

**T.C.  
ADNAN MENDERES ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ  
BİYOİSTATİSTİK YÜKSEK LİSANS PROGRAMI**

**PARÇALI REGRESYON İLE POLİNOM REGRESYON  
ANALİZLERİNİN KARŞILAŞTIRILMASI**

**BUĞRA VAROL  
YÜKSEK LİSANS TEZİ**

**DANIŞMAN  
Prof. Dr. İmran KURT ÖMÜRLÜ**

**AYDIN-2017**

## KABUL VE ONAY SAYFASI

T.C. Adnan Menderes Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı Yüksek Lisans Programı çerçevesinde Buğra VAROL tarafından hazırlanan "Parçalı Regresyon ile Polinom Regresyon Analizlerinin Karşılaştırılması" başlıklı tez, aşağıdaki jüri tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.

Tez Savunma Tarihi: 11/07/17

### Ünvanı, Adı ve Soyadı

Prof. Dr. İmran KURT ÖMÜRLÜ  
(Tez Danışmanı)

### Üniversitesi

Adnan Menderes Üniversitesi

### İmza



Prof. Dr. Mevlüt TÜRE

Adnan Menderes Üniversitesi



Prof. Dr. Mehmet N. ORMAN

Ege Üniversitesi



### ONAY:

Bu tez Adnan Menderes Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca yukarıdaki jüri tarafından uygun görülmüş ve Sağlık Bilimleri Enstitüsü'nün .....<sup>26.07.2017</sup> tarih ve 25..sayılı oturumunda alınan ..X...no'lu Yönetim Kurulu kararıyla kabul edilmiştir.



Prof. Dr. Ahmet CEYLAN  
Enstitü Müdürü

## TEŞEKKÜR

Yüksek lisans öğrenimim süresince ve tezimi tamamlama aşaması boyunca yardımlarıyla kendilerini yanımda hissettiğim değerli hocalarıma, aileme ve çalışma arkadaşlarıma teşekkürlerimi sunmak istiyorum:

Tez çalışmamın tüm aşamalarında bilgi ve deneyimleriyle, yardım ve desteğini hiçbir zaman esirgemeyen danışman hocam Sayın Prof. Dr. İmran KURT ÖMÜRLÜ'ye,

Tez çalışmam boyunca değerli bilgilerinden yararlandığım, görüş ve önerileri ile yol gösteren Sayın Prof. Dr. Mevlüt TÜRE'ye,

Tez yazım aşamasında verdikleri moral ve zorlandığım noktalardaki destekleri sayesinde tezimi bitirmeme yardımcı olan Biyoistatistik Anabilim Dalı'ndaki Araştırma Görevlisi arkadaşlarım Fulden CANTAŞ ve Hakan ÖZTÜRK'e,

Tez hazırlama sürecinde gösterdiği sabır, anlayış ve verdiği destek ile aldığım her nefeste sevgisini hissettiğim sevgili eşim Gülnur VAROL'a,

Beni yetiştirip bugünlere getiren, maddi-manevi desteklerini hiçbir zaman esirgemeyen ve haklarını ödeyemeyeceğim sevgili annem İsmet VAROL'a, babam Oğuz Sabri VAROL'a ve ablam Tuğçe VAROL KAYABAŞI'na sonsuz teşekkürlerimi sunarım.

# İÇİNDEKİLER

	Sayfa
KABUL VE ONAY SAYFASI.....	i
TEŞEKKÜR .....	ii
İÇİNDEKİLER.....	iii
SİMGELER VE KISALTMALAR DİZİNİ .....	v
ŞEKİLLER DİZİNİ .....	vi
TABLolar DİZİNİ.....	viii
ÖZET .....	ix
ABSTRACT .....	x
1. GİRİŞ.....	1
2. GENEL BİLGİLER.....	3
2.1. Tek değişkenli Polinom Regresyon Analizi .....	3
2.1.1. Polinom Regresyon Modelleri.....	3
2.1.1.1. Karesel Regresyon Modeli .....	3
2.1.1.2 Kübik Regresyon Modeli .....	4
2.1.2 En Küçük Kareler Yöntemi .....	5
2.2 Parçalı Regresyon .....	6
2.2.1 Dönüm Noktası.....	6
2.2.1.1 Sabit Dönüm Noktası .....	7
2.2.1.2 Değişken Dönüm Noktası.....	7
2.2.2 “+” Fonksiyonları .....	8
2.2.3 Parçalı Regresyon Modelleri .....	9
2.2.3.1 Karesel Parçalı Regresyon.....	13
2.2.3.2 Kübik Parçalı Regresyon .....	14
2.2.3.3 Karma Parçalı Regresyon .....	16
2.3 Model Değerlendirme Kriterleri .....	18
2.3.1 Belirtme Katsayısı .....	18
2.3.2 Hata Kareler Ortalaması .....	18
2.3.3 Akaike Bilgi Kriteri .....	19
2.3.4 Bayes Bilgi Kriteri.....	20
3. GEREÇ VE YÖNTEM.....	21
3.1 Kullanılan Programlar .....	21
3.2 Simülasyon Uygulaması.....	21

3.2.1 İkinci Dereceden Simülasyon Algoritması.....	21
3.2.2 Üçüncü Dereceden Simülasyon Algoritması.....	22
3.3 Gerçek Veri Seti Uygulaması .....	22
3.3.1 Tüberküloz Veri Seti .....	23
3.3.2 Kızamık Veri Seti .....	24
4. BULGULAR .....	26
4.1 Simülasyon Sonuçları .....	26
4.1.1 İkinci Dereceden Türetilen Veriler ile Yapılan Simülasyon Sonuçları.....	26
4.1.2 Üçüncü Dereceden Türetilen Veriler ile Yapılan Simülasyon Sonuçları.....	29
4.2 Gerçek Uygulama Sonuçları.....	32
4.2.1 Tüberküloz Veri Seti .....	32
4.2.1.1 Karesel Regresyonla Model Tahmini .....	32
4.2.1.2 Kübik Regresyonla Model Tahmini .....	32
4.2.1.3 Parçalı Regresyonla Model Tahmini .....	33
4.2.1.4 Regresyon Modellerinin Karşılaştırılması.....	36
4.2.2 Kızamık Veri Seti .....	37
4.2.2.1 Kübik Regresyonla Model Tahmini .....	37
4.2.2.2 Parçalı Regresyonla Model Tahmini .....	37
4.2.2.3 Regresyon Modellerinin Karşılaştırılması.....	40
5. TARTIŞMA.....	41
6. SONUÇ VE ÖNERİLER .....	45
KAYNAKLAR.....	47
ÖZGEÇMİŞ	

## SİMGELER VE KISALTMALAR DİZİNİ

ABK	:	Akaike bilgi kriteri
$y$	:	Bağımlı değişken
$\hat{y}$	:	Bağımlı değişkenin tahmin değeri
$x$	:	Bağımsız değişken
$k$	:	Bağımsız değişkenin derecesi
$\mathbf{x}$	:	Bağımsız değişken matrisi
$\mathbf{b}_0$	:	Regresyon katsayı vektörü
BBK	:	Bayes bilgi kriteri
$R^2$	:	Belirtme katsayısı
$n$	:	Örneklem birim sayısı
$t$	:	Dönüm noktası değeri
$m$	:	Dönüm noktası sayısı
EKK	:	En küçük kareler
HKO	:	Hata kareler ortalaması
$e$	:	Örneklem hata terimi
$\mathbf{e}$	:	Hata terimi matrisi
$b_{0j}$	:	Regresyon katsayısı
$k + md + 1$	:	Sabit terim dahil parçalı regresyonun parametre sayısı
sd	:	Serbestlik derecesi
$\mathbf{S}$	:	Parçalı regresyon modeli bağımlı değişken vektörü
$s_b$	:	Tahminin standart hatası
$( )^T$	:	Transpoz işlemi
$( )^{-1}$	:	Ters alma işlemi
$\varepsilon$	:	Hata terimi
$N$	:	Birim sayısı
$\beta$	:	Parametre katsayısı
$d$	:	"+" fonksiyonunun derecesi
$b_{lv}$	:	"+" fonksiyonunun regresyon katsayısı
$\mathbf{T}_m$	:	"+" fonksiyonu matrisleri
$\mathbf{b}_m$	:	"+" fonksiyonunun regresyon katsayı vektörleri

## ŞEKİLLER DİZİNİ

	Sayfa
Şekil 2.1 Karesel regresyon modelinin noktasal dağılımı .....	4
Şekil 2.2 Kübik regresyon modelinin noktasal dağılımı .....	4
Şekil 2.3 Parçalı regresyonda sabit dönüm noktası .....	7
Şekil 2.4 Parçalı regresyon aday dönüm bölgeleri .....	8
Şekil 2.5 Parçalı regresyon genel dağılımında süreklilik kısıtlamaları .....	10
Şekil 2.6 Karesel+Karesel yapıda $[x_1, x_n]$ aralığında $t_1$ dönümü .....	13
Şekil 2.7 Kübik+kübik yapıda $[x_1, x_n]$ aralığında $t_1$ dönümü .....	15
Şekil 2.8 Karesel+kübik yapıda $[x_1, x_n]$ aralığında $t_1$ dönümü .....	17
Şekil 3.1 Tüberküloz vaka sayısının yaşa göre dağılımı .....	23
Şekil 3.2 Kızamık vaka sayısının zamana göre dağılımı .....	25
Şekil 4.1 Karesel regresyon ve karesel+karesel yapıdaki parçalı regresyon için $R^2$ değerlerinin kutu grafiği .....	27
Şekil 4.2 Karesel regresyon ve karesel+karesel yapıdaki parçalı regresyon için HKO değerlerinin kutu grafiği .....	27
Şekil 4.3 Karesel regresyon ve karesel+karesel yapıdaki parçalı regresyon için ABK değerlerinin kutu grafiği .....	28
Şekil 4.4 Karesel regresyon ve karesel+karesel yapıdaki parçalı regresyon için BBK değerlerinin kutu grafiği .....	28
Şekil 4.5 Kübik regresyon ve parçalı regresyon modelleri için $R^2$ değerlerinin kutu grafigi .....	30
Şekil 4.6 Kübik regresyon ve parçalı regresyon modelleri için HKO değerlerinin kutu grafigi .....	30
Şekil 4.7 Kübik regresyon ve parçalı regresyon modelleri için ABK değerlerinin kutu grafigi .....	31
Şekil 4.8 Kübik regresyon ve parçalı regresyon modelleri için BBK değerlerinin kutu grafigi .....	31
Şekil 4.9 Tüberküloz vaka sayısının yaşa göre dağılımı ve dönüm noktası aday bölgeleri .....	34
Şekil 4.10 Yaşa göre değişen tüberküloz vaka sayılarının regresyon modelleri ile tahmin edilen ve gözlenen değerlerinin gösterimi .....	36
Şekil 4.11 1970-2015 yılları arası kızamık vaka sayılarının noktasal dağılımı .....	38

Şekil 4.12 Kızamık veri setindeki 1970-2015 yılları arasındaki kızamık vaka sayılarının regresyon modelleri ile tahmin edilen ve gözlenen değerlerinin gösterimi .....40





## TABLULAR DİZİNİ

	<b>Sayfa</b>
Tablo 3.1 Değişen yaş değerlerine göre toplam tüberküloz vaka sayısı .....	24
Tablo 3.2 Zamana göre değişen kızamık vaka sayısı .....	25
Tablo 4.1 Karesel ve parçalı regresyon modellerinin $R^2$ , HKO, ABK ve BBK değerlerinin tanımlayıcı istatistikleri ve karşılaştırma sonuçları .....	26
Tablo 4.2 Kübik regresyon ve parçalı regresyon modellerinin $R^2$ , HKO, ABK ve BBK değerlerinin tanımlayıcı istatistikleri ve karşılaştırma sonuçları.....	29
Tablo 4.3 Karesel regresyon parametre tahminleri .....	32
Tablo 4.4 Kübik regresyon parametre tahminleri.....	33
Tablo 4.5 Karesel+karesel+karesel yapıdaki parçalı regresyon parametre tahminleri.....	35
Tablo 4.6 Oluşturulan regresyon modellerinin $R^2$ , HKO, ABK ve BBK değerleri .....	36
Tablo 4.7 Kübik regresyon parametre tahminleri.....	37
Tablo 4.8 Karesel+karesel yapıdaki parçalı regresyon parametre tahminleri .....	39
Tablo 4.9 Oluşturulan regresyon modellerinin $R^2$ , HKO, ABK ve BBK değerleri .....	40

## ÖZET

### PARÇALI REGRESYON İLE POLİNOM REGRESYON ANALİZLERİNİN KARŞILAŞTIRILMASI

**Varol B. Adnan Menderes Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı Yüksek Lisans Tezi, Aydın, 2017.**

Bu çalışmada hem simülasyon verisi hem de gerçek veri setleri kullanılarak tek değişkenli polinom regresyon analizi ile karesel ve kübik parçalı regresyon analizlerinin karşılaştırılması hedeflendi.

Çalışmanın uygulama basamağında R yazılım programı kullanılarak simülasyon uygulaması için algoritmalar oluşturuldu. Hem bu algoritmalar yardımıyla türetilmiş veri setleri hem de tüberküloz ve kızamık veri setleri kullanılarak oluşturulan polinom ve parçalı regresyon modellerinin tahmin performansları; belirtme katsayısı ( $R^2$ ), hata kareler ortalaması (HKO), Akaike bilgi kriteri (ABK) ve Bayes bilgi kriteri (BBK) değerlerine göre karşılaştırıldı.

Simülasyon uygulaması ve gerçek veri setleri ile yapılan uygulamaların sonunda en uygun dönüm noktalarına göre oluşturulan tüm parçalı regresyon modellerinin  $R^2$  değerlerinin polinom regresyonlardan daha yüksek; HKO, ABK ve BBK değerlerinin ise daha düşük olduğu belirlendi.

Sonuç olarak bağımlı ve bağımsız değişken arasındaki ilişkinin inceleneceği tüm veri setleri için sayı ve konumu en uygun şekilde belirlenen dönüm noktalarına göre oluşturulan parçalı regresyon modelleri, polinom regresyon modellerine göre daha yüksek performanslı tahminleme yapmaktadır.

**Anahtar Kelimeler:** Polinom Regresyon, Parçalı Regresyon, Dönüm Noktası, Simülasyon.

## ABSTRACT

### COMPARISON OF PIECEWISE REGRESSION AND POLYNOMIAL REGRESSION ANALYSES

**Varol B. Adnan Menderes University Institute of Medical Sciences Biostatistics  
Department Master of Science Thesis, Aydın, 2017.**

In this study, it was aimed to compare quadratic and cubic piecewise regression analyses and univariate polynomial regression analysis using both simulated data and real data sets.

In the application step of the study, algorithms are created by using R software for simulation practice. Estimation performances of the polynomial and piecewise regression models created using both data sets generated by means of these algorithms and real data sets are compared according to coefficient of determination ( $R^2$ ), mean square error (MSE), Akaike information criteria (AIC) and Bayes information criteria (BIC). At the end of the simulation applications and the real data sets, the  $R^2$  values of all piecewise regression models formed with respect to the most suitable knots are higher than those of polynomial regression; MSE, AIC and BIC values were found to be lower.

As a result, for all data sets to examine the relationship between dependent and independent variables, the piecewise regression models that are formed according to the knots determined optimally by the number and the position make higher performance estimations than the polynomial regression models.

**Keywords:** Polynomial Regression, Piecewise Regression, Knot, Simulation.

# 1. GİRİŞ

Sağlık alanında yapılan bilimsel çalışmalarda bağımlı ve bağımsız değişken arasındaki ilişkinin incelenmesi için yaygın olarak kullanılan doğrusal regresyon modellerinin yanı sıra polinom regresyon modellerinden de yararlanılmaktadır. Ancak bağımlı ve bağımsız değişken arasındaki ilişkinin noktasal dağılımı, polinom regresyon modelleri ile ifade edilemeyecek derecede sapmalar gösteriyorsa, bu modellerin dağılımı ifade etme gücü düşmektedir. Bu gibi durumlarda parçalı regresyon modelleri kullanılmaktadır (Wainer, 1971; Seber ve Wild, 2003). Noktasal dağılımın parçalı regresyonlarla incelenmesi, araştırmacıya düşük dereceli polinomlarla uğraşma, bağımsız değişkenin herhangi bir aralığı için tahmin yapabilme, daha esnek eğriler elde etme ve bilinen modellerle açıklanamayacak kadar karmaşık dağılımları uygun dönüm noktalarına göre rahatlıkla modelleyebilme olanakları sağlamaktadır (Gallant ve Fuller, 1973; Eubank, 1999).

Bir veya daha fazla parçanın kırılma noktalarında birleştirildiği parçalı regresyon, istatistiksel bir teknik olarak yaygın bir şekilde kullanılmaktadır. Parçalı fonksiyonları oluşturan polinomlar, bilgisayar programlarında kolaylıkla birleştirilebilir. Bu nedenle, özellikle deneysel verilerin ya da modellenen eğrilerin tahmininde parçalı polinomların kullanımı uygundur (DeBoor ve Rice, 1968).

Literatürlere bakıldığında, araştırmacılar tarafından parçalı regresyonların kullanılmasına yönelik çalışmaların yapıldığı görülmektedir. Schoenberg (1946), parçalı fonksiyon ifadesini matematik alanında ilk defa kullanan bilim adamıdır. Genel matematikte parçalı fonksiyonların önemi, diferansiyel eşitliklerde parçalı fonksiyonların kullanımı ve parçalı teorisi üzerine çalışmıştır. Ahlberg ve Nilson (1953), parçalı yaklaşımının temel kavramlarını ve hata kareler ortalaması bakımından uyum özelliklerini teorik olarak incelemişlerdir. Smith (1979), “+” fonksiyonlarını kullanarak sabit dönüm noktalı parçalı regresyonların teorisini ele almış ve farklı dereceli polinom parçalarından oluşan karma parçalı regresyon modellerini göstermiştir. Parçalı regresyon modelleri sabit dönüm noktalı olarak ele alındığında, bu modellerin herhangi bir istatistik paket programında en küçük kareler yöntemiyle tahmin edilebileceğini belirtmiştir. Mills (1996), ilaç denemelerinde parçalı regresyonların nasıl kullanılabileceği, fonksiyonları oluştururken nelere dikkat edilmesi gerektiği, dönüm noktalarının nasıl belirleneceği gibi temel kavramlar üzerinde çalışmıştır. Park (1978), Birçok uygulamada tek bir polinom fonksiyonla tüm deney alanı üzerinde yeterli tahmin sağlanamadığından tek bir polinom fonksiyonla bağımlı değişkenin yeterince açıklanamadığını belirtmiştir. Bunun gibi durumlarda parçalı regresyon modellerinden yararlanılması gerektiğini izah etmiştir. Farebtother ve Schyns (2002),

parçalı regresyonlarla eğri tahminleme ve parçalı regresyonların avantaj ve dezavantajları üzerinde durmuşlardır. Seber ve Wild (2003), parçalı regresyonları teorik olarak açıklamış ve dönüm noktalarının sayı ve konumunun belirlenmesinde dikkat edilecek hususları görsel inceleme sonucu tespit etmişlerdir. Mulla (2007), kareli parçalı regresyonları kullanarak serum albümin ve hastane ölüm oranı riski arasındaki ilişkiyi araştırmıştır.

Bu çalışmanın amacı, simülasyon verisi kullanarak tek değişkenli polinom regresyon analizi ile karesel ve kübik parçalı regresyon analizlerinin performanslarını karşılaştırmaktır. Ayrıca tüberküloz ve kızamık hastalıkları ile ilgili veri setlerinde bu yöntemlerin kullanımı ve karşılaştırılması da yapılacaktır. Oluşturulan modellerin performansları belirtme katsayısı ( $R^2$ ), hata kareler ortalaması (HKO), Akaike bilgi kriteri (ABK) ve Bayes bilgi kriterine (BBK) göre değerlendirilecektir.



## 2. GENEL BİLGİLER

### 2.1. Tek Değişkenli Polinom Regresyon Analizi

Regresyon analizinin amacı bağımlı değişken ile bir ya da daha çok bağımsız değişken arasında kurulan modelin parametrelerini tahmin ederek, bağımsız değişkenin gözlenen değerleri için bağımlı değişkenin alacağı değeri belirlemektir (Freedman, 2005; Alpar, 2017). Tek bir bağımsız değişkenin kullanıldığı regresyon türüne tek değişkenli regresyon analizi denir. Tek değişkenli polinom regresyon eğrisinin denklemi aşağıdaki gibi hesaplanır (Fan, 1996; Rawlings ve ark, 1998):

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_k x_i^k + \varepsilon_i \quad i = 1, 2, \dots, N$$

Bu eşitlikte,

$y_i$ : Bağımlı değişkenin  $i$ . gözlem değerini,

$x_i$ : Bağımsız değişkenin  $i$ . gözlem değerini,

$N$ : Birim sayısını,

$\beta_0$ : Regresyon sabitini,

$\beta_1, \beta_2, \dots, \beta_k$ : Regresyon katsayılarını,

$k$ : Bağımsız değişkenin derecesini,

$\varepsilon_i$ : Modele ait hata terimini göstermektedir.

