

35018

**REGRESYON MODELLERİNDE UÇ DEĞER İNCELEMEDE
DAĞILIMA BAĞLI OLMAYAN BİR YÖNTEM**

YÜKSEK LİSANS TEZİ
(İSTATİSTİK)

Mehmet YILDIZ
TEMMUZ 1996

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Bu tezin Yüksek Lisans tezi olarak uygun olduğunu onaylarım.



Danışman

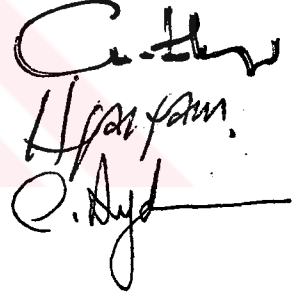
Doç. Dr. Hamza GAMGAM

Sınav Jürisi

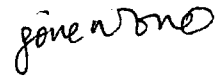
Başkan : Doç. Dr. Müslüm EKİNİ

Üye : Doç. Dr. Hamza GAMGAM

Üye : Yrd. Doç. Dr. Celal AYDIN



Bu Tez Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım
Esaslarına uygundur.



REGRESYON MODELLERİNDE UÇ DEĞER İNCELEMEDE

DAĞILIMA BAĞLI OLMAYAN BİR YÖNTEM

(Yüksek Lisans Tezi)

Mehmet YILDIZ

GAZİ ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

TEMMUZ 1996

ÖZ

Bu çalışmanın giriş kesiminde etkili değerler ve uç değerlerle ilgili bazı çalışmalara yer verilmiş ve bunlar için test istatistikleri tanıtılmıştır. Birinci bölümde ise, uç değerlerin budanması ve tahminlerin budanmış Bootstrap Yığınınından seçilecek örneklerle dayandırılması yöntemi tanıtılmıştır. Uç değerlerin örneğe girmesine engel olmak için Bootstrap Örneğinin çapını küçük seçmek, Bootstrap Yığınının az budanması ve Bootstrap Yığınının fazla budanması yöntemleri kullanılabilir. Bu çalışmada Bootstrap Yığınının budanması ile ilgilenilmiş ve budama yöntemleriyle ilgili test istatistikleri de tanıtılmıştır. Bu yöntem tahmin edilen artıkların Bootstrap dağılımına dayanır. Ayrıca Bootstrap tahminleri ve bozulmanın ölçüsü tanıtılmıştır. İkinci bölümde Bootstrap Yığınının budanması ile ilgili bir uygulamaya ve simülasyon çalışmasına yer verilmiştir. Uygulamada $\alpha=0.15$ için Bootstrap yığınınından budamalar yapılmış ve Budanmış Bootstrap yığını elde edilmiştir. Son bölümde ise bu uygulamadan elde edilen bulgulara yer verilmiştir.

Bilim kodu : 406.02.01

Anahtar kelimeler: Uç değer, Etkili değer, Artık tahmini, Bootstrap

Sayfa adedi : 67

Tez yöneticisi : Doç. Dr. Hamza GAMGAM

***A NON- DISTRIBUTIONAL METHOD FOR THE EXAMINATION
OF OUTLIER IN REGRESSION MODELS***

(M. Sc. Thesis)

Mehmet YILDIZ

GAZİ UNIVERSITY

INSTITUTE OF SCIENCE AND TECHNOLOGY

JULY 1996

ABSTRACT

In the introduction section of this study, several study related with influential value and outliers have been given, and for them the test statistics have been defined. In the first chapter, trimming of outliers and the method of making the estimations depend on the samples which will be selected from the trimmed Bootstrap Population have been explained. Selecting small Bootstrap sample size, mild trimming of the Bootstrap Population and severe trimming of the Bootstrap Population methods can be used to prevent including the outliers in the sample. In this study, trimming of the Bootstrap Population has been interested and also test statistics related with trimming methods have been defined. This method depends on Bootstrap distribution of the estimated residuals. Also, Bootstrap estimations and measure of contamination have been explained. In the second chapter, an application and a simulation study related with trimming of the Bootstrap Population have been performed. In the application, for $\alpha=0.15$ trims from Bootstrap Population have been made and trimmed Bootstrap Population has been obtained. In the last chapter, conclusions obtained from this application have been given.

