.(1968). *Estimates of the Regression Coefficeient Based on Kendall's*, (Board Edit). C. 41, 67, Germany.

- [23] Speath,H.(1991). *Linear Regression*, (Oxford Edit), C. 1346, 175-180, England.
- [24] Stone,C.J.(1977). *Consistens Nonparametric Regression*, The Annals of Statistics, C.3, 595-620.
- [25] Rousseeuw, P.J. (1984), "Least Median of Squares Regression", Journal of the American Statistical Association, 79, 871-880.
- [26] Nair Kr, Shrivastava Mp, 1942. *On a Simple Method of Curve Fitting*. Sankhya, 6, 121–132.
- [27] Kıroğlu,G.(2001). *Uygulamalı Parametrik Olmayan İstatistiksel Yöntemler*, Adir Kitap yayınları, 22-35,Ankara.
- [28] Peter, H.,Stephen, S.(1990).*On Variance Estimation in Nonparametric Regression Biometrika*, (Academica Edit), C. 77, Sayı. 2, 415-419. New York.
- [29] Takan,C.(1957). *Anadolu Üniversitesi. Bilim ve Teknoloji Dergisi*, C. 1, 59-66.
- [30] Oğuzlar, A.(1999). *Parametrik olmayan regresyon analizi*, Cumhuriyet Üni Yük Lis Tezi.
- [31] Wolfgang.H., Enno, M.(1993). *Comparing Nonparametric Versus Paremetrics Regression*, (Akedemik yayınları), 1926-1947, Türkiye.
- [32] Yıldız,N.,Topal,M.(2011). *Nonparametrik Regresyon Metotlarının incelenmesi, Atatürk Üni. İİBF Dergisi*, 10. Ekonometri ve İstatistik Sempozyumu Özel Sayı. 341
- [33] Yıldız,N.,Topal,M.(2004).*Farlı Dağılışı Gösteren Verilerde Nonparemetrik ve Parametrik Regresyonun İncelenmesi*. Nuray tezcan İstanbul Üni. Yük.Lis.Tezi.
- [34] Myles,H.,Douglas,W.(1999).*Nonparametrics Statistical Methods*, (Wily editor), C. 33, 23-32. Atalanta.
- [35] Ibragimov,I.A.,Hasminskii,R.Z.(1980).*On Nonparametric Estimation of Regression*, (Springer Edit), 810-814. New York.
- [36] Aron, J.L., Schwartz,IB.,Theor,J.(1984). *Seasonality and Period-doubling Bifurcations* doi.[10.1016/s0022-5193\(84\)80150-2](https://doi.org/10.1016/s0022-5193(84)80150-2)
- [37] Monney,Z.(1996).*Bootstrappin A Nonparametric Statistical*. Sage uni.paper book, C.57, 44.
- [38] Wolfgang, H.,Theo, G.(1984). *Robust Nonparatetric Function Fitting*, Technical University, C.12, 45-51, Munich.
- [39] Pater John,B., Freedman D.(1981). *Some Asymptotic Theory Fort he Bootstrap*, The Annals of Statistics,C. 5, 1196-1217.
- [40] Clark R.(2014). *Bootstrap Methods*, A Guide to Doing Statistics in Second Language Research Using SPSS and R book,(Second Edition), 25-27, New York.
- [41] Jun,S.(1996). *Bootstrap Model Selectios, Regresyon analizi*, Tümay yayınları. C.1, 45-56.,Ankara.
- [42] Lahiti, N.(1992). *On Bootstrapping*, (Second Edit), C.25, 15-16, USA.
- [43] Bradley,E. – Gail,G.(1983). *Aleisurely Look at The Bootstrap*, The American Statistician, C. 37, Sayı.1, 06-48. New York.
- [44] Breadley, E.(1982). *The Jackknife te Bootstrap and other Resampling Plans book*, Bradley Efron Department of Statistics Stanford University C. 30, 55-68, California.
- [45] Maritz,J.S.(1979). *On Theil's Method in Distribution-Free Regressions Australian Journal of Stati*, Akademi yayınları, C. 7, 75-77, İstanbul.
- [46] Jeremy,B.(1996).*Recent Developments in Bootstrapping Times Series* (Wiley Edit), 2-34. Atalanta
- [47] Okutan,D.(2009). *Bootstrap Yönteminin Regresyon Analizinde Kullanımı*, 6. İstatistik kongresi