#### 2.1.1. Polinom Regresyon Modelleri

Polinom regresyon analizi, regresyon modelinde yer alan bağımsız değişkenin, bağımlı değişken üzerine olan etkisinin toplanabilir olmadığı (eğrisel, üstel, çarpımsal vb.) varsayımına dayanan regresyon yöntemlerini içerir (Rawlings ve ark, 1998; Aggrey, 2002). Bu başlık altında karesel ve kübik yapıdaki polinom regresyon modelleri ele alınacaktır.

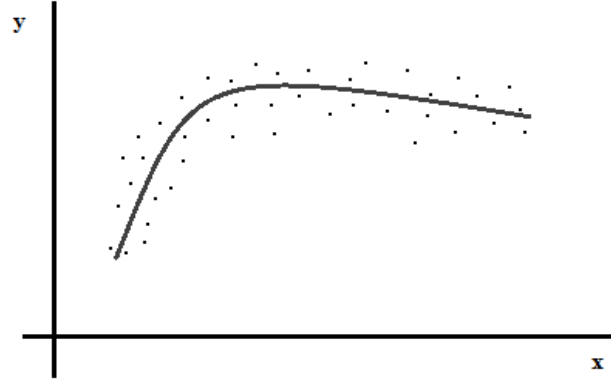
##### 2.1.1.1. Karesel Regresyon Modeli

İki değişken arasındaki sebep-sonuç ilişkisini ikinci dereceden bir model ile ortaya koyan yönteme karesel regresyon analizi denir (Rawlings ve ark, 1998; Seber ve Wild, 2003).

$\beta_0, \beta_1$  ve  $\beta_2$  parametrelerinin örneklemden elde edilen kestirimleri " $b_0, b_1$  ve  $b_2$ ", hata terimi ( $\varepsilon$ )'nin örneklemden elde edilen değeri " $e$ " olarak ele alındığında,  $n$  örneklem birim sayısı olmak üzere tek değişkenli regresyon eğrisinin denklemi,

$$y_i = b_0 + b_1x_i + b_2x_i^2 + e_i \quad i = 1,2, \dots, n$$

şeklinde ifade edilir (Rawlins ve ark, 1998; Lee ve Brooks, 2006; Draper ve Smith, 2014). Böyle bir modeli temsil eden noktasal dağılımın grafiksel gösterimi Şekil 2.1’de verilmiştir.



Şekil 2.1 Karesel regresyon modelinin noktasal dağılımı

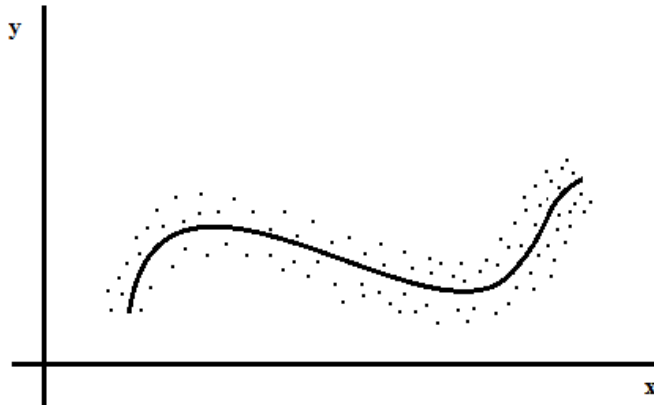
### 2.1.1.2 Kübik Regresyon Modeli

İki değişken arasındaki sebep-sonuç ilişkisini üçüncü dereceden bir model ile ortaya koyan yöntem kübik regresyon analizi denir (Rawlins ve ark, 1998; Seber ve Wild, 2003).

Tek değişkenli kübik regresyon eğrisinin denklemi,

$$y_i = b_0 + b_1x_i + b_2x_i^2 + b_3x_i^3 + e_i \quad i = 1,2, \dots, n$$

şeklinde ifade edilir (Rawlins ve ark, 1998; Lee ve Brooks, 2006; Draper ve Smith, 2014). Böyle bir modeli temsil eden noktasal dağılımın grafiksel gösterimi Şekil 2.2’de verilmiştir.



Şekil 2.2 Kübik regresyon modelinin noktasal dağılımı

### 2.1.2 En Küçük Kareler Yöntemi

Regresyon analizinde verilen noktaları temsil eden en iyi denklemi bulma işleminde yaygın olarak kullanılan yöntem en küçük kareler (EKK) yöntemidir. EKK yöntemi hem birinci dereceden fonksiyonlar için hem de polinomlar için kullanılabilir (Hartley ve Brooker, 1965; Seber ve Wild, 1989; Rawlins ve ark, 1998).

Tahmin edilmek istenilen polinomun derecesi  $k$  olmak üzere regresyon denklemi,

$$\hat{y}_i = b_0 + b_1x_i + b_2x_i^2 + \dots + b_kx_i^k \quad i = 1, 2, \dots, n$$

olarak yazılır. Her bir  $\{(x_i; y_i): i = 1, \dots, n\}$  veri çifti yukarıdaki denklemde yerine konulduğunda aşağıdaki denklemler elde edilir (Ramsey, 1969; Freeman ve Skapura, 1991):

$$\begin{aligned}\hat{y}_1 &= b_0 + b_1x_1 + b_2x_1^2 + \dots + b_kx_1^k \\ \hat{y}_2 &= b_0 + b_1x_2 + b_2x_2^2 + \dots + b_kx_2^k \\ &\vdots \\ \hat{y}_n &= b_0 + b_1x_n + b_2x_n^2 + \dots + b_kx_n^k\end{aligned}$$

Elde edilen denklemler  $\hat{\mathbf{y}} = \mathbf{xb}$  matris formunda yeniden tanımlanınca aşağıdaki eşitlik elde edilir (Lind ve ark, 2004):

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^k \\ 1 & x_2 & x_2^2 & \dots & x_2^k \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^k \end{bmatrix} * \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix}$$

Bağımlı değişkenin tahmininde oluşan hatalardan dolayı yukarıdaki modele bir hata vektörü ( $\mathbf{e}$ ) eklendiğinde gözlenen  $y_i$  değerleri elde edilir. Böyle bir durumda model  $\mathbf{y} = \mathbf{xb} + \mathbf{e}$  şeklinde yeniden tanımlanır (Jang ve ark, 1997):

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^k \\ 1 & x_2 & x_2^2 & \dots & x_2^k \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^k \end{bmatrix} * \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

EKK yönteminin temelini, toplam sapmaların karelerinin toplamını en küçük yapan hata değerlerinin bulunması oluşturmaktadır. Hata kareler toplamı aşağıdaki eşitlikle belirtilmektedir (Chen ve ark, 1991; Lehmann ve ark, 1998):



$$E(\mathbf{b}) = \sum_{i=1}^n e_i^2 = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{xb})'(\mathbf{y} - \mathbf{xb})$$

Hata kareler toplamının minimum olması için her bir  $b_j$ , ( $j = 1, 2, \dots, k$ ) katsayısına göre birinci türevlerinin sıfır olması istenir. Buna göre:

$$E(\mathbf{b}) = (\mathbf{y} - \mathbf{xb})'(\mathbf{y} - \mathbf{xb}) = \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{x}'\mathbf{y} + \mathbf{b}'\mathbf{x}'\mathbf{xb} \text{ olduğundan;}$$

$$\frac{\partial E(\mathbf{b})}{\partial \mathbf{b}} = -2\mathbf{x}'\mathbf{y} + 2\mathbf{x}'\mathbf{xb} = 0,$$

$$\mathbf{x}'\mathbf{xb} = \mathbf{x}'\mathbf{y}$$

eşitliği elde edilir. Bu eşitlikten hata kareler toplamını minimum yapan parametre tahminleri,

$$\mathbf{b} = (\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{y}$$

denklemleri ile hesaplanır (Hartley ve Brooker, 1965; Jang ve ark, 1997).

## 2.2 Parçalı Regresyon

Bağımlı ve bağımsız değişkenlere ait noktasal dağılımların dönüm olarak adlandırılan belirli noktalardan parçalara ayrılarak incelenmesi durumu “parçalı regresyon” olarak adlandırılır (Wainer, 1971; Poirier 1973; Porth, 1984).

Veri çiftleri  $\{(x_i; y_i): i = 1, \dots, n\}$  şeklinde olan bir veri setinde çok sayıda  $(x, y)$  veri noktalarını tek bir eğri ile tahmin etmek her zaman uygun olmayabilir. Nokta sayısı arttıkça noktasal dağılımdan sapmalar artacak ve  $x$  ile  $y$  arasındaki ilişkiyi temsil eden polinomun derecesi artış göstereceği için oluşturulan modelin tahmin gücü düşük olacaktır. Böyle bir durumda parçalı regresyon yaklaşımı önerilmektedir. Parçalı regresyon, veri noktalarını belirlenen dönüm noktalarına göre çeşitli aralıklara bölerek her bir aralıkta uygun dereceden polinomlarla yaklaşım yapma esasına dayanır (Poirier, 1973; Seber ve Wild, 1988; Chan, 1989; Karadavut ve ark, 2004).

### 2.2.1 Dönüm Noktası

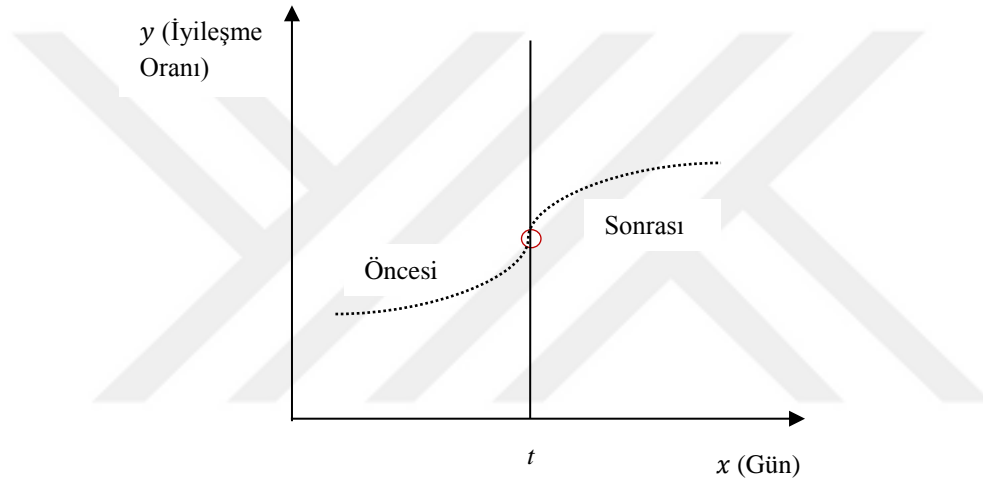
Noktasal dağılımın parçalara ayrıldığı noktalar “dönüm” olarak adlandırılmaktadır. (Marsh ve Cormier, 2002). Dönüm noktası araştırmacı tarafından araştırmanın başlangıcında (sabit dönüm) ya da araştırma sonunda elde edilen noktasal dağılımın incelenmesiyle (değişken

dönüm) belirlenir. İlk dönüm noktasından itibaren dönümler arasında oluşturulan fonksiyonlar ise  $d$ 'inci dereceden polinomlardır (Hawkins ve Douglas, 1980; Ruppert, 2002).

### 2.2.1.1 Sabit Dönüm Noktası

Dönüm noktası araştırmacı tarafından denemenin başında belirleniyorsa bu tür dönümler “sabit dönüm noktası” olarak adlandırılır (DeBoor ve Rice, 1968; Smith, 1979; Eubank, 1999).

Örneğin; bir hastalık nedeniyle kullanılan “A” ilacının, hastalığın günlük iyileşme düzeyini  $t$ 'inci günden itibaren ne kadar ve hangi yönde etkilediği inceleniyorsa, araştırmacı tarafından belirlenen  $t$ 'inci gün sabit dönüm noktası olarak seçilmelidir (Şekil 2.3).



Şekil 2.3 Parçalı regresyonda sabit dönüm noktası

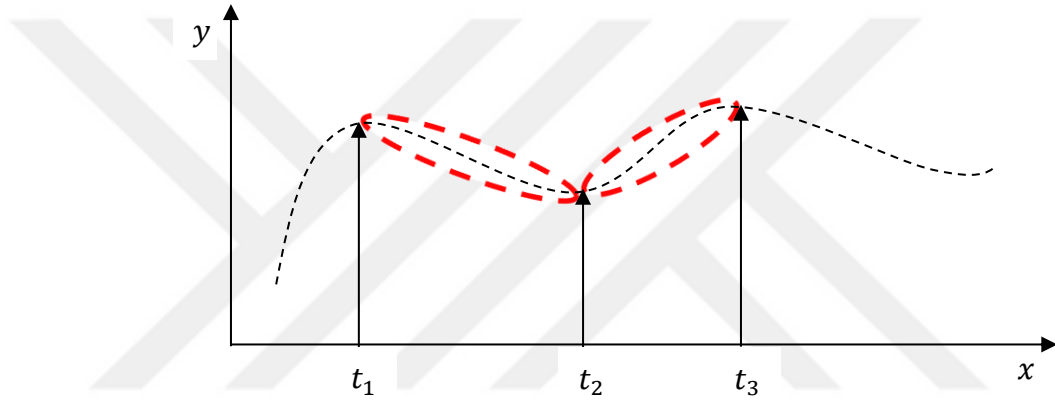
### 2.2.1.2 Değişken Dönüm Noktası

Dönüm noktası önceden bilinmiyorsa ve araştırma sonucunda elde edilen noktasal dağılımın incelenmesiyle belirleniyorsa bu tür dönümler “değişken dönüm noktası” olarak adlandırılır (DeBoor ve Rice, 1968; Wold, 1974; Eubank, 1999). Sağlık alanında yapılan çalışmalardan elde edilen verilere ait dağılımlarda doğrusal bir parçalama yapılacaksa uygun yöntemlerle dönüm noktasının yerini tam olarak belirlemek mümkündür. Ancak çalışmalarda genellikle dağılımlar doğrusal değildir. Bu nedenle dönüm noktasının kesin olarak bilinmesi zordur (Agarwal ve Studden, 1978; Lindstrom ve Marry, 1999; Eubank, 1999).

Dönüm noktasının yerinin ve sayısının tespit edilmesi, dağılımın şekli ile yakından ilgilidir. Parçalı regresyonlarda dönüm noktasının yaklaşık yeri ve sayısı genellikle görsel incelemeyle tespit edilebilir (Marsh ve Cormier, 2002; Seber ve Wild, 2003). Noktasal dağılımın görsel olarak incelenmesi; uygulanacak parçalama şekli, nasıl bir fonksiyonun

(karesel ya da kübik) uygulanacağı ve noktasal dağılımın kaç parçaya ayrılarak inceleneceği konularında araştırmacıya önemli ipuçları verir (Barrodale ve Young, 1996).

Noktasal dağılımı incelerken dikkat edilecek husus, ani yön değişimlerinin olduğu bölgelerin muhtemel dönüm bölgeleri olmasıdır. Ayrıca parçalamada kullanılacak fonksiyon da, noktasal dağılımın neresinde dönüm noktası ya da noktalarının oluşturulacağı konusunda belirleyici olacaktır (Wold, 1974; Eubank, 1984). Şekil 2.4’de görüldüğü gibi hem aşağı bükeyliğin başlangıç noktası ile bitiş noktası aralığı ( $t_1-t_2$  aralığı), hem de yukarı bükeyliğin başlangıç noktası ile bitiş noktası aralığı ( $t_2-t_3$  aralığı) parçalama için aday dönüm bölgeleridir ve bu bölgeler içerisinden veriyi en iyi temsil eden, en uygun sayıda dönüm noktası seçimi yapılabilir. (Şahin, 2009).



Şekil 2.4 Parçalı regresyon aday dönüm bölgeleri

### 2.2.2 “+” Fonksiyonları

Parçalı regresyonları oluşturabilmek için “+” fonksiyonları yaygın olarak kullanılmaktadır. “+” fonksiyonları ile belirlenen dönüm noktalarına göre regresyon modelleri parçalara ayrılabilir. Dönüm noktası  $t$ , parçalı fonksiyona ait bağımsız değişken  $(x - t)$  olmak üzere  $(x - t)_+$  ile ifade edilen bir “+” fonksiyonu,

$$(x - t)_+ = \begin{cases} x - t, & x > t \\ 0, & x \leq t \end{cases}$$

şeklinde tanımlanır. Bu fonksiyon,  $t$  dönüm noktasına eşit ya da  $t$  dönüm noktasından önceki bağımsız değişken değerleri için “0” değerini alırken,  $t$  dönüm noktasından sonraki bağımsız değişken değerleri için “ $x - t$ ” değerine eşit olur. Böylece “+” fonksiyonu ile verilen bir ifade, parçalı regresyon modelinin ilgili dönüm noktasından önceki kısmına etki etmez (Powell, 1969; Smith, 1979; Wegman ve Wright, 1983).

### 2.2.3 Parçalı Regresyon Modelleri

Noktasal dağılımların farklı bölgeleri, aynı fonksiyona ya da farklı fonksiyonlara ait dağılımlar gösterirken, belirlenen herhangi bir dönüm noktasından önce ve sonra oluşturulan fonksiyonların bu dönüm noktasında aynı “ $y_i$ ” değerini vermesi durumu “sürekli parçalı regresyon” olarak adlandırılır (Poirier, 1973; Chan, 1989; Oturaç, 2008).

Parçalı regresyonları oluşturan her bir parçanın süreksizliğe neden olan kendine ait sabit terimi vardır. Parçalı regresyon modeline süreklilik kısıtlamaları getirilerek sürekliliği bozan katsayılar modelden çıkarılabilir (Alkan, 2009).

Herhangi bir kısıtlama getirilmemiş parçalı regresyon genel formu aşağıdaki gibidir (Smith, 1979; Marsh, 1989; Wegman ve Wright, 1983; Markov, 2002):

$$S(x_i) = y_i = \sum_{j=0}^k b_{0j}x_i^j + \sum_{l=1}^m \sum_{v=0}^d b_{lv}(x_i - t_l)_+^v + e_i \quad i = 1, 2, \dots, n$$

$x$ : Bağımsız değişken,

$t$ : Dönüm noktası,

$b_{0j}$ : Regresyon katsayısı,

$b_{lv}$ : “+” fonksiyonunun regresyon katsayısı,

$k$ : Bağımsız değişkenin derecesi,

$d$ : “+” fonksiyonunun derecesi,

$e$ : Hata terimi,

$m$ : Dönüm sayısı,

$n$ : Örneklem birim sayısı

Ayrıca,

$$(x - t)_+ = \begin{cases} (x - t), & x > t \\ 0, & x \leq t \end{cases}$$

koşulu vardır ve  $k \leq d$ 'dir (Smith, 1979). Yukarıdaki denklem başka bir ifade ile,

$$\begin{aligned} S(x_i) = & b_{00} + b_{01}x_i + \dots + b_{0k}x_i^k \\ & + [b_{10}(x_i - t_1)_+^0 + b_{11}(x_i - t_1)_+^1 + b_{12}(x_i - t_1)_+^2 + \dots + b_{1d}(x_i - t_1)_+^d] \\ & + [b_{20}(x_i - t_2)_+^0 + b_{21}(x_i - t_2)_+^1 + b_{22}(x_i - t_2)_+^2 + \dots + b_{2d}(x_i - t_2)_+^d] + \dots \\ & + [b_{m0}(x_i - t_m)_+^0 + b_{m1}(x_i - t_m)_+^1 + b_{m2}(x_i - t_m)_+^2 + \dots + b_{md}(x_i - t_m)_+^d] \\ & + e_i \end{aligned}$$

olarak gösterilebilir. Bu denkleme süreklilik kısıtlamaları getirilip  $b_{10}, b_{20}, \dots, b_{m0}$  sabitleri modelden çıkarılırsa model aşağıdaki gibi olur (Studden ve ark, 1969; Smith, 1979; Wegman ve Wright, 1983; Markov, 2002):

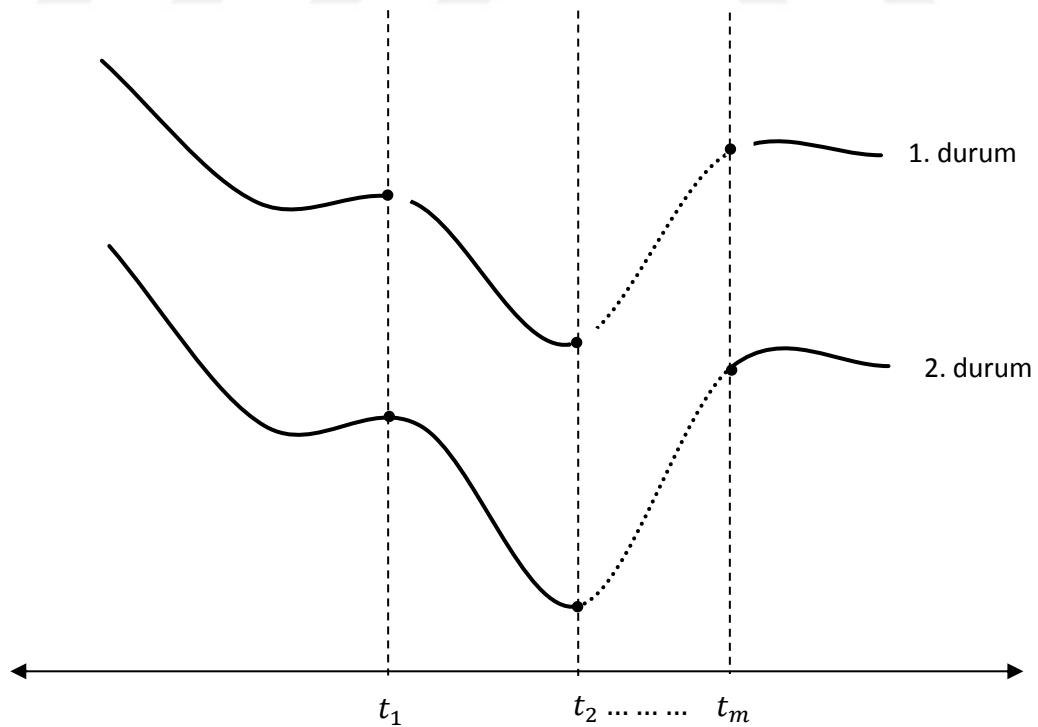
$$S(x_i) = y_i = \sum_{j=0}^k b_{0j}x_i^j + \sum_{l=1}^m \sum_{v=1}^d b_{lv}(x_i - t_l)_+^v + e_i$$

Başka bir ifade ile,

$$\begin{aligned} S(x_i) = & b_{00} + b_{01}x_i + \dots + b_{0k}x_i^k + [b_{11}(x_i - t_1)_+ + b_{12}(x_i - t_1)_+^2 + \dots + b_{1d}(x_i - t_1)_+^d] \\ & + [b_{21}(x_i - t_2)_+ + b_{22}(x_i - t_2)_+^2 + \dots + b_{2d}(x_i - t_2)_+^d] + \dots \\ & + [b_{m1}(x_i - t_m)_+ + b_{m2}(x_i - t_m)_+^2 + \dots + b_{md}(x_i - t_m)_+^d] \\ & + e_i \end{aligned}$$

olarak yazılabilir. Böylece model sürekli bir hale gelmiş olur (Smith, 1979; Alkan, 2009).