Science Code : 406.02.01

Key Words : Outlier, Influential value, Predicted residuals, Bootstrap

Page Number : 67

Adviser : Doç. Dr. Hamza GAMGAM

TEŞEKKÜR

Yüksek Lisans tez çalışmamın tüm aşamalarında beni yönlendiren ve yardımlarını esirgemeyen çok değerli hocam ve tez danışmanım Doç. Dr. Hamza GAMGAM'a ve bu çalışmam esnasında yardım ve desteklerini esirgemeyen değerli arkadaşlarım Sayın Arzu Özmen'e, Sayın Ertan Burkan'a, mesai arkadaşlarıma, dostlarıma ve aileme sonsuz teşekkürler.



TABLolarIN LİSTESİ

<u>Tablo No</u>	<u>Sayfa</u>
Tablo 1. Bootstrap Örneđi İindeki en az bir U Deđerin $\hat{e}_{pi}$ 'yi Bozma Yüzdeleri	17
Tablo 2. Uygulama iin Veri Tablosu	39
Tablo 3. Veri iin Regresyon Analizi Tablosu	40
Tablo 4. Tablo 2'deki Verinin Budanmasına İlişkin İşlem Sonuçları	41
Tablo 5. $n=30$, $k=2$ iken Artıklar, Artık Kareler, v_{ii} , r_i ve t_i^2 'lere İlişkin İşlem Sonuçları	43
Tablo 6. $n=21$ ve $k=2$ iin Budanmış Bootstrap Yığınınına İlişkin Bazı İşlemlerin Sonuçları	50
Tablo 7. $n=21$, $k=2$ iin Budanmış Bootstrap Yığınınına İlişkin Bazı Regresyon Analizi Sonuçları	51
Tablo 8. Budanmış Bootstrap Yığını iin $n=21$, $k=2$ iken Artıklar, Artık Kareler, v_{ii} , r_i ve t_i^2 'ler	52
Tablo 9. Bootstrap Örneđinin apı $n_b=7$ iin Bootstrap Tahminlerine İlişkin Simölasyon Sonuçları Tablosu	53

ŞEKİLLERİN LİSTESİ

<u>Sekil No</u>	<u>Sayfa</u>
Şekil 1. $(y_j, x'_j) \in \text{ALT}$ için $hx_{\text{ALT}}(i,j) \gg 0$ ve $\lambda_j \geq 0$ iken (y_j, x'_j) 'nin $E_{y/x}(\hat{e}_{pi})$ 'yi etkilemesi	12
Şekil 2. $(y_j, x'_j) \in \text{ALT}$ için $hx_{\text{ALT}}(i,j) \ll 0$ ve $\lambda_j \leq 0$ iken (y_j, x'_j) 'nin $E_{y/x}(\hat{e}_{pi})$ 'yi etkilemesi	13
Şekil 3. j 'ler için $hx_{\text{ALT}}(i,j) \ll 0$ ve $\lambda_j < 0$ yada $hx_{\text{ALT}}(i,j) \gg 0$ ve $\lambda_j > 0$ ise (y_j, x'_j) 'nin $E_{y/x}(\hat{e}_{pi})$ 'yi etkilemesi	13
Şekil 4. $X \in \text{Bootstrap Örneği}$ için her hangi bir uç değerin $\sum_{xx}$, σ_{xy} , ve β 'yi etkilemesi	18
Şekil 5. Uç Y gözlem değerlerinin σ_{xy} ve β 'nin tahminini etkilemesi	19
Şekil 6. Bootstrap Yığının etkilemeyen fakat Bootstrap Örneğini etkileyen gözlemler	20
Şekil 7. Budama Kriterleri (1.8), (1.10) ve (1.12) için Budanmış Bootstrap Yığını	21
Şekil 8. Veri için Serpme Diyagramı	40
Şekil 9. Budanmış Bootstrap Yığını için Serpme Diyagramı.....	51