- [48] Aral A., Morales, D. (2018). *Bernstein type operatöre*, C.12, ss.5,doi:10.7153/jmi-2018-12-64
- [49] Videnakii,V.S.(1990).*Bernatein Polynomial*, Dergipark makale, C. 5 Sayı. 123, 12-25.
- [50] Birou,M.(2014). *Note about some general King-.type operators*, Annals of the Tiberiu Popoviciu Seminar, C. 12, 3-16.
- [51] Gonska,H.P.(2005).*Ramarks on an article of J.P. King*, Comment. Math. Univ. C.44, 555-563.
- [52] Beran R.,Muni,S.(1985). *Botstrap Test and Confidence Regions For Functions of a Covariance Matrix*,The Annals of Statistics,C. 2, 95-115.
- [53] Akkuş,Ö.(2008). *İki düzeyli bağımlı değişken modelinin yarı parametrik tahmini*, İstatistikçiler dergisi say.1, 135-143.
- [54] Griffin, J.(1962). *Statistics Methods and Aplication Rinehart*, Çukurova Üni.Yük Lis.Tezi, Savaş Okur, 259-273, Adana
- [55] Gonska,H.P.,Paşa,I.(2009). *General King –Type Operators Of King Type*, (Academic Edi) ,C.5,25-60
- [56] Ruppert,D.,Wand,M.P.,Carroll,R.C.(2003). *Semiparametric Regression*, Cambridge University Press C. 1356, 55-60.
- [57] Koutsoyiannis,A.(1992).*Ekonometri Kuramı*.Verso yayıncılık, 42-55, İstanbul.
- [58] Ertek, T.(2000).*Ekonomiye Giriş*, İktisada Giriş, Beta yayınları, 150. Ankara.
- [59] Pagan,A.,Aman,U.(1999).*Nonparametric*, Cambridge Universit, (Press Edit),C.1, 25-27. Cambridge
- [60] Fukunaga,K.,Modi,A.(1975).*BranchandBoundAlgorithm*,Englewood, doi.10.1002/9780470400531.eorms0116
- [61] Ertek, T.(2000).*Ekonomiye Giriş*, İktisada Giriş, Beta yayınları, 150. Ankara.
- [62] Atakan,C.(2003).*Gerçek Hata Oranına İlişkin Güven Aralığı*,Anadolu Üniversitesi Bilim Teknoloji Dergisi. C.5, 12 -15. Eskişehir.
- [63] Şengül,M.(1999). *Yeniden Örnekleme Metodunda Nonparametrik Yaklaşım*, DergiPark, Makale C. 7 Sayı. 1, 75-79
- [64] Alabarda,Y.(2020).*Covid19casesofSomaliawithglobalbackgrund*,WOHOsitesindenerişildi, Erişim: <https://www.who.int/emergencies/crises/som/en/>.
- [65] Peng,L,Yang,W., Zhang,D.,Zhuge,C.,Hong,L. (2020). *Epidemic analysis of COVID-19 in Chinaby dynamical modeling*. arXiv preprint arXiv, Erişim: <https://arxiv.org/>
- [66] Aron, J.L.,Schwartz.I.B.,Teheor J. (1984), doi. 10.1016/s0022-5193(84)80150-2

ÖZGEÇMİŞ

Yücel GÜL

KİŞİSEL BİLGİLER

Doğum Yeri : Elazığ
Doğum Yılı : 1992
Uyruğu : Türkiye Vatandaşı
Adres : ELAZIĞ
E-posta : yucel_gl@hotmail.com

EĞİTİM BİLGİLERİ

Lisans : Fırat Üniversitesi İstatistik Bölümü / 2011 – 2015
Lise : Hıdır Sever Lisesi / 2007 – 2011

İŞ DENEYİMİ

2018 – 2019 : Kuveyt Türk Katılım Bankası