Şekil 2.5'deki dağılım, 1. durumda parçalı regresyon genel modelini temsil etmektedir ve süreksizdir. 2. durumda ise yukarıda açıklanan süreklilik kısıtlamaları uygulanmış ve sürekli hale getirilmiş parçalı regresyon modelini temsil etmektedir (Smith, 1979).



Şekil 2.5 Parçalı regresyon genel dağılımında süreklilik kısıtlamaları

Her bir  $\{(x_i; S(x_i)): i = 1, \dots, n\}$  veri çifti yukarıdaki sürekli hale getirilen denklemde yerine konulduğunda aşağıdaki sürekli parçalı denklemler elde edilir (Ramsey, 1969; Freeman ve Skapura, 1991):

$$\begin{aligned}
S(x_1) &= b_{00} + b_{01}x_1 + \dots + b_{0k}x_1^k + [b_{11}(x_1 - t_1)_+ + b_{12}(x_1 - t_1)_+^2 + \dots + b_{1d}(x_1 - t_1)_+^d] \\
&\quad + [b_{21}(x_1 - t_2)_+ + b_{22}(x_1 - t_2)_+^2 + \dots + b_{2d}(x_1 - t_2)_+^d] + \dots \\
&\quad + [b_{m1}(x_1 - t_m)_+ + b_{m2}(x_1 - t_m)_+^2 + \dots + b_{md}(x_1 - t_m)_+^d] + e_1 \\
S(x_2) &= b_{00} + b_{01}x_2 + \dots + b_{0k}x_2^k + [b_{11}(x_2 - t_1)_+ + b_{12}(x_2 - t_1)_+^2 + \dots + b_{1d}(x_2 - t_1)_+^d] \\
&\quad + [b_{21}(x_2 - t_2)_+ + b_{22}(x_2 - t_2)_+^2 + \dots + b_{2d}(x_2 - t_2)_+^d] + \dots \\
&\quad + [b_{m1}(x_2 - t_m)_+ + b_{m2}(x_2 - t_m)_+^2 + \dots + b_{md}(x_2 - t_m)_+^d] + e_2 \\
&\quad \vdots \\
S(x_n) &= b_{00} + b_{01}x_n + \dots + b_{0k}x_n^k + [b_{11}(x_n - t_1)_+ + b_{12}(x_n - t_1)_+^2 + \dots + b_{1d}(x_n - t_1)_+^d] \\
&\quad + [b_{21}(x_n - t_2)_+ + b_{22}(x_n - t_2)_+^2 + \dots + b_{2d}(x_n - t_2)_+^d] + \dots \\
&\quad + [b_{m1}(x_n - t_m)_+ + b_{m2}(x_n - t_m)_+^2 + \dots + b_{md}(x_n - t_m)_+^d] + e_n
\end{aligned}$$

Elde edilen denklemleri matris formunda göstermek için kullanılacak vektör ve matrisleri tanımlayalım:

Bağımlı değişken vektörü:

$$\mathbf{S} = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{bmatrix}$$

Regresyon katsayı vektörü:

$$\mathbf{b}_0 = \begin{bmatrix} b_{00} \\ b_{01} \\ \vdots \\ b_{0k} \end{bmatrix}$$

Bağımsız değişken matrisi:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^k \\ 1 & x_2 & x_2^2 & \dots & x_2^k \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^k \end{bmatrix}$$

“+” fonksiyonunun regresyon katsayı vektörleri:

$$\mathbf{b}_1 = \begin{bmatrix} b_{11} \\ b_{12} \\ \vdots \\ b_{1d} \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} b_{21} \\ b_{22} \\ \vdots \\ b_{2d} \end{bmatrix}, \dots, \mathbf{b}_m = \begin{bmatrix} b_{m1} \\ b_{m2} \\ \vdots \\ b_{md} \end{bmatrix}$$

“+” fonksiyonu matrisleri:

$$\mathbf{T}_1 = \begin{bmatrix} (x_1 - t_1)_+ & (x_1 - t_1)_+^2 & \dots & (x_1 - t_1)_+^d \\ (x_2 - t_1)_+ & (x_2 - t_1)_+^2 & \dots & (x_2 - t_1)_+^d \\ \vdots & \vdots & \dots & \vdots \\ (x_n - t_1)_+ & (x_n - t_1)_+^2 & \dots & (x_n - t_1)_+^d \end{bmatrix},$$

$$\mathbf{T}_2 = \begin{bmatrix} (x_1 - t_2)_+ & (x_1 - t_2)_+^2 & \dots & (x_1 - t_2)_+^d \\ (x_2 - t_2)_+ & (x_2 - t_2)_+^2 & \dots & (x_2 - t_2)_+^d \\ \vdots & \vdots & \dots & \vdots \\ (x_n - t_2)_+ & (x_n - t_2)_+^2 & \dots & (x_n - t_2)_+^d \end{bmatrix},$$

$$\mathbf{T}_m = \begin{bmatrix} (x_1 - t_m)_+ & (x_1 - t_m)_+^2 & \dots & (x_1 - t_m)_+^d \\ (x_2 - t_m)_+ & (x_2 - t_m)_+^2 & \dots & (x_2 - t_m)_+^d \\ \vdots & \vdots & \dots & \vdots \\ (x_n - t_m)_+ & (x_n - t_m)_+^2 & \dots & (x_n - t_m)_+^d \end{bmatrix}$$

Hata terimi matrisi:

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

Yukarıdaki vektör ve matrisler yardımıyla oluşturulan parçalı regresyon modeli,

$$\mathbf{S} = \mathbf{x}\mathbf{b}_0 + [\mathbf{T}_1\mathbf{b}_1 + \mathbf{T}_2\mathbf{b}_2 + \dots + \mathbf{T}_m\mathbf{b}_m] + \mathbf{e}$$

biçiminde gösterilir. Eşitlikte  $\mathbf{B} = \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_m \end{bmatrix}$  ve  $\mathbf{X}=(\mathbf{x}, \mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_m)$  olarak yazılırsa model,

$$\mathbf{S} = \mathbf{XB} + \mathbf{e}$$

olarak tanımlanır. Sonuç olarak yukarıdaki eşitlik ile denklem birinci dereceden bir regresyon modeli haline getirilmiştir. Model parametresi en küçük kareler tahmin edicisi,

$$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}$$

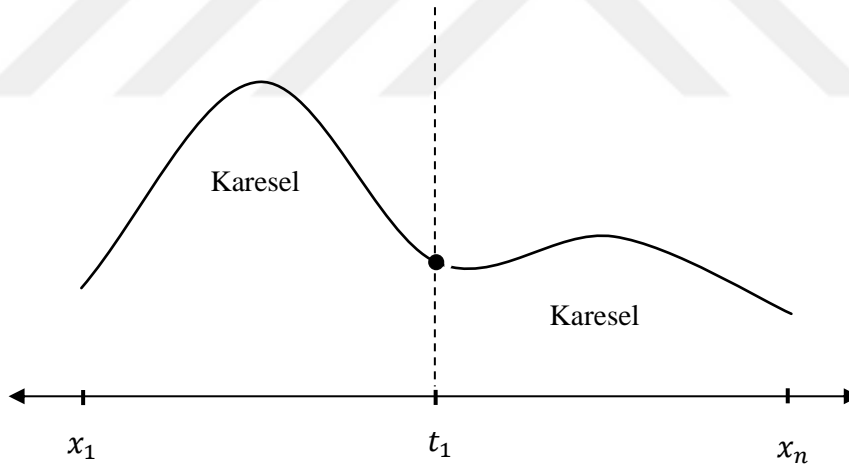
ile tahmin edilir (Alkan, 2002; Ruppert, 2002).

### 2.2.3.1 Karesel Parçalı Regresyon

Parçalı bir regresyon modelini oluşturan tüm kısmi parçaların karesel yapıda olduğu regresyon türü karesel parçalı regresyon olarak adlandırılır (Smith, 1979). İkinci dereceden, herhangi bir süreklilik kısıtlaması getirilmemiş,  $m$  tane dönüm noktasına sahip karesel parçalı regresyon genel denklemi,

$$S(x_i) = y_i = \sum_{j=0}^2 b_{0j}x_i^j + \sum_{l=1}^m \sum_{v=0}^2 b_{lv}(x_i - t_l)_+^v + e_i \quad i = 1, 2, \dots, n$$

olarak ifade edilir (Smith, 1979; Wegman ve Wright, 1983; Marsh, 1989). Verilen  $[x_1, x_n]$  aralığından seçilen  $t_1$  gibi bir dönüm noktasına ve karesel+karesel yapıda iki kısmi parçaya ayrılan bir dağılım Şekil 2.6'daki gibi gösterilir.



Şekil 2.6 Karesel+Karesel yapıda  $[x_1, x_n]$  aralığında  $t_1$  dönümü

Şekil 2.6'daki gibi tek dönüm noktası ve karesel+karesel yapıda iki kısmi parça içeren parçalı regresyon denklemi aşağıdaki gibi yazılır (Smith, 1979; Wegman ve Wright, 1983):

$$S(x_i) = \sum_{j=0}^2 b_{0j}x_i^j + \sum_{v=0}^2 b_{1v}(x_i - t_1)_+^v + e_i$$



Bu denklemde  $(x_i - t_1)_+$  fonksiyonu,

$$(x_i - t_1)_+ = \begin{cases} (x_i - t_1), & x_i > t_1 \\ 0, & x_i \leq t_1 \end{cases}$$

değerini alır (Smith, 1979; Eubank, 1984). Bu eşitlik için oluşturulan aralık fonksiyonları,

$$S(x) = \begin{cases} b_{00} + b_{01}x + b_{02}x^2 & , x \leq t_1 \\ b_{00} + b_{01}x + b_{02}x^2 + b_{10} + b_{11}(x - t_1) + b_{12}(x - t_1)^2 & , t_1 < x \end{cases}$$

olarak yazılır (Freund ve ark, 2006; Harrell, 2017). Buradan tüm noktasal dağılım için, yani  $[x_1, x_n]$  aralığı için regresyon denklemi,

$$S(x_i) = b_{00} + b_{01}x_i + b_{02}x_i^2 + b_{10}(x_i - t_1)_+^0 + b_{11}(x_i - t_1)_+^1 + b_{12}(x_i - t_1)_+^2 + e_i$$

olarak yazılır (Smith, 1979; Marsh, 1989). Yukarıdaki model için süreklilik kısıtlaması getirilip sabit katsayısı kısıtlanmış bir model elde edilebilir. Dolayısıyla  $b_{10}$  katsayısı çıkarılarak aşağıdaki denklem elde edilir (Smith, 1979; Wegman ve Wright, 1983):

$$S(x_i) = \sum_{j=0}^2 b_{0j}x_i^j + \sum_{v=1}^2 b_{1v}(x_i - t_1)_+^v + e_i$$

Böyle bir durumda aralık fonksiyonları,

$$S(x) = \begin{cases} b_{00} + b_{01}x + b_{02}x^2 & , x \leq t_1 \\ b_{00} + b_{01}x + b_{02}x^2 + b_{11}(x - t_1) + b_{12}(x - t_1)^2 & , t_1 < x \end{cases}$$

olarak ifade edilir (Freund ve ark, 2006; Harrell, 2017). Böylece tüm noktasal dağılım için sabit katsayısı kısıtlanmış, tek dönüm noktalı, karesel+karesel yapıda iki kısmi parça içeren aşağıdaki parçalı regresyon denklemi elde edilir ve model sürekli bir hale getirilmiş olur (Smith, 1979; Wegman ve Wright, 1983):

$$S(x_i) = b_{00} + b_{01}x_i + b_{02}x_i^2 + b_{11}(x_i - t_1)_+^1 + b_{12}(x_i - t_1)_+^2 + e_i$$

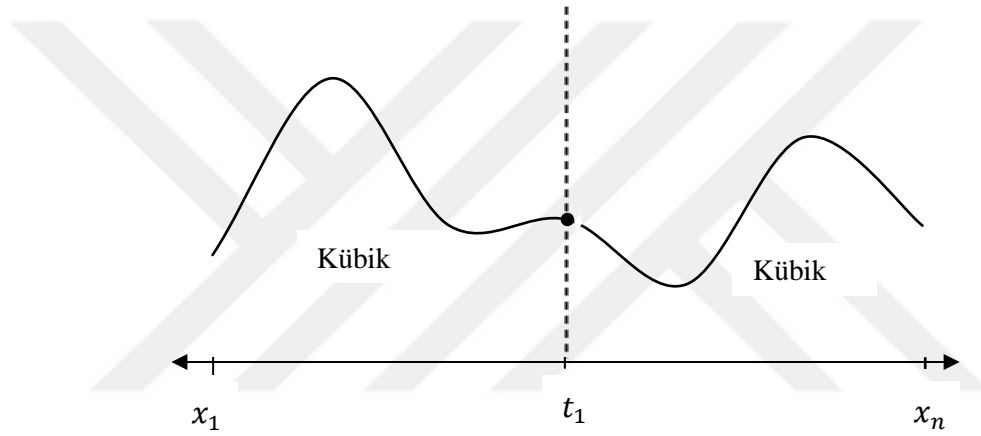
### 2.2.3.2 Kübik Parçalı Regresyon

Parçalı bir regresyon modelini oluşturan tüm kısmi parçaların kübik yapıda olduğu regresyon türü kübik parçalı regresyon olarak adlandırılır (Smith, 1979). Kübik parçalı fonksiyonu ilk olarak 1957 yılında Holladay tarafından tanıtılmıştır (Moreton, 1992). Kübik parçalı fonksiyonda, noktasal dağılımın her alt aralığındaki noktalardan geçen fonksiyonlar,

üçüncü dereceden polinom olarak seçilmektedir (Oturanç ve ark, 2008; Şahin, 2009). Herhangi bir süreklilik kısıtlaması getirilmemiş, m tane dönüm noktasına sahip kübik parçalı regresyon genel denklemi,

$$S(x_i) = y_i = \sum_{j=0}^3 b_{0j}x_i^j + \sum_{l=1}^m \sum_{v=0}^3 b_{lv}(x_i - t_l)_+^v + e_i \quad i = 1, 2, \dots, n$$

olarak ifade edilir (Smith, 1979; Wegman ve Wright, 1983). Verilen  $[x_1, x_n]$  aralığından seçilen  $t_1$  gibi bir dönüm noktasına ve kübik+kübik yapıda iki kısmi parçaya ayrılan bir dağılım Şekil 2.7'deki gibi gösterilir.



Şekil 2.7 Kübik+kübik yapıda  $[x_1, x_n]$  aralığında  $t_1$  dönümü

Şekil 2.7'deki gibi tek dönüm noktalı ve kübik+kübik yapıda iki kısmi parça içeren parçalı regresyon denklemi aşağıdaki gibi yazılır (Smith, 1979; Marsh, 1989; Berberoğlu, 2011):

$$S(x_i) = \sum_{j=0}^3 b_{0j}x_i^j + \sum_{v=0}^3 b_{1v}(x_i - t_1)_+^v + e_i$$

Bu denklemde  $(x_i - t_1)_+$  fonksiyonu,

$$(x_i - t_1)_+ = \begin{cases} (x_i - t_1), & x_i > t_1 \\ 0, & x_i \leq t_1 \end{cases}$$

değerini alır (Powell, 1969; Smith, 1979; Eubank, 1984). Bu eşitlik için oluşturulan aralık fonksiyonları,

$$S(x) = \begin{cases} b_{00} + b_{01}x + b_{02}x^2 + b_{03}x^3 & , x \leq t_1 \\ b_{00} + b_{01}x + b_{02}x^2 + b_{03}x^3 + b_{10} + b_{11}(x - t_1) + b_{12}(x - t_1)^2 + b_{13}(x - t_1)^3 & , t_1 < x \end{cases}$$

olarak yazılır (Freund ve ark, 2006; Harrell 2017). Buradan tüm noktasal dağılım için denklem,

$$S(x_i) = b_{00} + b_{01}x_i + b_{02}x_i^2 + b_{03}x_i^3 + b_{10}(x_i - t_1)_+^0 + b_{11}(x_i - t_1)_+^1 + b_{12}(x_i - t_1)_+^2 + b_{13}(x_i - t_1)_+^3 + e_i$$

olarak yazılır. Yukarıdaki model için süreklilik kısıtlaması getirilerek sabit katsayısı kısıtlanmış bir model elde edilebilir. Dolayısıyla  $b_{10}$  katsayısı çıkarılarak aşağıdaki denklem elde edilir:

$$S(x_i) = \sum_{j=0}^3 b_{0j}x_i^j + \sum_{v=1}^3 b_{1v}(x_i - t_1)_+^v + e_i$$

Böyle bir durumda aralık fonksiyonları,

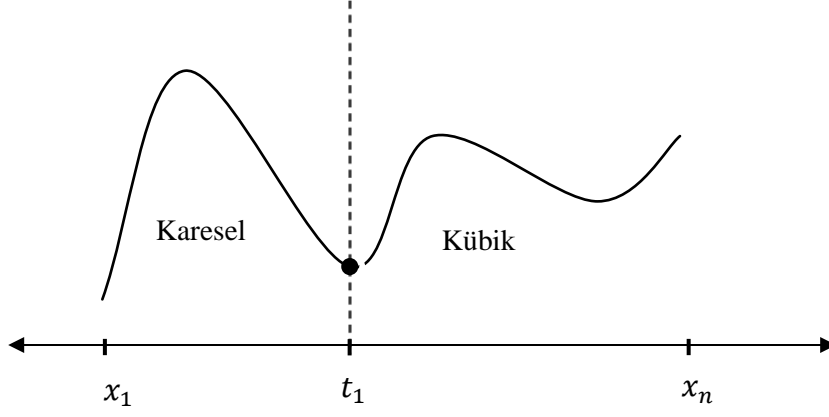
$$S(x) = \begin{cases} b_{00} + b_{01}x + b_{02}x^2 + b_{03}x^3 & , x \leq t_1 \\ b_{00} + b_{01}x + b_{02}x^2 + b_{03}x^3 + b_{11}(x - t_1) + b_{12}(x - t_1)^2 + b_{13}(x - t_1)^3 & , t_1 < x \end{cases}$$

olarak ifade edilebilir (Freund ve ark, 2006; Harrell 2017). Böylece tüm noktasal dağılım için sabit katsayısı kısıtlanmış, tek dönüm noktalı, kübik+kübik yapıda iki kısmi parça içeren aşağıdaki parçalı regresyon denklemi elde edilir ve model sürekli bir hale getirilmiş olur (Smith, 1979; Berberoğlu, 2011):

$$S(x_i) = b_{00} + b_{01}x_i + b_{02}x_i^2 + b_{03}x_i^3 + b_{11}(x_i - t_1)_+^1 + b_{12}(x_i - t_1)_+^2 + b_{13}(x_i - t_1)_+^3 + e_i$$

### 2.2.3.3 Karma Parçalı Regresyon

Bazı durumlarda noktasal dağılımın ilk bölgesi farklı dereceden, diğer bölgeleri daha farklı dereceden parçalar ile ifade edilebilir (Smith, 1979). Verilen  $[x_1, x_n]$  aralığından seçilen  $t_1$  dönüm noktasına sahip böyle bir dağılım Şekil 2.8'deki gibi gösterilebilir.