İÇİNDEKİLER

Sayfa No

ÖZET	I
ABSTRACT	II
TEŞEKKÜR	III
TABLoların LİSTESİ	IV
ŞEKİLLERİN LİSTESİ	V

GİRİŞ	1
-------------	---

BÖLÜM 1

UÇ DEĞER İNCELEMEDE DAĞILIMA BAĞLI OLMAYAN BİR YÖNTEM	15
--	----

1.1. Yeniden Örnekleme Yöntemi	15
1.2. Bazı Uç Değer Budama Yöntemleri	16
1.2.1. Bootstrap Örneğinin Çapını Küçük Seçmek	16
1.2.2. Bootstrap Yığınının Az Budanması	18
1.2.3. Bootstrap Yığınının Fazla Budanması	22
1.3. Bootstrap Tahminleri	32
1.4. Şartlı Hipotez Testi	35
1.5. Bozulmanın Ölçülmesi	36

BÖLÜM 2

UYGULAMA	38
----------------	----

BÖLÜM 3

SONUÇ VE TARTIŞMA 59

KAYNAKLAR 65

ÖZGEÇMİŞ 67



GİRİŞ

Regresyonda örneği oluşturan gözlemler serpm diyagramı üzerinde birer nokta ile temsil edilirler. Serpm diyagramındaki bu noktalar veri merkezine mahalonobis uzaklığına göre incelenirler. Serpm diyagramında bulunan noktaların bir yada daha fazlası veri merkezine uzaklıklarına göre etkili(influential) değer veya uç değer(outlier) olabilir. Örneği oluşturan gözlemlerden bazıları veri merkezine uzak olabilir. Bu gözlemler regresyon modeline uygun ise; bu tür gözlemlere etkili değer adı verilir. Etkili gözlemi işlem dışı tutarak parametre tahmini yapılabilir. Bu parametre tahmini ile tüm gözlemlerden bulunan parametre tahmin değeri arasında önemli ölçüde fark görülür. Örneği oluşturan gözlemler içinde bazı gözlemler veri merkezinin uzağında bulunabilir. Bu gözlemler regresyon modeline uygunluk göstermiyorsa; bu tür gözlemlere uç değer adı verilir. Uç değerler parametre tahminlerini önemli ölçüde değiştirebilen gözlemlerdir. Aynı zamanda artışı büyük olan gözlemlerdir.

Uç değer in olmadığı bir regresyon modeli;

$$Y_{n \times 1} = X_{n \times k} \beta_{k \times 1} + e_{n \times 1} \quad (0.1)$$

şeklindedir. Burada,

Y, yanıt vektörü,

X, k sayıda tahmin edici değişkenin oluşturduğu matris,

β , tahmin edici deęişkenlerle ilgili regresyon katsayıları vektörü,

ve

e , modelin rasgele hata vektörü
dür.

Uç deęer var ise bu durumda model,

$$Y_{n \times 1} = X_{n \times k} \beta_{k \times 1} + I_{n \times n} \lambda + e_{n \times 1} \quad (0.2)$$

formunda yazılabilir. Burada λ , veriye uygun model (0.2) iken, e 'nin dağılımındaki 'konum deęiştirme (location shift)' ile ilgili 'kayma parametresi (shift paramater)' dir(1).

Sadece bir uç deęerin olduęu varsayımı altında, $e \sim n(0, \sigma^2)$ iken, bu uç deęer ile ilgili gözlemin anlamlılığı için yapılabilecek testler oldukça çoktur(2,3,4,5).