Şekil 2.8 Karesel+kübik yapıda  $[x_1, x_n]$  aralığında  $t_1$  dönümü

Şekil 2.8'deki gibi tek dönüm noktalı ve karesel+kübik yapıda iki kısmi parça içeren parçalı regresyon denklemi aşağıdaki gibi yazılabilir:

$$S(x_i) = y_i = \sum_{j=0}^2 b_{0j}x_i^j + \sum_{v=0}^3 b_{1v}(x_i - t_1)_+^v + e_i \quad i = 1, 2, \dots, n$$

Bu denklemde  $(x_i - t_1)_+$  fonksiyonu,

$$(x_i - t_1)_+ = \begin{cases} (x_i - t_1), & x_i > t_1 \\ 0, & x_i \leq t_1 \end{cases}$$

değerini alır (Barrodale ve Young, 1966; Smith, 1979; Eubank, 1984). Böyle bir durumda tek dönüm noktası içeren bir model için aralık fonksiyonları,

$$S(x) = \begin{cases} b_{00} + b_{01}x + b_{02}x^2 & , x \leq t_1 \\ b_{00} + b_{01}x + b_{02}x^2 + b_{10} + b_{11}(x - t_1) + b_{12}(x - t_1)^2 + b_{13}(x - t_1)^3 & , t_1 < x \end{cases}$$

olarak yazılır (Marsh, 1989; Freund ve ark, 2006). Buradan tüm noktasal dağılım için denklem,

$$S(x_i) = b_{00} + b_{01}x_i + b_{02}x_i^2 + b_{10}(x_i - t_1)_+^0 + b_{11}(x_i - t_1)_+^1 + b_{12}(x_i - t_1)_+^2 + b_{13}(x_i - t_1)_+^3 + e_i$$

olarak yazılır. Yukarıdaki modele süreklilik kısıtlaması getirilerek  $b_{10}$  katsayısı çıkarılırsa aşağıdaki denklem elde edilir:

$$S(x_i) = \sum_{j=0}^2 b_{0j}x_i^j + \sum_{v=1}^3 b_{1v}(x_i - t_1)_+^v + e_i$$

Böyle bir durumda tek dönüm noktası içeren karma yapıdaki bir model için aralık fonksiyonları,

$$S(x) = \begin{cases} b_{00} + b_{01}x + b_{02}x^2 & , x \leq t_1 \\ b_{00} + b_{01}x + b_{02}x^2 + b_{11}(x - t_1) + b_{12}(x - t_1)^2 + b_{13}(x - t_1)^3 & , t_1 < x \end{cases}$$

şeklinde ifade edilebilir. Böylece tüm noktasal dağılım için sabit katsayısı kısıtlanmış, tek dönüm noktalı, karesel+kübik yapıda iki kısmi parça içeren aşağıdaki parçalı regresyon denklemi elde edilir ve model sürekli bir hale getirilmiş olur (Marsh, 1989; Sheather, 2009):

$$S(x_i) = b_{00} + b_{01}x_i + b_{02}x_i^2 + b_{11}(x_i - t_1)_+^1 + b_{12}(x_i - t_1)_+^2 + b_{13}(x_i - t_1)_+^3 + e_i$$

### 2.3 Model Değerlendirme Kriterleri

Bu tez çalışmasında modellerin performanslarını karşılaştırmak için dört kriter kullanılmıştır. Bu kriterler  $R^2$ , HKO, ABK ve BBK'dir.

#### 2.3.1 Belirtme Katsayısı

$R^2$ , oluşturulan modelde bağımlı değişkendeki değişimin yüzde kaçının bağımsız değişken tarafından açıklanabildiğinin ifadesidir. 0 ve 1 aralığında değerler alabilen ve 1'e yaklaştığında tahmin edilen regresyon eğrisinin verilere uyumunun daha iyi olduğunu gösteren  $R^2$  değeri aşağıdaki denklem ile elde edilir (Gujarati, 1999; Young, 2000; Günel, 2003; Ucal, 2006):

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Burada,

y: Gözlenen değerleri,

$\hat{y}$ : Tahmin edilen değerleri ifade eder.

#### 2.3.2 Hata Kareler Ortalaması

HKO, hata kareler toplamının hataya ait serbestlik derecesine bölünmesiyle elde edilir ve polinom regresyon analizi için aşağıdaki gibi hesaplanır (Allen, 1971; Mood ve ark, 1974; Lehmann ve ark, 1998; Başar ve Oktay, 2004):

$$HKO = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - (k + 1)}$$

Parçalı regresyon modeli için ise HKO aşağıdaki denklem ile elde edilir:

$$HKO = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - (k + md + 1)}$$

Burada,

$y$ : Gözlenen değerleri,

$\hat{y}$ : Tahmin edilen değerleri,

$k + 1$ : Sabit terim dahil polinom regresyonun parametre sayısını,

$k + md + 1$ : Sabit terim dahil parçalı regresyonun parametre sayısını,

$n$ : Örneklem birim sayısını,

$m$ : Dönüm sayısını gösterir.

Hata kareler ortalamasının düşük olması istenen bir durumdur. Çünkü hata kareler ortalamasının düşük olması, regresyona ait F test istatistiğinin yüksek çıkmasına ve dolayısı ile “regresyon modeli anlamlı değildir” şeklinde kurulan  $H_0$  hipotezinin büyük bir F değeri ile reddedilmesi anlamına gelir. Diğer bir ifade ile, oluşturulan modelin veri setini temsil etme düzeyinin yüksek olduğu söylenebilir (Gujarati, 1999; Başar ve Oktay, 2004; Lee ve Brooks, 2006).

### 2.3.3 Akaike Bilgi Kriteri

ABK farklı modeller arasında en uygununu seçmeye yarayan bir kriterdir. Akaike tarafından 1974 yılında literatüre kazandırılmış olan ABK değeri daha küçük olan modelin daha uygun olduğu ifade edilir (Akaike, 1974; Pan, 2001; Aladağ ve Eğrioglu, 2010)

ABK'yı polinom regresyon analizi için aşağıdaki formülle hesaplamak mümkündür (Gujarati, 1999; Geyer, 2003; Larget, 2003):

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

Parçalı regresyon analizi için ise ABK aşağıdaki denklem ile elde edilir (Geyer, 2003; Larget, 2003):

$$ABK = n \left[ 1 + \ln 2\pi + \ln \left( \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \right) \right] + 2(k + md + 1)$$

### 2.3.4 Bayes Bilgi Kriteri

Bu ölçüt 1978 yılında Gideon Schwarz tarafından önerildiği için Schwarz ölçütü olarak da bilinir. BBK'yı polinom regresyon analizi için aşağıdaki formülle hesaplamak mümkündür (Schwarz ve Gideon, 1978; Larget, 2003):

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

Parçalı regresyon analizi için ise BBK aşağıdaki denklem ile elde edilir (Geyer, 2003, Larget, 2003):

$$BBK = n \left[ 1 + \ln 2\pi + \ln \left( \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \right) \right] + \ln(n) (k + md + 1)$$

Model karşılaştırmalarında her zaman en küçük BBK değerini veren model tercih edilir (Gujarati, 1999; Aladağ ve Eğrioğlu, 2013; Ucal, 2006).

## GEREÇ VE YÖNTEM

### 3.1 Kullanılan Programlar

Çalışmada analizler R 3.3.1 programında gerçekleştirildi. Veri setlerinin türetilmesinde R programlama dilinden yararlanıldı ve *stats*, *plyr*, *reshape*, *ggplot2*, *ggthemes* paketleri kullanıldı.

### 3.2 Simülasyon Uygulaması

Bu çalışmada iki farklı simülasyon algoritması kullanıldı. Bu algoritmaların veri türetimi aşamasında birtakım farklılıklar bulunmaktadır. Polinom ve sürekli parçalı regresyon analiz yöntemlerinin karşılaştırılması  $n=100$  birimlik, 1000 tekrarlı simülasyon çalışması sonunda hesaplanan  $R^2$ , HKO, ABK, ve BBK değerlerine göre gerçekleştirildi.

#### 3.2.1 İkinci Dereceden Simülasyon Algoritması

1. Rassal olarak  $x \sim U(1,100)$  dağılımından bağımsız değişken değerleri türetildi.
2. Rassal olarak  $t \sim U(40,60)$  dağılımından bir tane dönüm noktası ve  $e \sim U(-15,15)$  dağılımından hata terimi değerleri türetildi.
3.  $z = x - t$  değişkeni oluşturuldu.
4.  $y = -10 + 15 * x - 0,22 * x^2 + 15 * z + 0,07 * z^2 - e$  denklemi ile  $y$  bağımlı değişkeni oluşturuldu.
5. Karesel regresyon modeli EKK yöntemi ile tahmin edildi ve bu modelin  $R^2$ , HKO, ABK ve BBK değerleri hesaplandı.
6.  $t$  dönüm noktasına göre iki parçaya ayrılan ve ilk parçası yalnızca  $x$  değişkeni ile, ikinci parçası ise  $x$  ve  $z$  değişkenleri ile oluşturulan, karesel+karesel yapıda iki kısmi parça içeren bir parçalı regresyon modeli EKK yöntemi ile tahmin edildi. Modelin  $R^2$ , HKO, ABK ve BBK değerleri hesaplandı.
7. 1. – 6. adımlar 1000 kez tekrar edildi.
8. 1000 tekrar sonucunda elde edilen  $R^2$ , HKO, ABK ve BBK değerlerinin dağılım yapısına uygun tanımlayıcı istatistikleri hesaplandı.
9.  $R^2$  değeri en büyük olan ve HKO, ABK ve BBK değerleri en küçük olan model en uygun model olarak seçildi.



### 3.2.2 Üçüncü Dereceden Simülasyon Algoritması

1. Rassal olarak  $x \sim U(1,100)$  dağılımından bağımsız değişken değerleri türetildi.
2. Rassal olarak  $t_1 \sim U(30,50)$  ve  $t_2 \sim U(51,70)$  dağılımından iki farklı dönüm noktası ve  $e \sim U(-4,4)$  dağılımından hata terimi değerleri türetildi.
3.  $z_1 = x - t_1$  ve  $z_2 = x - t_2$  değişkenleri oluşturuldu.
4.  $y = -10 + 7 * x - 0,2 * x^2 + 0,0012 * x^3 + 5,2 * z_1 - 0,02 * z_1^2 + 0,00009 * z_1^3 - 2 * z_2 + 0,01 * z_2^2 - 0,00009 * z_2^3 - e$  denklemi ile  $y$  bağımlı değişkeni oluşturuldu.
5. Kübik regresyon modeli EKK yöntemi ile tahmin edildi ve bu modelin  $R^2$ , HKO, ABK ve BBK değerleri hesaplandı.
6.  $t_1$  dönüm noktasına göre iki parçaya ayrılan ve ilk parçası yalnızca  $x$  değişkeni ile, ikinci parçası ise  $x$  ve  $z_1$  değişkenleri ile oluşturulan, karesel+kübik yapıda iki kısmi parça içeren bir parçalı regresyon modeli EKK yöntemi ile tahmin edildi. Modelin  $R^2$ , HKO, ABK ve BBK değerleri hesaplandı.
7. Kübik+kübik yapıda iki kısmi parça içeren bir parçalı regresyon modeli tahmin etmek için kullanılacak dönüm noktası  $t_3 = \frac{t_1+t_2}{2}$  olarak belirlendi.
8.  $z_3 = x - t_3$  bağımsız değişkeni oluşturuldu.
9.  $t_3$  dönüm noktasına göre iki parçaya ayrılan ve ilk parçası yalnızca  $x$  değişkeni ile, ikinci parçası ise  $x$  ve  $z_3$  değişkenleri ile oluşturulan kübik+kübik yapıda iki kısmi parça içeren bir parçalı regresyon modeli EKK yöntemi ile tahmin edildi ve oluşturulan modelin  $R^2$ , HKO, ABK ve BBK değerleri hesaplandı.
10. 1. – 9. adımlar 1000 kez tekrar edildi.
11. 1000 tekrar sonucunda elde edilen  $R^2$ , HKO, ABK ve BBK değerlerinin dağılım yapısına uygun tanımlayıcı istatistikleri hesaplandı.
12.  $R^2$  değeri en büyük olan ve HKO, ABK ve BBK değerleri en küçük olan model en uygun model olarak seçildi.

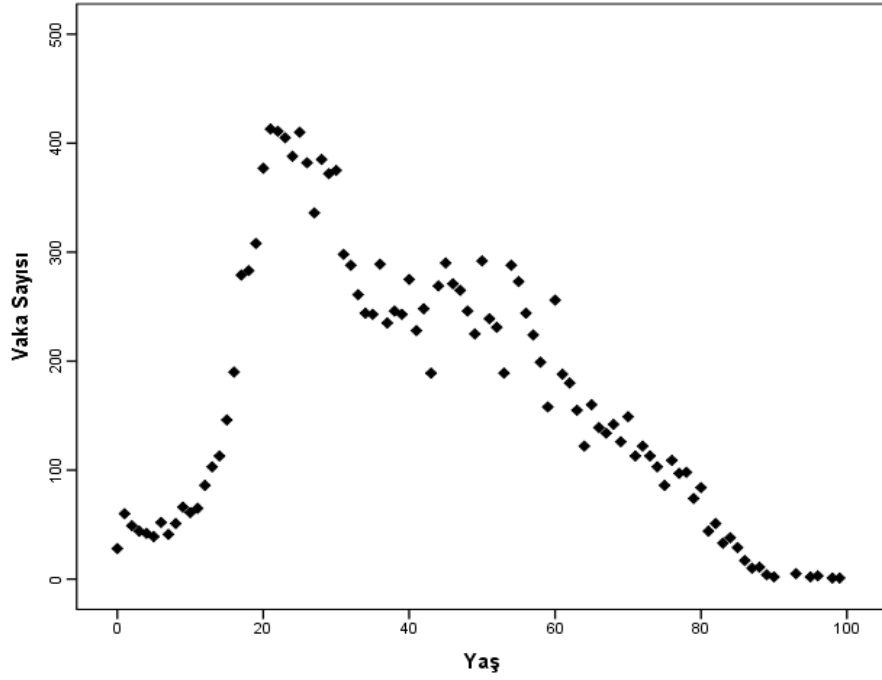
### 3.3 Gerçek Veri Seti Uygulaması

Bu kısımda tüberküloz ve kızamık vaka sayılarına ilişkin gerçek veri setleri kullanıldı. Polinom ve sürekli parçalı regresyon modellerinin parametre tahminleri yapıлып modeller oluşturuldu. Parçalı regresyon modellerinde görsel inceleme ile belirlenmiş dönüm noktaları kullanıldı. Oluşturulan regresyon modellerinin performanslarını karşılaştırmak için model

belirleme kriteri olarak kullanılacak deęerler hesaplandı.  $R^2$  deęeri en byk olan ve HKO, ABK, BBK deęerleri en kk olan model en uygun regresyon modeli olarak kabul edildi.

### 3.3.1 Tberkloz Veri Seti

alıřmada kullanılan 2010 yılına ait tberkloz veri seti, Trkiye Halk Saęlıęı Kurumu'nun "Trkiye'de Verem Savařı 2012 Raporu" bařlıklı alıřmasından alındı. Veri seti 96 birimden oluřmaktadır ve 0-99 arasında deęiřen yař deęerlerine gre toplam tberkloz vaka sayılarını iermektedir. Regresyon modellerinin oluřturulması iin bu veri setinden toplam vaka sayısı deęiřkeni baęımlı deęiřken olarak, yař deęiřkeni ise baęımsız deęiřken olarak alındı (Tablo 3.1). Yařa gre deęiřen tberkloz vaka sayılarının daęılımı Őekil 3.1'de verilmiřtir.



Őekil 3.1 Tberkloz vaka sayısının yařa gre daęılımı

Tablo 3.1 Değişen yaş değerlerine göre toplam tüberküloz vaka sayısı

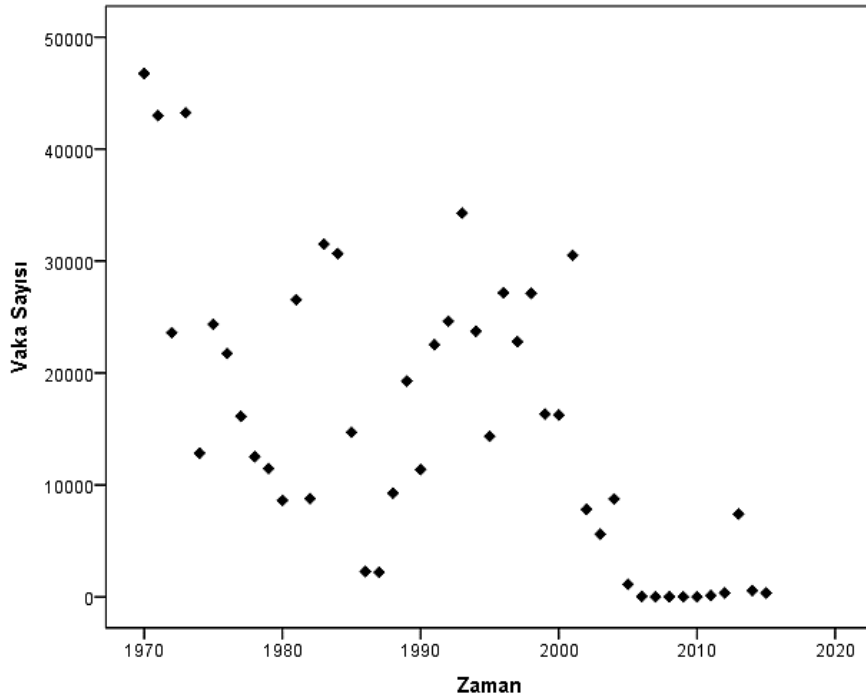
YAŞ	TOPLAM VAKA SAYISI	YAŞ	TOPLAM VAKA SAYISI	YAŞ	TOPLAM VAKA SAYISI
0	28	32	288	64	122
1	60	33	261	65	160
2	49	34	244	66	139
3	44	35	243	67	134
4	42	36	289	68	142
5	39	37	235	69	126
6	52	38	246	70	149
7	41	39	243	71	113
8	51	40	275	72	122
9	66	41	228	73	113
10	61	42	248	74	103
11	65	43	189	75	86
12	86	44	269	76	109
13	103	45	290	77	97
14	113	46	271	78	98
15	146	47	265	79	74
16	190	48	246	80	84
17	279	49	225	81	44
18	283	50	292	82	51
19	308	51	239	83	33
20	377	52	231	84	38
21	413	53	189	85	29
22	411	54	288	86	17
23	405	55	273	87	10
24	388	56	244	88	11
25	410	57	224	89	4
26	382	58	199	90	2
27	336	59	158	93	5
28	385	60	256	95	2
29	372	61	188	96	3
30	375	62	180	98	1
31	298	63	155	99	1

### 3.3.2 Kızamık Veri Seti

Çalışmada kullanılan kızamık veri seti, Türkiye Halk Sağlığı Kurumu'nun internet sitesindeki "Aşı İle Önlenebilir Hastalıklar Daire Başkanlığı İstatistiksel Verileri" sayfasından alındı. Veri seti 46 birimden oluşmaktadır ve 1970-2015 yılları arasında değişen kızamık vaka sayılarını içermektedir. Regresyon modellerinin oluşturulması için bu veri setinden vaka sayısı değişkeni bağımlı değişken olarak, zaman değişkeni ise bağımsız değişken olarak alındı (Tablo 3.2). Zamana göre değişen kızamık vaka sayılarının noktasal dağılımı Şekil 3.2'de verilmiştir.

Tablo 3.2 Zamana göre deęişen kızamık vaka sayısı

Zaman	Vaka Sayısı	Zaman	Vaka Sayısı
1970	46.761	1993	34.285
1971	43.000	1994	23.733
1972	23.601	1995	14.351
1973	43.249	1996	27.171
1974	12.836	1997	22.795
1975	24.347	1998	27.120
1976	21.740	1999	16.329
1977	16.123	2000	16.244
1978	12.517	2001	30.509
1979	11.471	2002	7.810
1980	8.618	2003	5.600
1981	26.547	2004	8.744
1982	8.778	2005	1.119
1983	31.515	2006	34
1984	30.666	2007	3
1985	14.695	2008	4
1986	2.267	2009	4
1987	2.194	2010	7
1988	9.279	2011	111
1989	19.273	2012	349
1990	11.372	2013	7.405
1991	22.521	2014	565
1992	24.626	2015	342



Şekil 3.2 Kızamık vaka sayısının zamana göre dağılımı

## 4. BULGULAR

### 4.1 Simülasyon Sonuçları

Algoritma kullanılarak yapılan 1000 tekrarlı simülasyon uygulamasında, oluşturulan regresyon modellerinin  $R^2$ , HKO, ABK ve BBK değerleri açısından karşılaştırılmaları için önce bu değerlerin normal dağılıma uygun olup olmadığı Kolmogorov-Smirnov testi ile belirlendi.

#### 4.1.1 İkinci Dereceden Türetilen Veriler ile Yapılan Simülasyon Sonuçları

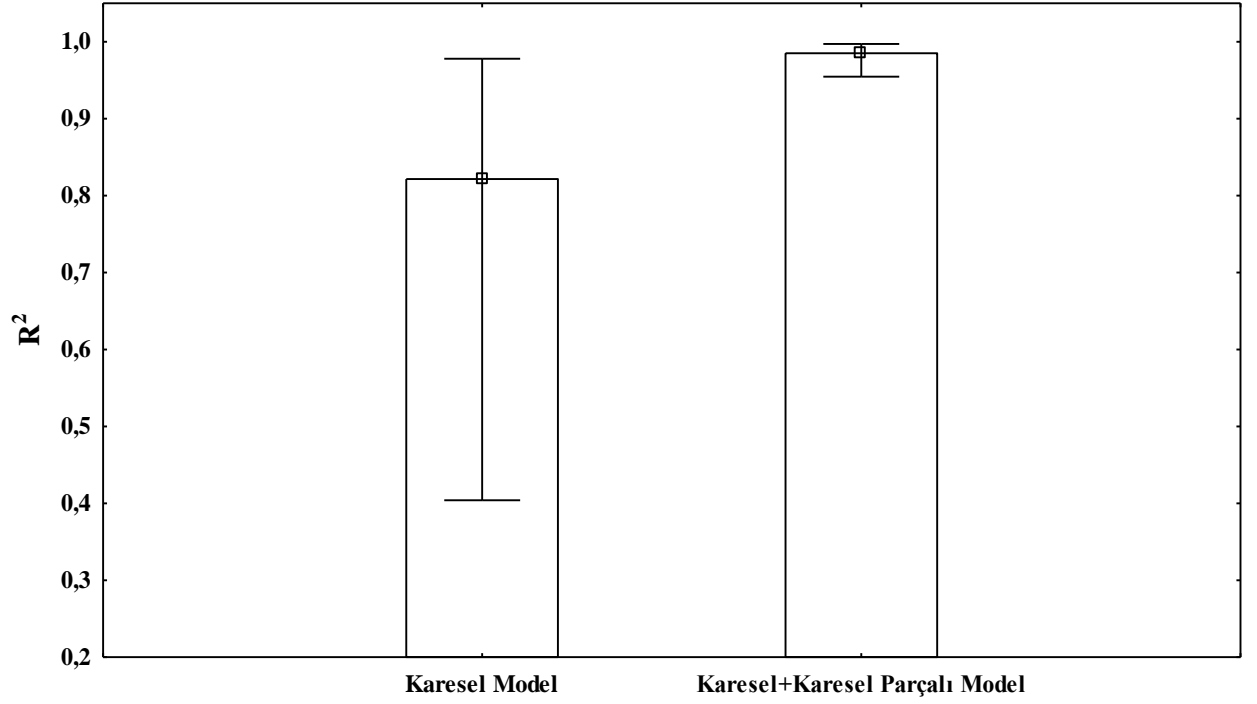
İkinci dereceden türetilen veriler ile yapılan çalışmada karesel regresyon modeline göre  $R^2$ , HKO, ABK ve BBK değerlerinin hiçbirinin normal dağılmadığı; karesel+karesel yapıdaki parçalı regresyon modeline göre  $R^2$  değerinin normal dağılmadığı, HKO, ABK ve BBK değerlerinin normal dağıldığı tespit edildi. Bu nedenle modeller arasında bu değerler bakımından fark olup olmadığının belirlenmesi amacıyla Mann Whitney U testi uygulandı. Tanımlayıcı istatistikler medyan (25. - 75. persantil) olarak belirtildi (Tablo 4.1).

Tablo 4.1 Karesel ve parçalı regresyon modellerinin  $R^2$ , HKO, ABK ve BBK değerlerinin tanımlayıcı istatistikleri ve karşılaştırma sonuçları

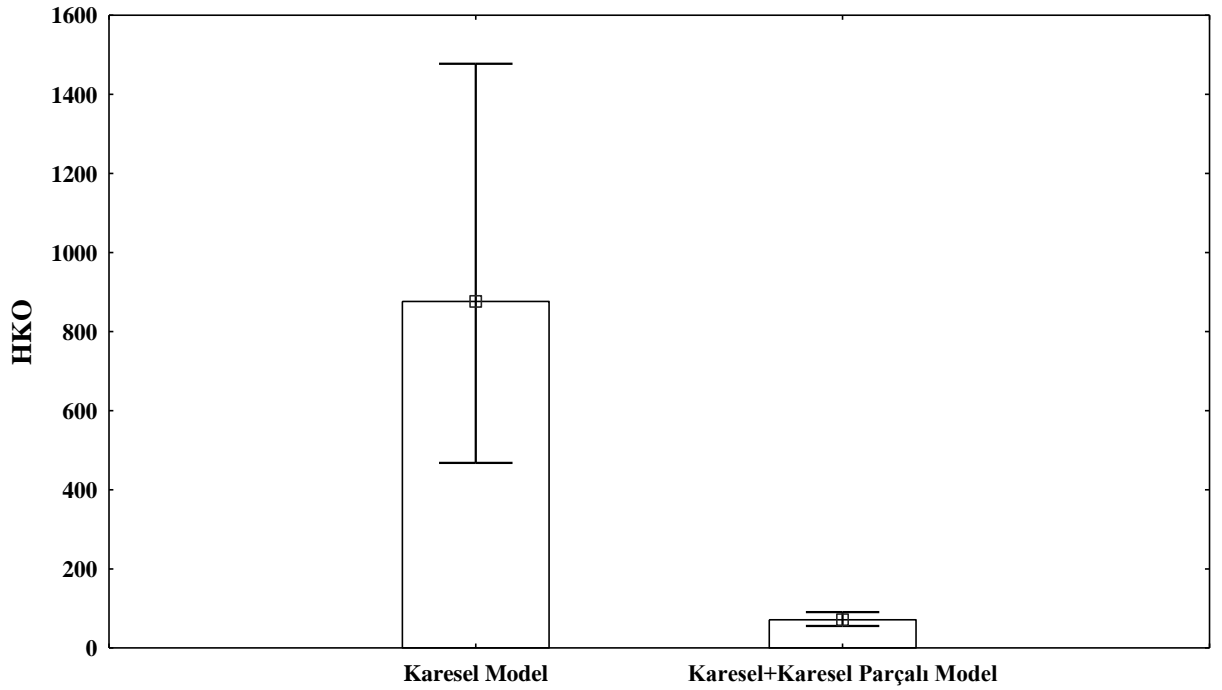
	MODEL		
	Karesel	Karesel+Karesel Parçalı	p
<b><math>R^2</math></b>	0,82 (0,65 - 0,93)	0,98 (0,97 - 0,99)	<b>&lt;0,001</b>
<b>HKO</b>	876,32 (760,00 - 1050,96)	71,30 (67,02 - 76,45)	<b>&lt;0,001</b>
<b>ABK</b>	969,36 (955,12 - 987,54)	722,48 (716,30 - 729,45)	<b>&lt;0,001</b>
<b>BBK</b>	979,78 (965,54 - 997,96)	738,11 (731,93 - 745,08)	<b>&lt;0,001</b>

Oluşturulan modeller  $R^2$ , HKO, ABK ve BBK değerleri bakımından istatistiksel olarak birbirinden farklı bulundu ( $p < 0.001$ ). Simülasyon sonucu elde edilen modellere ait değerler incelendiğinde karesel+karesel yapıdaki parçalı regresyon modelinin daha yüksek  $R^2$  değerine ve daha düşük HKO, ABK ve BBK değerlerine sahip olduğu görüldü (Tablo 4.1).

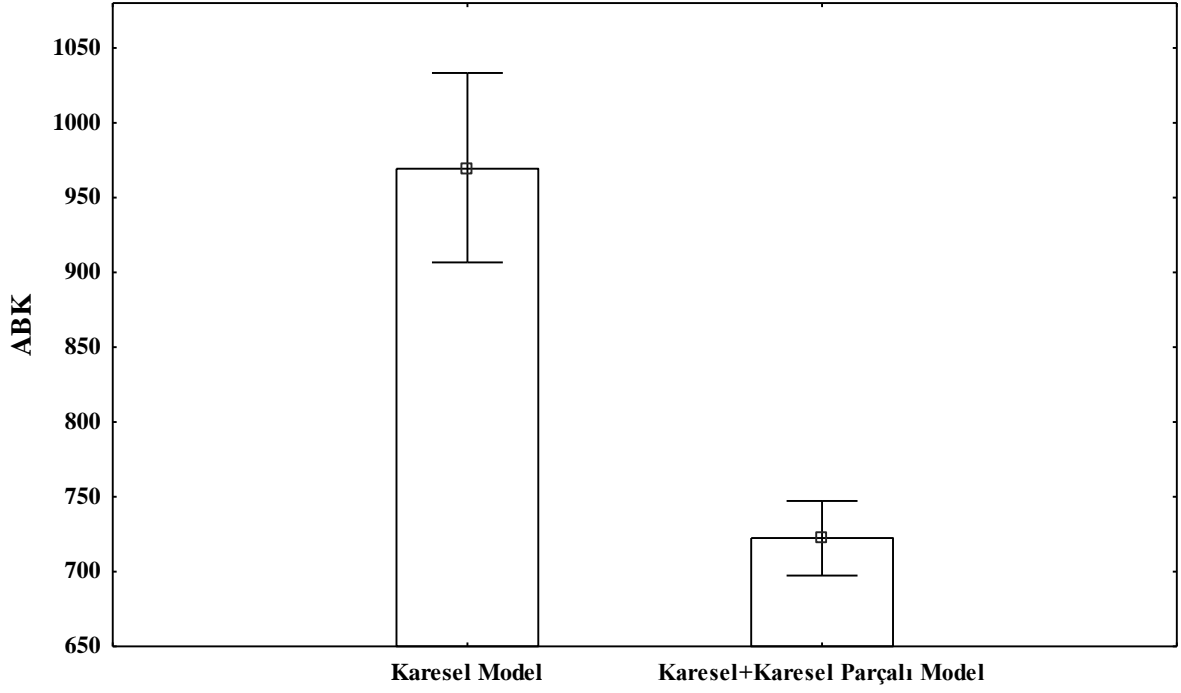
Polinom ve parçalı regresyon modeli için kutu grafikleri incelendiğinde, karesel regresyon modeline ait kutu grafiklerinin medyandan daha geniş bir yayılım gösterdiği görülürken, karesel+karesel parçalı regresyon modeline ait kutu grafiklerinin medyana daha yakın bir yayılım gösterdiği görülmüştür (Şekil 4.1 – Şekil 4.4).



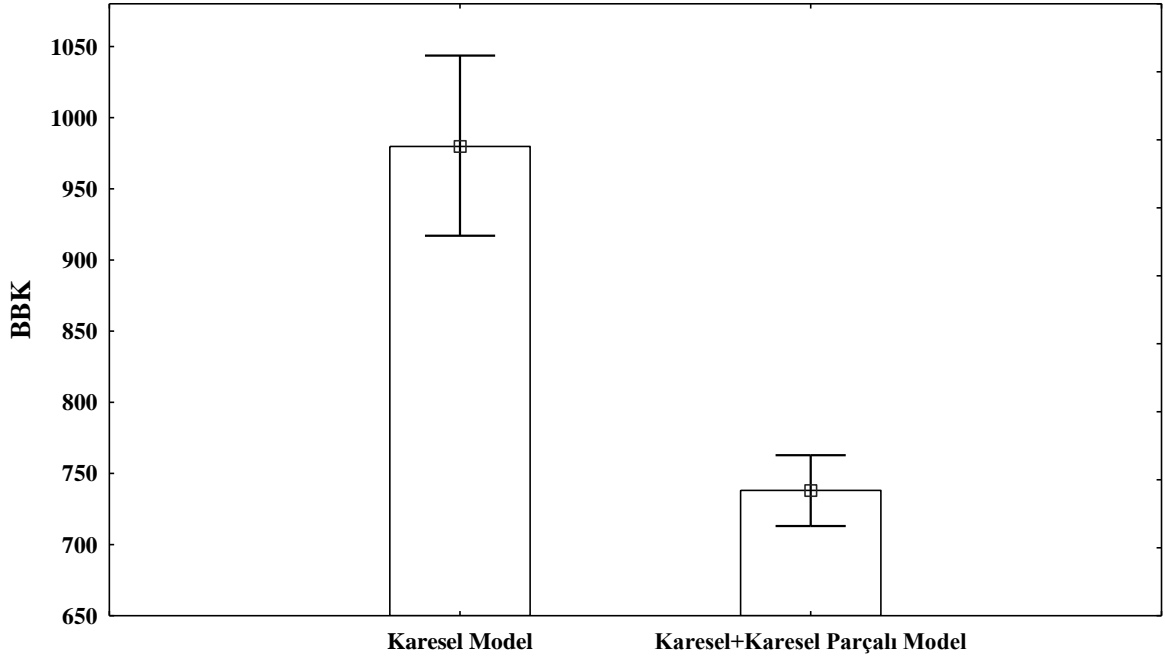
Şekil 4.1 Karesel regresyon ve karesel+karesel yapıdaki parçalı regresyon için  $R^2$  değerlerinin kutu grafiği



Şekil 4.2 Karesel regresyon ve karesel+karesel yapıdaki parçalı regresyon için HKO değerlerinin kutu grafiği



Şekil 4.3 Karesel regresyon ve karesel+karesel yapıdaki parçalı regresyon için ABK değerlerinin kutu grafiği



Şekil 4.4 Karesel regresyon ve karesel+karesel yapıdaki parçalı regresyon için BBK değerlerinin kutu grafiği

#### 4.1.2 Üçüncü Dereceden Türetilen Veriler ile Yapılan Simülasyon Sonuçları

Üçüncü dereceden türetilen veriler ile yapılan çalışmada kübik regresyon modeline göre  $R^2$  ve HKO değerlerinin normal dağılmadığı, ABK ve BBK değerlerinin normal dağıldığı; karesel+kübik parçalı regresyon modeline göre  $R^2$ , HKO, ABK ve BBK değerlerinin hiçbirinin normal dağılmadığı; kübik+kübik parçalı regresyon modeline göre yalnızca HKO değerinin normal dağılmadığı,  $R^2$ , ABK ve BBK değerlerinin normal dağıldığı tespit edildi. Bu nedenle modeller arasında bu değerler bakımından fark olup olmadığının belirlenmesi amacıyla Kruskal Wallis analizi uygulandı. Tanımlayıcı istatistikler medyan (25. - 75. persantil) olarak belirtildi (Tablo 4.2).

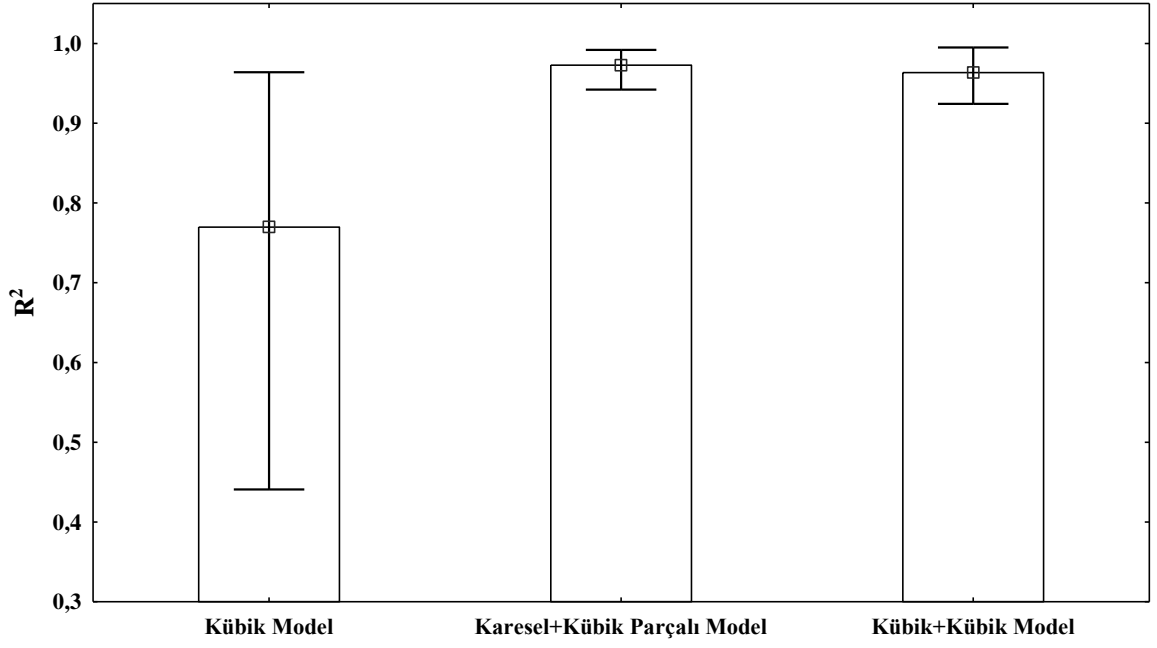
Tablo 4.2 Kübik regresyon ve parçalı regresyon modellerinin  $R^2$ , HKO, ABK ve BBK değerlerinin tanımlayıcı istatistikleri ve karşılaştırma sonuçları

	MODEL			
	Kübik	Karesel+Kübik Parçalı	Kübik+Kübik Parçalı	p
<b><math>R^2</math></b>	0,77 (0,69 - 0,85)	0,97 (0,96 - 0,98)	0,96 (0,95 - 0,97)	<0,001
<b>HKO</b>	60,14 (52,24 - 70,59)	7,68 (6,88 - 8,72)	9,74 (8,53 - 11,4)	<0,001
<b>ABK</b>	703,46 (689,37 - 719,47)	501,62 (490,59 - 514,39)	527,44 (514,18 - 543,15)	<0,001
<b>BBK</b>	716,47 (702,40 - 732,50)	519,85 (508,83 - 532,63)	548,28 (535,02 - 564,00)	<0,001

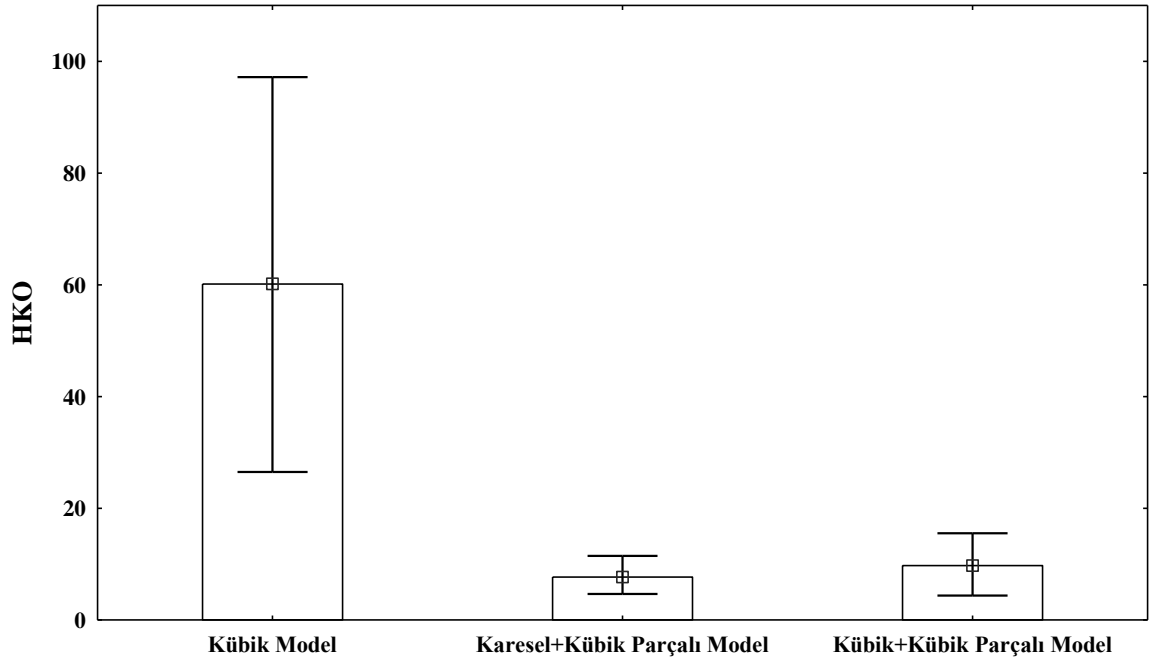
Oluşturulan modeller  $R^2$ , HKO, ABK ve BBK değerleri bakımından istatistiksel olarak birbirinden farklı bulundu ( $p < 0.001$ ). Ayrıca yapılan çoklu karşılaştırmalar sonucunda  $R^2$ , HKO, ABK ve BBK değerleri için tüm modellerin birbirinden istatistiksel olarak farklı olduğu sonucuna varıldı (tüm regresyon modellerinin birbiriyle karşılaştırılmaları için  $p < 0,001$ ). Parçalı regresyon modellerinin daha yüksek  $R^2$  değerine ve daha düşük HKO, ABK ve BBK değerlerine sahip olduğu görüldü. En yüksek  $R^2$  değeri karesel+kübik parçalı regresyon modelinde görülürken, aynı şekilde en düşük HKO, ABK ve BBK değerleri de karesel+kübik parçalı regresyon modelinde görüldü (Tablo 4.2).