Tek bir tane gözlemin veriden çıkartılmasına dayalı olan tanılama(teşhis etme) yöntemleri çoklu regresyon analizinde sıkça kullanılır. Birden fazla gözlemin veriden çıkartılması ile ilgili yöntemler de oldukça fazla olmalarına rağmen, kombinasyon sayısının fazla olması nedeniyle, çok az kullanılırlar. Fakat bazı durumlarda, birden fazla sayıda uç deęerin yapısı(pattern), tek bir tane gözlemin çıkartılması ile açığa çıkartılamaz. Bu gibi durumlarda gizleme-maskeleyme(masking) olduęu söylenir(6). Bu tür problemle karşılaşıldığında Atkinson(1986) uç deęerlerin incelenmesinde

gizlememe-maskelememe(unmasking) yönteminin kullanılmasını önermektedir(6). Bu sağlam(robust) yöntem, oldukça iyi özelliklere sahiptir ve uç değerlerin belirlenmesi için bir araştırma aracı olarak sunulmuştur. Birden fazla gözlemin veriden çıkartılması ile ilgili olan bu yöntem, etkili gözlemler ve uç değerlerin varlığının doğrulanması için uygulanabilir(6). Hawkins, Bradu, ve Kass(1984) normallik varsayımı altında çok sayıda uç değerlerin ortaya çıkarılması için yöntemler geliştirdiler(4).

Etkili gözlemler kümesinin belirlenmesi Andrews ve Pregibon(1978); Belsley, Kuh ve Welsch(1980); Cook ve Weisberg(1982); John ve Geisser(1983) ve Gray ve Ling(1984) tarafından incelenmiş ve bu çalışmalarda etkili gözlemler için önemli bulgular elde edilmiştir(7,8,9,2,10).

Eğer tek bir uç değer var ise λ vektörünün $n-1$ elemanı sıfır ve bir elemanı da sıfırdan farklıdır. Uç değer sayısı birden fazla ise uç değer sayısı kadar λ_i sıfırdan farklı olur. Örneğin, uç değer sayısı iki ise sıfırdan farklı λ_i 'nin sayısı da ikidir. λ_i (λ vektörünün i . elemanı) sıfıra eşit değilse i . gözlem uzak konumdadır. Bu noktanın ağırlığı sıfırdan farklı λ_i uzaklığına orantılıdır. λ vektöründe sıfırdan farklı λ_i değerlerinin sayısı veri içinde mevcut olan uç noktaların sayısını gösterir. Hangi gözlemlerin uzak konumda oldukları konusunda bir önsel bilgi var ise, bu durumda normallik varsayımı altında belirli bir anlamlılık düzeyinde ilgili λ_i 'lerin sıfırdan anlamlı şekilde farklı olup olmadıkları konusunda karar vermede yardımcı

olacak bir test mevcuttur(11). Eğer böyle bir önsel bilgi mevcut değilse, Bonferroni tekniği kullanılabilir(9, 12,13).

Örnek içinde uç değerler var ise, bilinen En Küçük Kareler tahmin edicisi,

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (0.3)$$

sapmalı bir tahmin edicidir. Yani X verilmiş iken ($E_{y/x}$ ile ifade edilen) β 'nın koşullu beklenen değeri,

$$E_{y/x}(\hat{\beta}) = \beta + (X'X)^{-1} X'\lambda \quad (0.4)$$

dır. Model (0.1) için bilinen artıklar, yani

$$\hat{e} = Y - X\hat{\beta} \quad (0.5)$$

uç değerlerin varlığı durumunda, sıfır ortalama vektörüne sahip değildir(14). Bu artıklar için,

$$E_{y/x}(\hat{e}) = \lambda - X(X'X)^{-1} X'\lambda \quad (0.6)$$

dir. Bundan dolayı veri içinde uç değer olmasının hem β 'nın En Küçük Kareler tahminini hem de artıkları oldukça etkilediği açıktır. Bu etki gizlenen gözlemlerin tespit edilememesi durumunda çok fazla olabilir. Örneğin X'in i. satırı x_i olmak üzere,

$$\lambda_i - x_i(X'X)^{-1} X'\lambda = 0 \quad (0.7)$$

ise, uzakta olan i. gözlem ($\lambda \neq 0$) gizlenebilir(masking). Uç değer olduğu düşünülen bir gözlemin gerçekte uç değer olup olmadığının belirlenmesi için başka bir test istatistiği daha önerilmiştir(1). Diğer yandan, eğer

$$x_i(X'X)^{-1} X'\lambda \quad (0.8)$$

sıfırdan anlamlı derecede farklı ise i. gözlem veri içinde yutulabilir(swapping). Bu nedenle gizlemenin veya yutulmanın olduğu durumlarda, uç değerleri incelerken sadece bilinen artıkları kullanmak pek sağlıklı olmayabilir(1).