Polinom ve parçalı regresyon modelleri için kutu grafikleri incelendiğinde, kübik regresyon modeline ait kutu grafiklerinin medyandan daha geniş bir yayılım gösterdiği görülürken, parçalı regresyon modellerine ait kutu grafiklerinin medyana daha yakın bir yayılım gösterdiği görülmüştür (Şekil 4.5 – Şekil 4.8).

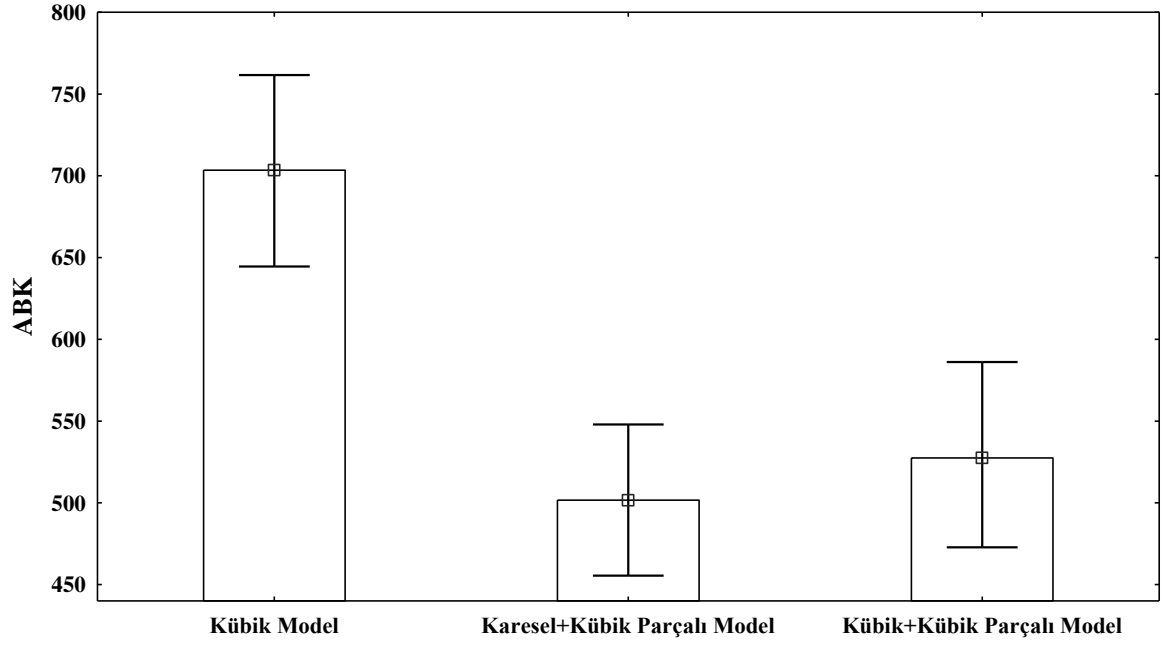




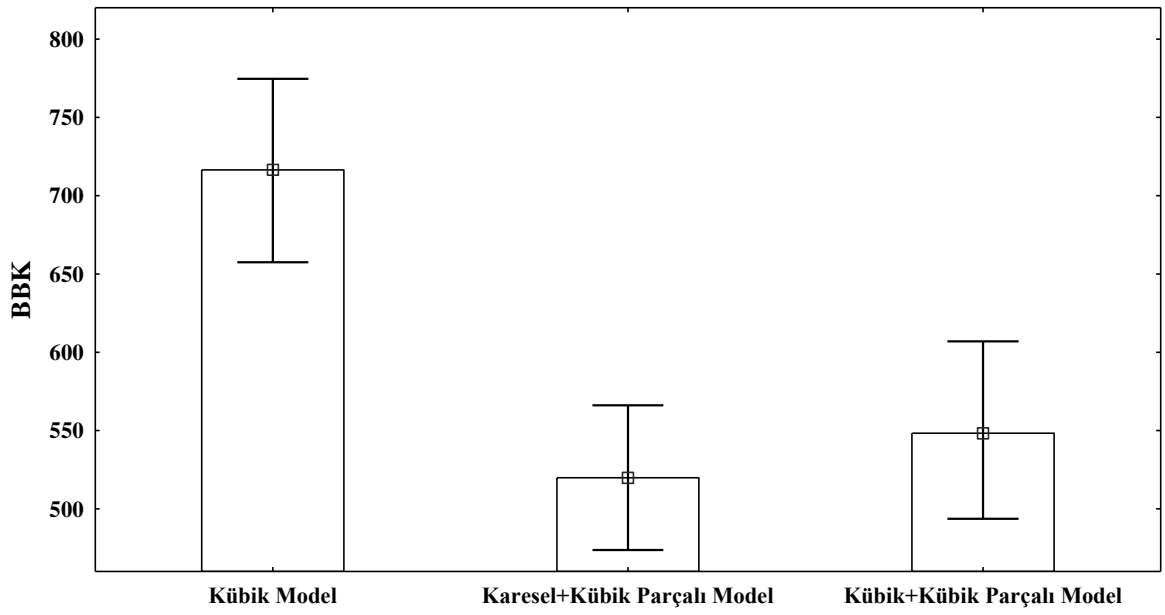
Şekil 4.5 Kübik regresyon ve parçalı regresyon modelleri için  $R^2$  değerlerinin kutu grafiği



Şekil 4.6 Kübik regresyon ve parçalı regresyon modelleri için HKO değerlerinin kutu grafiği



Şekil 4.7 Kübik regresyon ve parçalı regresyon modelleri için ABK değerlerinin kutu grafiği



Şekil 4.8 Kübik regresyon ve parçalı regresyon modelleri için BBK değerlerinin kutu grafiği

## 4.2 Gerçek Uygulama Sonuçları

### 4.2.1 Tüberküloz Veri Seti

Şekil 3.1’de görüldüğü gibi tüberküloz veri setindeki yaşa göre değişen vaka sayılarının dağılımı kübik yapıya daha uygun olduğu için polinom ve parçalı regresyonda karesel ve kübik modeller incelendi.

#### 4.2.1.1 Karesel Regresyonla Model Tahmini

Yaşa göre değişen tüberküloz vaka sayılarını tahmin etmek için aşağıdaki ikinci dereceden regresyon denklemi oluşturuldu:

$$y_i = b_0 + b_1x_i + b_2x_i^2 + e_i \quad i = 1,2,\dots,96$$

Oluşturulan modelde  $b_0$  regresyon modelinin sabiti,  $b_1$  ve  $b_2$  regresyon modelinin eğim katsayıları veya bağımsız değişkenin katsayı değerleri,  $e$  ise modelin hata terimidir.

Regresyon denklemi EKK yöntemi ile tahmin edildi. Elde edilen sonuçlara göre karesel regresyon modeli istatistiksel olarak anlamlı bulundu ( $F=101$ ;  $sd_1=2$ ,  $sd_2=93$ ;  $p<0,001$ ). Ayrıca modelin tüm katsayıları istatistiksel olarak anlamlı bulundu (Tablo 4.3).

Tablo 4.3 Karesel regresyon parametre tahminleri

	<b><i>b</i></b>	<b><i>s<sub>b</sub></i></b>	<b><i>t</i></b>	<b><i>p</i></b>
<b>Sabit</b>	60,42	20,60	2,93	0,004
<b><i>x</i></b>	10,53	1,00	10,51	<0,001
<b><i>x</i><sup>2</sup></b>	-0,13	0,01	-12,58	<0,001

Yapılan analiz sonucu elde edilen model aşağıdaki gibidir.

$$\hat{y}_i = 60,42 + 10,53x_i - 0,13x_i^2$$

#### 4.2.1.2 Kübik Regresyonla Model Tahmini

Yaşa göre değişen tüberküloz vaka sayılarını tahmin etmek için aşağıdaki üçüncü dereceden regresyon denklemi oluşturuldu:

$$y_i = b_0 + b_1x_i + b_2x_i^2 + b_3x_i^3 + e_i \quad i = 1,2,\dots,96$$

Oluşturulan modelde  $b_0$  regresyon modelinin sabiti,  $b_1$ ,  $b_2$  ve  $b_3$  regresyon modelinin eğim katsayıları veya bağımsız değişkenin katsayı değerleri,  $e$  ise modelin hata terimidir.

Regresyon denklemi EKK yöntemi ile tahmin edildi. Elde edilen sonuçlara göre kübik regresyon modeli istatistiksel olarak anlamlı bulundu ( $F=128$ ;  $sd_1=3$ ,  $sd_2=92$ ;  $p<0,001$ ). Ayrıca modelin tüm katsayıları istatistiksel olarak anlamlı bulundu (Tablo 4.4).

Tablo 4.4 Kübik regresyon parametre tahminleri

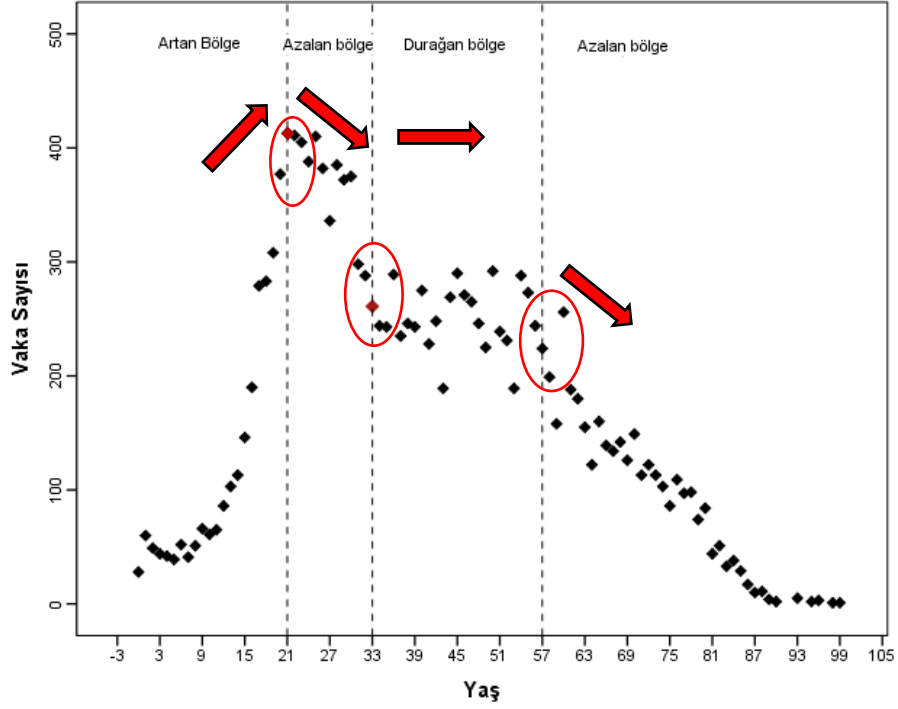
	<b>b</b>	<b>s<sub>b</sub></b>	<b>t</b>	<b>p</b>
<b>Sabit</b>	-44,39	21,20	-2,09	0,04
<b>x</b>	24,13	1,94	12,42	<0,001
<b>x<sup>2</sup></b>	-0,49	0,05	-10,24	<0,001
<b>x<sup>3</sup></b>	0,003	0,0003	7,66	<0,001

Yapılan analiz sonucu elde edilen model aşağıdaki gibidir.

$$\hat{y}_i = -44,39 + 24,13x_i - 0,49x_i^2 + 0,003x_i^3$$

#### 4.2.1.3 Parçalı Regresyonla Model Tahmini

Yaşa göre değişkenlik gösteren tüberküloz vaka sayısının dağılım grafiği Şekil 4.9'da verilmiştir. Şekil incelendiğinde, tüberküloz vaka sayısının 6-21 yaşları arasında artış gösterdiği, 21 yaşından sonra vaka sayısının azalmaya başladığı, 33-57 yaşları arasında durağan bir seyir izlediği ve 57 yaşından sonra vaka sayısında hızlı bir azalma olduğu görülmektedir. Sıçrama ve ani değişim bölgelerindeki her bir nokta birer aday dönüm noktasıdır. Bu nedenle iki dönüm noktası ve üç kısmi parçalı fonksiyon ile modellemeye karar verildi. Tüberküloz vaka sayısının sıçrama yaptığı ya da dağılım yönünde sapmaların olduğu bölgeler için yapılan denemeler sonucu 21 ile 33 yaşları parçalı regresyon uygulamak için en iyi dönüm noktaları olarak belirlendi.



Şekil 4.9 Tüberküloz vaka sayısının yaşa göre dağılımı ve dönüm noktası aday bölgeleri

Belirlenen dönüm noktalarına göre aşağıdaki parçalı fonksiyon yapısı oluşturulmuştur.

$$S(x) = \begin{cases} b_{00} + b_{01}x + b_{02}x^2 & ,0 \leq x \leq 21 \\ b_{00} + b_{01}x + b_{02}x^2 + b_{11}(x - 21) + b_{12}(x - 21)^2 & ,21 < x \leq 33 \\ b_{00} + b_{01}x + b_{02}x^2 + b_{11}(x - 21) + b_{12}(x - 21)^2 + b_{21}(x - 33) + b_{22}(x - 33)^2 & ,33 < x \leq 99 \end{cases}$$

Fonksiyona göre 0-21 yaşları arası ilk parçayı, 21-33 yaşları arası ikinci parçayı ve 33-99 yaşları arası üçüncü parçayı oluşturmaktadır. Belirlenen dönüm noktasına göre oluşturulan “+” fonksiyonları ile aşağıdaki gibi sabit katsayıları kısıtlanmış bir parçalı regresyon modeli kuruldu:

$$y_i = b_{00} + b_{01}x_i + b_{02}x_i^2 + b_{11}(x_i - 21)_+ + b_{12}(x_i - 21)_+^2 + b_{21}(x_i - 33)_+ + b_{22}(x_i - 33)_+^2 + e_i \quad i = 1, 2, \dots, 96$$

Oluşturulan modelde  $b_{00}$  regresyon modelinin sabiti,  $b_{01}$ ,  $b_{02}$ ,  $b_{11}$ ,  $b_{12}$ ,  $b_{21}$  ve  $b_{22}$  regresyon modelinin eğim katsayıları ya da bağımsız değişkenlerin ve “+” fonksiyonlarının katsayı değerleri,  $e$  ise modelin hata terimidir. Modeli kısıtlayan “+” fonksiyonları:

$$(x_i - 21)_+ = \begin{cases} (x - 21), & x_i > 21 \\ 0, & x_i \leq 21 \end{cases}$$

$$(x_i - 21)_+^2 = \begin{cases} (x - 21)^2, & x_i > 21 \\ 0, & x_i \leq 21 \end{cases}$$

$$(x_i - 33)_+ = \begin{cases} (x - 33), & x_i > 33 \\ 0, & x_i \leq 33 \end{cases}$$

$$(x_i - 33)_+^2 = \begin{cases} (x - 33)^2, & x_i > 33 \\ 0, & x_i \leq 33 \end{cases}$$

değerlerini alır. Böylece bu fonksiyonlar modelin ilgili dönüm noktasından önceki kısmına etki etmez.

Regresyon denklemi EKK yöntemi ile tahmin edildi. Elde edilen sonuçlara göre parçalı regresyon modeli istatistiksel olarak anlamlı bulundu ( $F=340$ ;  $sd_1=6$ ,  $sd_2=89$ ;  $p<0,001$ ). Ayrıca modelin tüm katsayıları istatistiksel olarak anlamlı bulundu (Tablo 4.5).

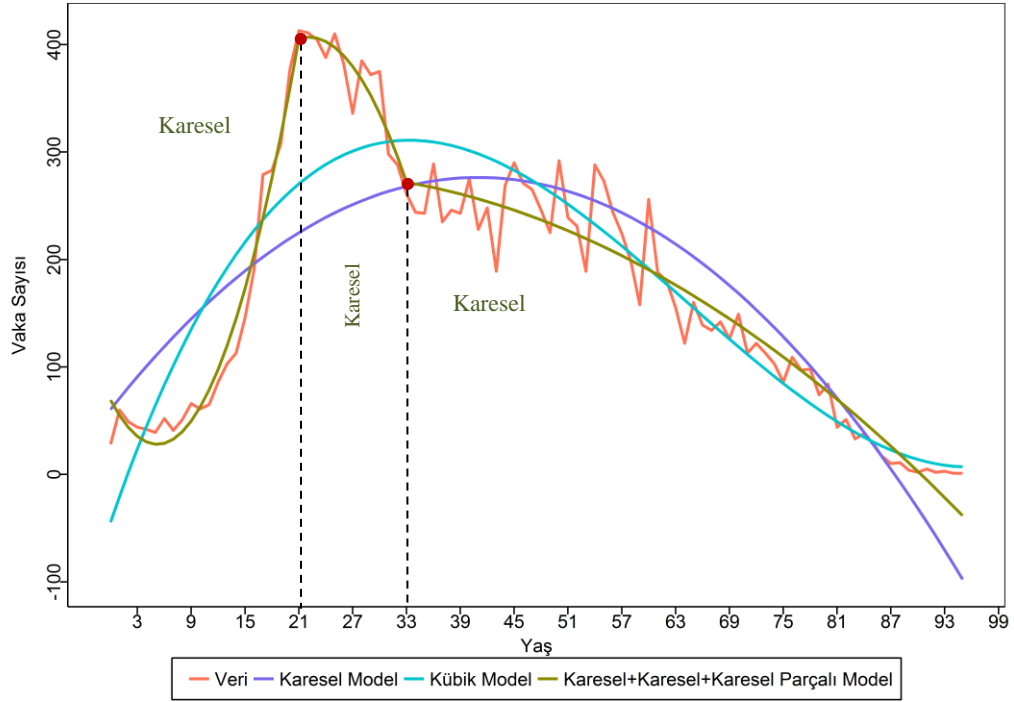
Tablo 4.5 Karesel+karesel+karesel yapıdaki parçalı regresyon parametre tahminleri

	<b>b</b>	<b>s<sub>b</sub></b>	<b>t</b>	<b>p</b>
<b>Sabit</b>	69,19	14,81	4,67	<0,001
<b>x</b>	-15,80	3,18	-4,97	<0,001
<b>x<sup>2</sup></b>	1,52	0,14	10,71	<0,001
<b>(x - 21)<sub>+</sub></b>	-45,35	7,37	-6,15	<0,001
<b>(x - 21)<sub>+</sub><sup>2</sup></b>	-2,65	0,37	-7,27	<0,001
<b>(x - 33)<sub>+</sub></b>	23,29	4,95	4,71	<0,001
<b>(x - 33)<sub>+</sub><sup>2</sup></b>	1,08	0,39	2,75	0,007

Yapılan analiz sonucu elde edilen model aşağıdaki gibidir:

$$\hat{y}_i = 69,19 - 15,80x_i + 1,52x_i^2 - 45,35(x_i - 21)_+ - 2,65(x_i - 21)_+^2 + 23,29(x_i - 33)_+ + 1,08(x_i - 33)_+^2$$

Oluşturulan modellerin grafiksel gösterimi Şekil 4.10'da verilmiştir. İşaretlenen noktalar parçalı regresyonun dönüm noktalarıdır.



Şekil 4.10 Yaşa göre değişen tüberküloz vaka sayılarının regresyon modelleri ile tahmin edilen ve gözlenen değerlerinin gösterimi

#### 4.2.1.4 Regresyon Modellerinin Karşılaştırılması

Oluşturulan regresyon modellerinin hesaplanan model belirleme kriterlerine göre karşılaştırma sonuçları Tablo 4.6’da verildi. Sonuçlar incelendiğinde parçalı regresyon modelinin diğer modellere göre daha yüksek  $R^2$  değerine ve daha düşük HKO, ABK ve BBK değerlerine sahip olduğu görüldü.

Tablo 4.6 Oluşturulan regresyon modellerinin  $R^2$ , HKO, ABK ve BBK değerleri

Regresyon Modeli	$R^2$	HKO	ABK	BBK
<b>Karesel</b>	0,68	4569,95	1089,45	1099,71
<b>Kübik</b>	0,81	2791,5	1044,13	1056,95
<b>Karesel+Karesel+Karesel Parçalı</b>	0,96	604,59	903,37	923,79

Elde edilen sonuçlara göre karesel+karesel+karesel yapıda üç kısmi parça ile oluşturulan parçalı regresyon modelinin yaşa göre değişen tüberküloz vaka sayısını tahmin etmede karesel ve kübik regresyon modellerine göre daha başarılı bir model olduğu sonucuna varıldı.

## 4.2.2 Kızamık Veri Seti

Şekil 3.2’de görüldüğü gibi kızamık veri setindeki zamana göre değişen vaka sayılarının gösterdiği dağılım kübik yapıya daha uygun olduğu için polinom regresyonda kübik, parçalı regresyonda karesel modeller incelendi.

### 4.2.2.1 Kübik Regresyonla Model Tahmini

Zamana göre değişen kızamık vaka sayısının dağılımını tahmin etmek için aşağıdaki kübik regresyon denklemi oluşturuldu:

$$y_i = b_0 + b_1x_i + b_2x_i^2 + b_3x_i^3 + e_i \quad i = 1,2, \dots, 46$$

Oluşturulan modelde  $b_0$  regresyon modelinin sabiti,  $b_1$ ,  $b_2$  ve  $b_3$  regresyon modelinin eğim katsayıları ya da bağımsız değişkenlerin katsayı değerleri,  $e$  ise modelin hata terimidir.

Regresyon denklemi EKK yöntemi ile tahmin edildi. Elde edilen sonuçlara göre karesel regresyon modeli istatistiksel olarak anlamlı bulundu ( $F=13,81$ ;  $sd_1=3$ ,  $sd_2=42$ ;  $p=0,002$ ). Ayrıca modelin tüm katsayıları istatistiksel olarak anlamlı bulundu (Tablo 4.7).

Tablo 4.7 Kübik regresyon parametre tahminleri

	<b>b</b>	<b>s<sub>b</sub></b>	<b>t</b>	<b>p</b>
<b>Sabit</b>	16 880 000 000	6 048 000 000	2,79	0,008
<b>x</b>	-25 420 000	9 107 000	-2,79	0,008
<b>x<sup>2</sup></b>	12 760	4571	2,79	0,008
<b>x<sup>3</sup></b>	-2,14	0,77	-2,79	0,008

Yapılan analiz sonucu elde edilen model aşağıdaki gibidir:

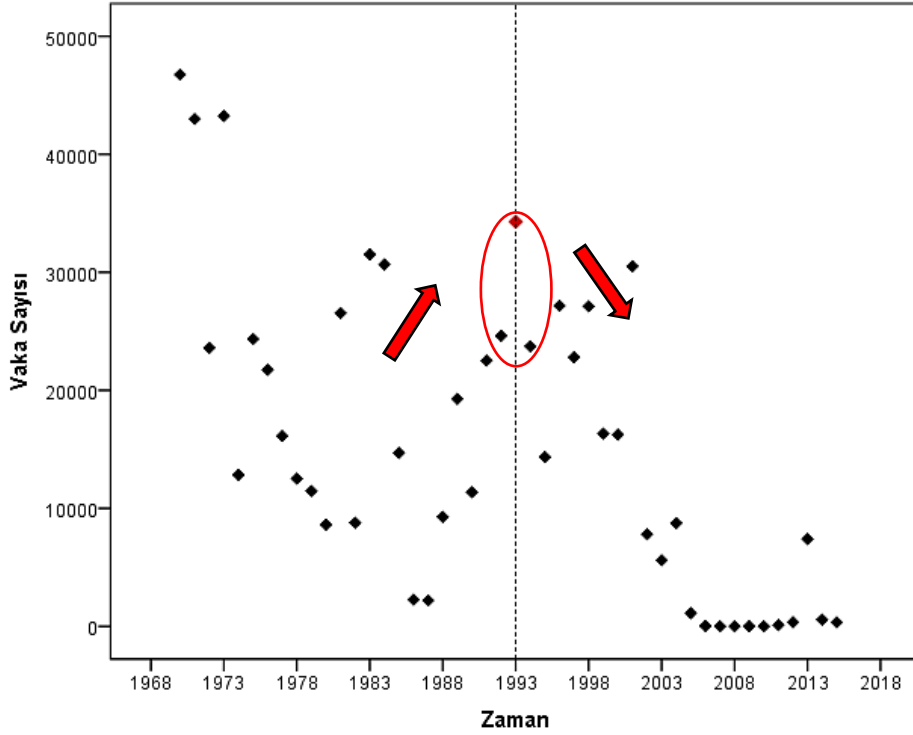
$$\hat{y}_i = 16\,880\,000\,000 - 25\,420\,000 x_i + 12\,760x_i^2 - 2,14x_i^3$$

### 4.2.2.2 Parçalı Regresyonla Model Tahmini

1970-2015 yılları arası kızamık vaka sayısının gösterdiği dağılımın grafiği Şekil 4.11’de verilmiştir. Dönüm noktası sıçrama ya da ani değişim noktalarına göre belirlendi. Şekil 4.11 incelendiğinde kızamık vaka sayısının belirli bölgelerde inişli çıkışlı bir seyir izlediği görülmektedir. Dolayısıyla bu bölgelerdeki noktaların her biri birer aday dönüm noktasıdır. Bu nedenle tek bir dönüm noktası ve iki kısmı parçalı fonksiyon ile modellemeye karar verildi ve



kızamık vaka sayısının sıçrama yaptığı noktalardan biri olan 1993 yılı yapılan denemeler sonucu parçalı regresyon uygulamak için en iyi dönüm noktası olarak belirlendi.



Şekil 4.11 1970-2015 yılları arası kızamık vaka sayılarının noktasal dağılımı

Belirlenen dönüm noktasına göre aşağıdaki parçalı fonksiyon yapısı oluşturuldu.

$$S(x) = \begin{cases} b_{00} + b_{01}x + b_{02}x^2 & ,1970 \leq x \leq 1993 \\ b_{00} + b_{01}x + b_{02}x^2 + b_{11}(x - 1993) + b_{12}(x - 1993)^2 & ,1993 < x \leq 2015 \end{cases}$$

Fonksiyona göre 1970-1993 yılları arası ilk parçayı, 1993-2015 yılları arası ikinci parçayı oluşturmaktadır. Belirlenen dönüm noktasına göre oluşturulan “+” fonksiyonları ile aşağıdaki gibi sabit katsayıları kısıtlanmış bir parçalı regresyon modeli kuruldu:

$$y_i = b_{00} + b_{01}x_i + b_{02}x_i^2 + b_{11}(x_i - 1993)_+ + b_{12}(x_i - 1993)_+^2 + e_i \quad i = 1, 2, \dots, 46$$

Oluşturulan modelde  $b_{00}$  regresyon modelinin sabiti,  $b_{01}$ ,  $b_{02}$ ,  $b_{11}$  ve  $b_{12}$  regresyon modelinin eğim katsayıları ya da bağımsız değişkenlerin ve “+” fonksiyonlarının katsayı değerleri,  $e$  ise modelin hata terimidir. Modeli kısıtlayan “+” fonksiyonları:

$$(x_i - 1993)_+ = \begin{cases} (x - 1993), & x_i > 1993 \\ 0, & x_i \leq 1993 \end{cases}$$

$$(x_i - 1993)_+^2 = \begin{cases} (x - 1993)^2, & x_i > 1993 \\ 0, & x_i \leq 1993 \end{cases}$$

değerlerini alır. Böylece bu fonksiyonlar modelin ilgili dönüm noktasından önceki kısmına etki etmez.

Regresyon denklemi EKK yöntemi ile tahmin edildi. Elde edilen sonuçlara göre parçalı regresyon modeli istatistiksel olarak anlamlı bulundu (F=18,63; sd<sub>1</sub>=4, sd<sub>2</sub>=41; p<0,001). Ayrıca modelin tüm katsayıları istatistiksel olarak anlamlı bulundu (Tablo 4.8).

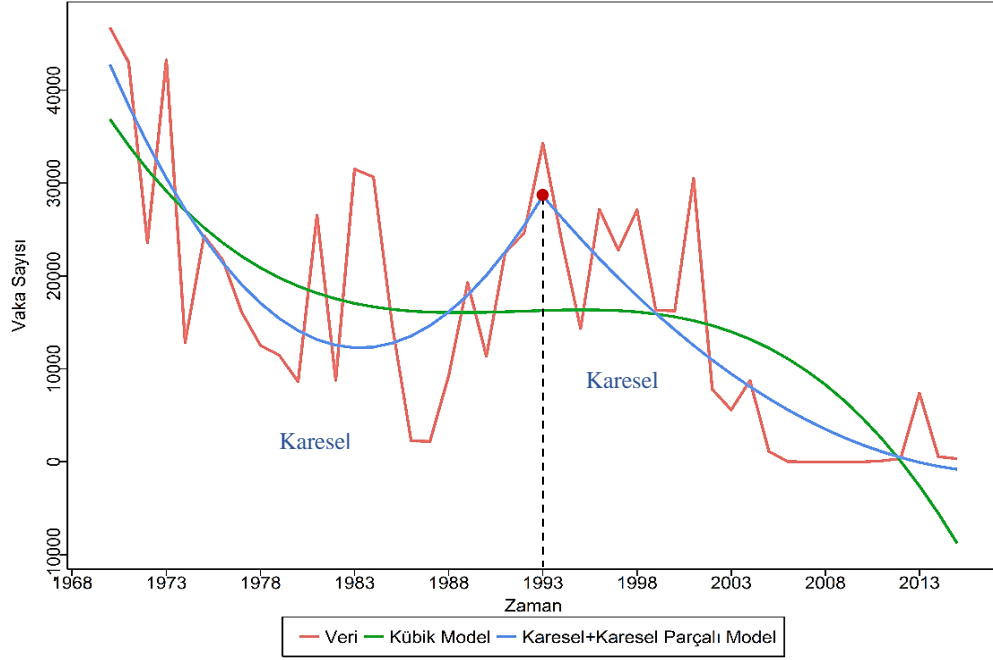
Tablo 4.8 Karesel+karesel yapıdaki parçalı regresyon parametre tahminleri

	<b>b</b>	<b>s<sub>b</sub></b>	<b>t</b>	<b>p</b>
<b>Sabit</b>	680 600 000	135 600 000	5,02	<0,001
<b>x</b>	-686 300	136 800	-5,02	<0,001
<b>x<sup>2</sup></b>	173	34,52	5,01	<0,001
<b>(x - 1993)<sub>+</sub></b>	-5757	1431	-4,02	<0,001
<b>(x - 1993)<sub>+</sub><sup>2</sup></b>	-125	40,56	-3,08	0,004

Yapılan analiz sonucu elde edilen model aşağıdaki gibidir.

$$\hat{y}_i = 680\,600\,000 - 686\,300x_i + 173x_i^2 - 5757(x - 1993)_+ - 125(x - 1993)_+^2$$

Oluşturulan modellerin grafiksel gösterimi Şekil 4.12'de verilmiştir. İşaretlenen nokta parçalı regresyonun dönüm noktasıdır.



Şekil 4.12 Kızamık veri setindeki 1970-2015 yılları arasındaki kızamık vaka sayılarının regresyon modelleri ile tahmin edilen ve gözlenen değerlerinin gösterimi

#### 4.2.2.3 Regresyon Modellerinin Karşılaştırılması

Oluşturulan regresyon modellerinin hesaplanan model belirleme kriterlerine göre karşılaştırma sonuçları Tablo 4.9’da verildi. Sonuçlar incelendiğinde parçalı regresyon modelinin daha yüksek  $R^2$  değerine ve daha düşük HKO, ABK ve BBK değerlerine sahip olduğu görüldü.

Tablo 4.9 Oluşturulan regresyon modellerinin  $R^2$ , HKO, ABK ve BBK değerleri

Regresyon Modeli	$R^2$	HKO	ABK	BBK
Kübik	0,50	82 550 036	979,07	988,22
Karesel+Karesel Parçalı	0,65	58 190 618	964,99	975,96

Elde edilen sonuçlara göre karesel+karesel yapıda iki kısmi parça ile oluşturulan parçalı regresyon modelinin, kızamık vaka sayısının zamana göre gösterdiği değişkenliğin dağılımını tahmin etmede kübik regresyon modeline göre daha başarılı bir model olduğu sonucuna varıldı.

## 5. TARTIŞMA

Parçalı regresyonlar birçok alanda kullanılmasına rağmen sağlık alanında kullanımı halen yaygınlaşmamıştır. Polinom regresyonların parçalı regresyonlara göre analiz kolaylığı sağlaması ve istatistiksel paket programlarda polinom regresyonlara daha çok yer verilmesi parçalı regresyon tekniğinin yaygınlaşmamasına neden olmuştur. Bu durum sağlık alanında çalışmalar yapan araştırmacıları polinom regresyon modellerini kullanmaya itmiştir. Oysaki parçalı regresyonları oluşturan polinomlar bilgisayar programlarında rahatlıkla birleştirilebilir ve araştırmacıya gerekli analiz kolaylığını sağlar. Bu çalışmada hem simülasyon verisi hem de gerçek veri setleri kullanılarak tek değişkenli polinom regresyon analizi ile karesel ve kübik parçalı regresyon analizlerinin performanslarının karşılaştırılması hedeflendi.

Literatürler incelendiğinde parçalı regresyonların kullanıldığı bazı çalışmalar dikkat çekmektedir. Garcia ve ark. (2007), Meksika’da bir firmanın 1 yıllık süt satışı miktarlarının (litre) aylık değişimini analiz etmek için EKK yöntemi ile parametrelerini tahmin ettikleri doğrusal, karesel, kübik regresyon ve kübik parçalı regresyon modellerini kullanmışlardır. Verilerin dağılımını doğrusal, karesel ve kübik regresyon modelleriyle tek bir parçada, kübik parçalı regresyon modeliyle de iki dönüm noktası belirleyerek üç parçada tahmin etmişlerdir. Firmanın daha önceki yıllarda satışlarında mevsimsel azalma ve artma yaşadığı ayları belirleyip bunlardan ikisini dönüm noktaları olarak kullanmışlardır (sabit dönüm noktaları). Modelleri  $R^2$  ve HKO değerlerine göre karşılaştırmış ve kübik parçalı regresyon modelinin satış miktarını tahmin etmede en etkin model olduğunu tespit etmişlerdir. Karadavut ve ark. (2004), bitki boyunun (cm) zamana göre değişiminin modellenmesi için 10’ar gün aralıklarla toplam 15 kez ölçtüğü 15 bitkinin boy ölçüm değerlerini kullanmışlar, EKK yöntemi ile parametrelerini tahmin ettikleri kübik regresyon ve karesel+kübik yapıdaki parçalı regresyon modellerinin performanslarını karşılaştırmışlardır. Çalışmada tek bir dönüm noktası kullanmışlar ve görsel inceleme sonucu belirledikleri 6. günü dönüm noktası olarak seçmişlerdir. Elde ettikleri sonuçlara göre parçalı regresyon modelinin daha yüksek  $R^2$  değerine ve daha düşük standart hata değerine sahip olduğunu ve büyümeyi dönemler halinde parçalı regresyonlarla incelemenin bütün olarak incelemeye göre daha uygun olduğunu bildirmişlerdir. Alkan (2004), Türkiye’de 1923-2010 yılları arası ihracatın ithalatı karşılama oranlarının yıllık değişimlerini EKK yöntemi ile parametrelerini tahmin ettiği polinom (karesel ve kübik) ve parçalı (karesel ve kübik) regresyonlarla incelemiştir. Çalışmada tüm parçalı modeller için iki dönüm noktası kullanmış ve dağılım grafiğinin görsel incelenmesi ile tespit ettiği bölgelerden yaptığı

denemeler sonucu en uygun noktaları dönüm noktası olarak belirlemiştir. Oluşturduğu modellerin performanslarını  $R^2$  ve HKO değerleri açısından karşılaştırmıştır. Elde ettiği bulgulara göre tüm parçalı regresyon modellerinin, polinom regresyon modellerine göre daha yüksek  $R^2$  ve daha düşük HKO değerlerine sahip olduğunu bildirmiştir. Hurley ve ark. (2004), parçalı regresyonlar ile basit regresyonların performanslarını karşılaştırmak için bir bağımlı ve 201 birimden oluşan,  $x \sim U(1,100)$  dağılımından türetilen bir bağımsız değişken kullanarak üçü karesel yapıda ikisi kübik yapıda olmak üzere oluşturdukları beş farklı yapıdaki veri seti ile 2000 tekrarlı simülasyon çalışması yapmışlardır. Her veri yapısı için doğrusal regresyon, polinom regresyon (kübik ve karesel) ve parçalı regresyon (doğrusal, karesel ve kübik) olmak üzere altı regresyon modeli kurmuşlardır. Her bir parçalı regresyon modeli için  $x=32$  ve  $x=68$  noktalarını sabit dönüm noktaları olarak belirlemişlerdir. Türettikleri veri setlerini tahmin etmek için, EKK yöntemi ile veri yapısına uygun doğrusal, polinom ve parçalı regresyon modelleri oluşturmuş; parçalı regresyon modellerinin daha yüksek  $R^2$  değerine ve daha düşük HKO değerine sahip olduklarını bildirmişlerdir. Mulla (2007), klinik araştırmalarda özellikle hastaya verilen ilaç dozu ve alınan yanıt ile ilgili çalışmalarda parçalı regresyon modelleme tekniğinin kullanılmasının araştırmacıya büyük faydalar sağlayacağını bildirmiş ve karesel parçalı regresyon modelini kullanarak serum albümin miktarı (g/100 ml) ile hasta ölüm oranı riski arasındaki ilişkiyi araştırmıştır. Bu çalışma için 1,1 ile 5,1 g/100 ml arasında değişen miktarlarda serum albümin verilen 117 hastaya ait kayıtları kullanmıştır. EKK yöntemi ile oluşturduğu parçalı regresyon modeli için görsel inceleme sonucu belirlediği iki noktayı dönüm noktaları olarak tespit etmiştir. Elde ettiği sonuçlara göre karesel parçalı regresyonun yüksek  $R^2$  değerine sahip olduğunu bildirmiştir. Wold (1989), 60 hasta ile yaptığı çalışmada oral yoldan alınan bir ilacın kandaki konsantrasyonunun zamana (saat) göre değişimini incelemiştir. Bu çalışmada her ikisini de EKK yöntemi ile oluşturduğu kübik parçalı regresyon modeli ve klasik bir model kullanmış, kübik parçalı regresyon modeli için görsel inceleme ile belirlediği bir dönüm noktası oluşturmuştur. Çalışmada elde ettiği bulgulara göre kübik parçalı regresyon modelinin 60 hastaya ait kandaki konsantrasyon değerlerinin tamamını gerçeğe yakın doğrulukta tahmin ettiğini; buna karşın klasik modelin ancak 25 hastaya ait değerlerde gerçeğe yakın tahmin sonuçları verdiğini, kalan 35 hastaya ait sonuçların gerçek konsantrasyon değerlerinden oldukça büyük farklılıklar içerdiğini ve bu farklılığın parçalı regresyon için belirlenen dönüm noktası etrafında toplandığını bildirmiştir. Farrel ve Walker (2017), İskoçya'da 1982-2014 yılları arasında kronik karaciğer hastalığından ölen kadın ve erkeklerin sayısını gösteren bir veri setine ait noktasal dağılımı tahmin etmek amacıyla EKK yöntemi ile parametrelerini tahmin ettikleri karesel, kübik ve kübik parçalı regresyon modellerini

kullanmışlardır. Kübik parçalı regresyon modeli için sabit dönüm noktaları kullanmış ve literatürdeki bilgilere dayanarak üç dönüm noktası belirlemişlerdir. Elde ettikleri bulgulara göre en yüksek  $R^2$  değerini kübik parçalı regresyon modelinden elde etmişlerdir. Çalışmamızda simülasyon verisi ve gerçek veri setleri kullanılarak EKK yöntemi ile polinom ve parçalı regresyon modelleri oluşturulmuş ve bu modellerin performansları karşılaştırılmıştır. Simülasyon uygulamasında veri üretim aşamasında oluşturulan kırılım noktalarına göre belirlenen sabit dönüm noktaları kullanılmış, oluşturulan tüm parçalı regresyon modelleri için dönüm noktası sayısı bir olarak belirlenmiştir. Gerçek veri setleri kullanılarak yapılan uygulamalarda ise değişken dönüm noktaları kullanılmış, dönüm noktası sayıları tüberküloz veri seti ile yapılan çalışmada iki, kızamık veri seti ile yapılan çalışmada bir olarak belirlenmiştir. Değişken dönüm noktalarının belirlenmesinde öncelikle görsel inceleme sonucu kaç dönüm noktası kullanılacağına karar verilmiş ve aday dönüm noktası bölgeleri tespit edilmiş; daha sonra yapılan denemeler sonucu en yüksek performansı veren noktalar dönüm noktaları olarak seçilmiştir. Aday bölge olarak belirlenen bir bölgeden dönüm noktası olarak denenen noktalardan hiçbirisi modelin tahmin gücüne anlamlı bir katkıda bulunmuyorsa o bölgeden dönüm noktası olarak herhangi bir seçim yapılmamıştır. Hem simülasyon verisi hem de gerçek veri setleri kullanılarak yapılan uygulama sonuçlarına göre parçalı regresyonların performansının, daha yüksek  $R^2$  ve daha düşük HKO, ABK, BBK değerleri ile polinom regresyonlara göre daha iyi olduğu görülmüştür.

Uygun dönüm noktalarının belirlenmesi yüksek performanslı tahminleme için oldukça önemlidir. Parkhurst ve ark (2002) tarafından yapılan çalışmada hem sabit hem de değişken dönüm noktalı parçalı regresyon modelleri kullanılmış ve oluşturulan tüm değişken dönüm noktalı modellerin, aynı yapıdaki sabit dönüm noktalı modellere göre daha yüksek tahmin gücüne sahip olduğu tespit edilmiştir. Ancak çalışmamızda simülasyon sonuçları, sabit dönüm noktaları kullanılarak yapılan analizlerden de yüksek tahmin gücüne sahip modellerin elde edilebileceğini göstermiştir. Önemli olan dönüm noktasının konumunun en uygun şekilde tespit edilmesidir. Şahin (2009), parçalı regresyonların ifade edilmesi aşamasında, dönüm noktalarının seçimi ve parçalamanın nasıl ve hangi fonksiyonlarla yapılacağı konusunda belirleyici olan üç faktörü ön plana çıkarmış ve bu faktörleri; noktasal dağılımın ifade edilmesinde kullanılacak polinomların derecesi, parçalamada kullanılacak dönüm sayısı ve dönüm noktası veya noktalarının konumu olarak sıralamıştır. Seber ve Wild (2003) ve Eubank (1988), dalgalı yapıdaki noktasal dağılımların parçalı regresyonlarla ifade edilmesinde derecesi düşük polinomlar kullanılırsa yüksek bir tahminleme sağlamak için daha fazla dönüm noktası sayısına ihtiyaç duyulacağını, dönüm noktası sayısı fazlalığının da oluşan aralıklarda ve

parametre sayısında artışa yol açacağını, bunun bir sonucu olarak da araştıracının daha karmaşık modellerle çalışmak zorunda kalacağını ve zaman kaybına uğrayacağını bildirmiştir. Ayrıca belirlenen en uygun dönüm noktalarına göre oluşturulan noktasal aralıklar için gereksiz yere yüksek dereceli polinomlar kullanılmasının da  $R^2$ 'de anlamlı bir artış sağlamadığı gibi parametre fazlalığına ve serbestlik derecesi kaybına neden olacağını tespit etmiştir. Wold (1989), dönüm noktası sayısı ve konumunun belirlenmesinde bazı noktalara dikkat çekmiş ve parçalı regresyonu oluşturan her bir aralığın en az 4-5 gözlem noktası içermesi gerektiğini, bu nedenle dönüm noktası sayısının mümkün olduğunca az seçilmesi gerektiğini bildirmiştir. Çalışmamızda kullanılan kızıyık veri setinin noktasal dağılımı kübik yapıya daha uygun olmasına rağmen, en uygun bir dönüm noktasına göre oluşturulan karesel+karesel yapıdaki parçalı regresyon modeli kullanılarak yüksek performansla tahmin edilmiştir. Dönüm noktası sayısında artış ihtiyacına neden olmayacaksa düşük dereceli polinomlarla çalışmak işlem kolaylığı sağlanması açısından istediğimiz bir durumdur. Çalışmamızda dönüm noktasının konumunun dağılımın şekli ile yakından ilgili olduğu gözlenmiş ve mümkün olduğunca az sayıda dönüm noktası ile oluşturulan alt aralıklara uygun dereceden polinomlar kullanılmıştır. Ayrıca görsel inceleme ile belirlenen aday dönüm bölgelerindeki her bir noktanın dönüm noktası olarak seçilmesi durumunda modelin performansına yapacağı katkının birbirine yakın olduğu tespit edilmiştir. Burada öncelikle dönüm noktası sayısı belirlenmeli ve dönüm noktası ya da noktalarının hangi aday bölgelerden seçileceğine karar verilmelidir. Daha sonra dönüm noktalarının konumu belirlenmeli, gerekirse aday bölgelerden denemeler yapılarak en yüksek tahmin gücünü sağlayan dönüm noktası ya da noktaları tespit edilmelidir.

## 6. SONUÇ VE ÖNERİLER

Bu çalışmada polinom ve parçalı regresyon yöntemleri; ikinci ve üçüncü dereceden türetilen veriler ile gerçekleştirilen simülasyon uygulamasında ve hem tüberküloz hem de kızamık hastalıkları ile ilgili veri setleri kullanılarak karşılaştırıldı.

İkinci dereceden türetilen veriler ile yapılan simülasyon uygulamasında karesel regresyon ve karesel+karesel parçalı regresyon modellerinin  $R^2$ , HKO, ABK ve BBK değerleri bakımından performansları istatistiksel olarak birbirinden farklı bulundu ( $p<0.001$ ). Karesel+karesel parçalı regresyon modelinin  $R^2$  değerinin karesel regresyon modeline göre daha yüksek; HKO, ABK ve BBK değerlerinin ise daha düşük olduğu gözlemlendi.

Üçüncü dereceden türetilen veriler ile yapılan simülasyon uygulamasında kübik regresyon, karesel+kübik parçalı regresyon ve kübik+kübik parçalı regresyon modellerinin  $R^2$ , HKO, ABK ve BBK değerleri bakımından performansları istatistiksel olarak birbirinden farklı bulundu ( $p<0.001$ ). Her iki parçalı regresyon modelinin de  $R^2$  değerinin kübik regresyon modeline göre daha yüksek; HKO, ABK ve BBK değerlerinin ise daha düşük olduğu gözlemlendi. Oluşturulan modeller incelendiğinde, en yüksek performanslı modelin karesel+kübik parçalı regresyon modeline ait olduğu görülürken en düşük performanslı modelin kübik regresyon modeline ait olduğu tespit edildi.

Gerçek veri setleri kullanılarak yapılan çalışmalarda da simülasyon çalışmasında olduğu gibi tüm model değerlendirme kriterlerine göre karşılaştırmalar yapıldığında parçalı regresyon modellerinin performanslarının, polinom regresyon modellerine göre daha iyi olduğu görüldü. Tüberküloz veri seti ile yapılan uygulamada karesel+karesel+karesel parçalı regresyon modelinin; kızamık veri seti ile yapılan uygulamada ise karesel+karesel parçalı regresyon modelinin en iyi performansa sahip modeller oldukları tespit edildi.

Parçalı regresyon yöntemlerinde dönüm noktalarının sayısının ve yerinin belirlenmesi, modelin tahmin gücünü etkileyen en önemli faktörlerdir. Dönüm noktası sayısının gereğinden fazla seçilmesi,  $R^2$ 'de anlamlı bir artış sağlamazken, parametre sayısında da artışa neden olacaktır. Bu durum hem araştırmacının analiz aşamasında işini zorlaştıracak hem de ihtiyaç duyulandan fazla aralık fonksiyonlarının oluşmasına neden olacaktır. Bu nedenle mümkün olduğunca az sayıda dönüm noktası kullanmaya özen gösterilmelidir. Yapılacak analizlerde değişken dönüm noktaları kullanılacaksa öncelikle noktasal dağılım ayrıntılı incelenmeli ve dönüm noktaları için aday bölgeler belirlenmelidir. Aday bölgelerin doğru tespit edilmesi durumunda bu bölgelerden dönüm noktası olarak seçilen her nokta birbirine yakın  $R^2$  değerine



sahip modellerin oluşmasını sağlayacaktır. Ancak bu bölgelerdeki tüm noktalar tek tek denenerek en uygun olanları tespit edilebilir. Böylece dönüm noktalarının yeri ve sayısı belirlenebilir. Eğer sabit dönüm noktaları kullanılacaksa da öncelikle çalışma ile ilgili literatür incelenmeli ve dönüm noktalarının tespiti buna göre yapılmalıdır. Örneğin bir hastalık ile ilgili yeni üretilen bir ilacın gün bazında iyileşme süresi üzerine etkisi araştırılıyorsa, daha önceden bu hastalık ile ilgili yapılan çalışmalarda kullanılan ilaçların hastalarda yaklaşık olarak kaçınıcı günden itibaren bir iyileşme sağladığı tespit edilmeli ve belirlenecek dönüm noktası da bu bilgi göz önünde bulundurularak belirlenmelidir. Ayrıca noktasal dağılımlarda aykırı gözlemler bulunması durumunda dönüm noktalarının bu bölgelerden seçilmemesi gereklidir. Aksi takdirde istatistiksel açıdan anlamsız modellerin oluşmasına neden olur.

Aşırı dalgalı yapıdaki noktasal dağılımlarda dönüm noktalarının konumunu belirlerken oluşacak alt aralıkların öncelikle kübik yapıya uygun polinomlarla ifade edilecek şekilde düzenlenmesi gerekir. Bunun nedeni aşırı dalgalı yapıdaki bir noktasal dağılımı kübik yapıda parçalarla ifade etmek daha az sayıda dönüm noktasının kullanılmasını gerektirir.

Sağlık alanında yapılan çalışmaların çoğunda polinom regresyon yöntemlerinin tercih edilmesine rağmen bu çalışma ile en uygun dönüm noktalı parçalı regresyonlarla veri analizinin istatistiksel açıdan üstünlük sağladığı uygulamalarla ortaya konmuştur. Bundan sonra yapılacak çalışmalarda, bağımlı ve bağımsız değişken arasındaki ilişkinin inceleneceği tüm veri setleri için parçalı regresyon yöntemi de güçlü bir alternatif olarak görülmelidir. Ayrıca sağlık alanında yapılacak araştırmalarda, daha yüksek performanslı tahminleme için parçalı regresyon kullanımı yaygınlaştırılmalıdır.

## KAYNAKLAR

- Agarwal GG, Studden WJ.** An algorithm for selection of design and knots in the response curve estimation by spline functions, Purdue University Press, Department of Statistics, Indiana, 1978, 30.
- Aggrey SE.** Comparison of three nonlinear and spline regression models for describing chicken growth curves. Poultry Science Association, 2002, 81(12), 1782-1788.
- Ahlberg JH, Nilson EN.** Convergence properties of the spline fit. *Journal of the Society for Industrial and Applied Mathematics*, 1953, 11(1), 95-104.
- Akaike H.** A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 1974, 19 (6), 716-723.
- Akkaya Ş, Pazarlıoğlu MV.** Ekonometri I, 4. Baskı, Anadolu Matbaacılık, İzmir, 2000, 579.
- Aladağ ÇH, Eğrioğlu E.** En iyi yapay sinir ağı mimarisinin belirlenmesinde uyarlamalı ağırlıklı bilgi kriterinin kullanılması. Türkiye İstatistik Kurumu 19. İstatistik Araştırma Sempozyumu Bildiriler Kitabı, 2010, 49-56.
- Allen DM.** Mean square error of prediction as a criterion for selecting variables. *Technometrics*, 1971, 469-475.
- Alkan O, Genc A, Oktay E, Celik AK.** Electricity consumption analysis using spline regression models: The case of a turkish province. *Asian Social Science*, 2013, 9(7), 231.
- Alkan Ö.** İhracatın ithalatı karşılama oranlarının parçalı regresyonlarla modellenmesi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 2012, 16(1), 497-511.
- Alpar R,** Uygulamalı Çok Değişkenli İstatistiksel Yöntemler, 5. Baskı, Detay Yayıncılık, Ankara, 2017, 840.
- Barrodale I, Young A.** A note on numerical procedures for approximation by spline functions. *Computer Journal*, 1966, 9(3), 319.
- Başar A, Oktay E.** Uygulamalı İstatistik 2 Kısa Teorik Bilgiler ve Çözülmüş Problemler, Aktif Yayınevi, Erzurum 2004, 297.
- Berberoğlu B, Berberoğlu CN.** Modeling the Structural Shifts in Real Exchange Rate with Cubic Spline Regression (CSR), Turkey 1987-2008, *International Journal of Business and Social Science*, 2011, 17(2).
- Chan SH.** Polynomial spline regression with unknown knots, The Ohio State University Department of Statistics Doctoral Thesis, USA, 1989, 185.
- Chen S, Cowan CF, Grant PM.** Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on neural networks*, 1991, 2(2), 302-309.

- DeBoor C, Rice JR.** Least Squares Cubic Spline Approximation I - Fixed Knots, Department of Computer Science Technical Reports, 1968, 30.
- DeBoor C, Rice JR.** Least Squares Cubic Spline Approximation II - Variable Knots, Department of Computer Science Technical Reports, 1968, 39.
- Draper NR, Smith H.** Applied Regression Analysis (3 rd ed), John Wiley&Sons, New York, 2014, 736.
- Eubank RL.** Nonparametric Regression and Spline Smoothing (2 nd ed), Marcel Dekker, New York, 1999, 343.
- Eubank RL.** Approximate regression models and splines. *Communications in Statistics - Theory and Methods*, 1984, 13(4), 433-484.
- Fan J, Gijbels I.** Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66, CRC Press, 1996, 336.
- Farebtother RW, Schyns M.** Visualizing statistical models and concepts, CRC Press, England, 2002, 280.
- Farrel R, Walker D.** PHI trend analysis guidance, Public Health & Intelligence, National Services Scotland, 2017, 50.
- Freedman DA.** Statistical Models: Theory and Practice, Cambridge University Press, New York, 2005, 424.
- Freeman JA, Skapura DM.** Neural networks: algorithms, applications and programming techniques, Addison-Wesley, 1991, 401.
- Freund RJ, Wilson WJ, Sa P.** Regression Analysis, Statistical Modeling of a Response Variable (2 nd ed), San Diego, Academic Press, 2006, 459.
- Gallant AR, Fuller WA.** Fitting segmented polynomial regression models whose join points have to be estimated. *Journal of the American Statistical Association*, 1973, 68, 144-147
- Garcia JL, Coronel NA, Martinez EA, Duque JA.** Using spline for demand forecasting, *Second International Congress in Engineering and Technology*, pp 37-56, October 2008, Chihuahua, Mexico: Ciudad Juarez.
- Geyer CJ.** Model selection in R. *Masters level theory notes*, 2003.
- Gujarati DN.** Use of dummy variables in testing for equality between sets of coefficients in two linear regressions. *The American Statistician*, 1970, 24(1), 50-52.
- Gujarati DN.** Basic Econometrics, 4th Edition, McGraw-Hill, New York 2003, 1002.
- Günel A.** Regresyon denkleminin başarısını ölçmede kullanılan belirleme katsayısı ve kritiği. *Doğuş Üniversitesi Dergisi*, 2003, 4(2), 133-140.

- Harrell FE.** Regression modeling strategies, Department of Biostatistics, Vanderbilt University School of Medicine Press, 2017, 570.
- Hartley HO, Brooker A.** Nonlinear least square estimation. *The Annals Of Mathematical Statistics*, 1965, 36, 638-650.
- Hawkins, Douglas M.** A Note on continuous and discontinuous segmented regressions. *Technometrics*, 1980, 22(3), 443-444.
- Hodson, Darek J.** Fitting segmented curves whose join points have to be estimated. *Journal Of The American Statistical Association*, 61(316), 1097-1125.
- Hurley D, Hussey J, McKeown R, Addy C.** An evaluation of splines in linear regression. *Innovations in Biostatistical Methods and Applications: The 132nd Annual Meeting*, pp 31, 6-10 November 2014, Washington.
- Jang JSR, Sun CT, Mizutani E.** Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence, Prentice Hall, 1997, 614.
- Jennrich, Robert I.** Asymptotic properties of non-linear least square estimators. *The Annals of Mathematical Statistics*, 1969, 40(2), 633-643.
- Karadavut U, Genç A, Tozluca A, Sinan A, Karakoca A, Aksoyak Ş, Palta Ç.** Parçalı regresyon yardımı ile bitki boyu-zaman ilişkisi parametrelerinin tahmini. *Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 2004, 8(1), 114-117.
- Kaps M and Lamberson WR.** Biostatistics for Animal Science, 2nd Edition, MPG Books Group, London 2009, 504.
- Koutsoyiannis A.** Ekonometri Kuramı, Şensel Ü, Şenesen G, Verso Yayıncılık, Ankara, 1989, 688.
- Large B.** Applied Regression Analysis, Statistics 333, Springer, 2003.
- Lee BE, Brooks DM.** Regression modeling strategies for microarchitectural performance and power prediction, Harvard Computer Science Group Technical Report, USA, 2006, 38.
- Lehmann EL, Casella, George.** Theory of Point Estimation (2nd ed.), Springer Texts in Statistics, 1998, 309-427.
- Lind DA, Marchal WG, Wathen SA.** Basic Statistics for Business and Economics (3rd ed.), New York: McGraw-Hill Higher Education, 1996, 581.
- Lindstrom, Mary J.** Penalized estimation of free-knot splines. *Journal of Computational and Graphical Statistics*, 1999, 8(2), 333-352.
- Markov D.** Information Content In Stock Market Technical Patterns: A Spline Regression Approach. University of Notre Dame Doctoral Thesis, Indiana 2002, 42.
- Marsh LC, Cormier DR.** Spline Regression Models, Sage Publications, California 2002, 69.

- Marsh LC.** Estimating the number and location of knots in spline regressions. *Journal Of Applied Business Research*, 1989, 2(3), 60-70.
- Mills D.** Improved mathematical methods for drug design, Institute of Food Research, UK, 1996, 54.
- Mood A, Graybill F, Boes D.** Introduction to the Theory of Statistics (3 rd ed.), McGraw-Hill, 1974, 229.
- Moreton HP.** Minimum curvature variation curves, networks and surfaces for fair free-form shape design, University of California Doctoral Thesis, Berkeley, 1992.
- Mulla ZD.** Spline regression in clinical research. *West Indian Medical Journal*, 2007, 56(1), 77-79.
- Oturanç G, Kurnaz A, Kiriş ME, Keskin Y.** Sayısal Analiz, 2. Baskı, Dizgi Ofset, Konya 2008, 397.
- Pan W.** Akaike's information criterion in generalized estimating equations. *Biometrics*, 2001, 57, 120–125.
- Park SH.** Experimental designs for fitting segmented polynomial regression models. *Technometrics*, 1978, 20(2), 151- 154.
- Parkhurst AM, Spiers DA, Hahn GL.** Spline models for estimating heat stress thresholds in cattle. *Annual Conference on Applied Statistics in Agriculture*, pp 148, 2002, Kansas State University.
- Pierce DA.** Least square estimation in the regression model with autoregressive-moving average errors. *Biometrika*, 1971, 58(2), 299-312.
- Poirier DJ.** Piecewise regression using cubic splines. *Journal of the American Statistical Association*, 1973, 68(343), 515-524.
- Porth RW.** Application of least square cubic splines to the analysis of edges. Rochester Institute of Technology Master of Science thesis, New York, 1984, 123.
- Powell MJD.** The local dependence of least squares cubic splines. *SIAM Journal on Numerical Analysis*, 1969, 6(3), 398-413.
- Ramsey JB.** Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society Series B (Methodological)*, 1969, 350-371.
- Rawlings D, John O, Pantula, Sastry G, Dickey, David A.** Applied Regression Analysis (2 nd ed.): A Research Tool, Springer Science & Business Media, 1998, 657.
- Ruppert D.** Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, 2002, 11(4), 735-757.

- Schoenberg IJ.** Contributions to the problem of approximation of equidistant data by analytic functions, *Quarterly of Applied Mathematics*, 1946, 4, 45–99 and 112–141.
- Seber GAF, Wild CJ.** Nonlinear Regression, John Wiley&Sons, New Jersey 2003, 768.
- Sheather SJ.** A Modern Approach to Regression with R, Springer Texts in Statistics, 2009, 331-369.
- Smith PL.** Splines as a useful and convenient statistical tool. *The American Statistician*, 1979, 33(2), 57-62.
- Schwarz, Gideon E.** Estimating the dimension of a model. *Annals of Statistics*, 1978, 6(2), 461-464.
- Schwetlick H, Schütze T.** Least squares approximation by splines with free knots. *BIT Numerical Mathematics*, 1995, 35(3), 361-384.
- Studden, William J, VanArman DJ.** Admissible designs for polynomial spline regression. *The Annals of Mathematical Statistics*, 1969, 40, 1557-1569.
- Şahin M.** Parçalı Regresyonlar Ve Tarımsal Alanlarda Kullanımı. Kahramanmaraş Sütçü İmam Üniversitesi Fen Bilimleri Enstitüsü Doktora Tezi, Kahramanmaraş, 2009.
- Türker ES, Can E.** Bilgisayar Uygulamalı Sayısal Analiz Yöntemleri, Değişim Yayınları, Adapazarı, 2002, 412.
- Ucal MŞ.** Ekonometrik model seçim kriterleri üzerine kısa bir inceleme. *ÇÜ İktisadi ve İdari Bilimler Dergisi*, 2006, 7(2), 41-57.
- Wainer H.** Piecewise regression, a simplified procedure. *British Journal of Mathematical and Statistical Psychology*, 1971, 24(1), 83-92.
- Wegman EJ, Wriqth IW.** Splines in statistics. *Journal of the American Statistical Association*, 1983, 78(382), 351-365.
- Wold S.** Spline function in data analysis. *Technometrics*, 1974, 16, 1-11.
- Young PH.** Generalized coefficient of determination. *The Journal of Cost Analysis & Management*, 2000, 2(1), 59-68.

## ÖZGEÇMİŞ

**Soyadı, Adı** : VAROL, Buğra  
**Uyruk** : T.C  
**Doğum yeri ve tarihi** : Konak / İZMİR, 17.07.1987  
**Telefon** : 0533 595 94 16  
**E-mail** : [bugravarol87@gmail.com](mailto:bugravarol87@gmail.com)  
**Yabancı Dil** : İngilizce (Konuşma: Orta, Yazma: İyi, Okuma: İyi)

### EĞİTİM

Derece	Kurum	Bulunduğu Yıllar
Lisans	İzmir Ege Üniversitesi, İstatistik Bölümü	2006 - 2010
Lise	Karşıyaka Şemikler Lisesi	2001 - 2004

### İŞ DENEYİMİ

Yıl	Yer/Kurum	Ünvan
2016 - Halen	İzmir Medical Park Hastanesi	İş Analisti