Bu uç değerlerin araştırılmasında önemli olabilecek bazı özet bilgiler ve student tipi artıklarla ilgili bilgiler aşağıda verilecektir.

v_{ii} , i. gözlemin veri merkezine olan uzaklığının ölçüsü,

v_{ij} , i. ve j. gözlemlerin veri merkezine uzaklığının ölçüsü,

x_i , X'in i. satırı

ve

x_j , X'in j. satırı

olmak üzere,

$$v_{ii} = x_i'(XX)^{-1}x_i \quad (0.9)$$

ve

$$v_{ij} = x_i'(XX)^{-1}x_j \quad (0.10)$$

dir. Student tipi artıklar,

$$r_i = \frac{\hat{e}_i}{\hat{\sigma}(1-v_{ii})^{1/2}} \quad (0.11)$$

şeklinde tanımlanır. Burada

$$\hat{\sigma} = \left(\frac{\sum (Y_i - \hat{Y}_i)^2}{n-p-1} \right)^{1/2} \quad (0.12)$$

dir. $E(r_i) = 0$ ve $Var(r_i) = 1$ dir. Artıkların varyansı,

$$Var(\hat{e}_i) = (1-v_{ii})\hat{\sigma}^2 \quad (0.13)$$

kovaryansı,

$$\text{Cov}(\hat{e}_i, \hat{e}_j) = -v_{ij}\hat{\sigma}^2 \quad (0.14)$$

ve korelasyonu ,

$$\text{Kor}(\hat{e}_i, \hat{e}_j) = \frac{-v_{ij}}{(1-v_{ii})^{1/2}(1-v_{jj})^{1/2}} \quad (0.15)$$

şeklinde tanımlanır. v_{ii} ve v_{ij} 'ler $V_{n \times n}$ 'nin elemanları olsun. $V_{n \times n}$ matrisi X 'e bağlı olarak

$$V_{n \times n} = X(X'X)^{-1}X' \quad (0.16)$$

yazılabilir. $V_{n \times n}$ simetrik ve idempotenttir.

$$V_{n \times n} = V_{n \times n}' \quad (0.17)$$

ve

$$V_{n \times n} = V_{n \times n} V_{n \times n} \quad (0.18)$$

eşitliği açıktır(14). Tam ranklı modellerde $\text{Rank}(V) = \text{rank}(X) = p'$ dir.

Köşegen elemanlarının izi(trace) ranka eşittir. Aşağıdaki gibi gösterilebilir.

$$\sum_{i=1}^n v_{ii} = \text{rank}(X) = p' \quad (0.19)$$

Bununla beraber,

$$\sum_i v_{ij}^2 = v_{ii} \text{ ve } \sum_i v_{ij} = \sum_j v_{ij} = 1$$

sonucu gösterilmiştir(14). Etkili gözlemler için hipotezler,

$$H_0: i .\text{gözlem etkili değildir.} \quad (0.20)$$

$$H_1: i. \text{ gözlem etkilidir.} \quad (0.21)$$

formunda yazılabilir. H_0 hipotezini test etmek için aşağıdaki gibi tanımlanan

Cook uzaklığından yararlanılır(9,14)

$$D_i = \frac{(\hat{\beta}_{-i} - \hat{\beta})'(X'X)(\hat{\beta}_{-i} - \hat{\beta})}{p'\hat{\sigma}^2} \quad (0.22)$$

$\hat{\beta}_{-i}$, Etkili olduğu düşünölen gözlemi işlem dışı bıraktıktan sonra bulunan parametre tahmin vektörü,

$\hat{\beta}$, Örneđi oluşturan gözlemlerin tümünden tahmin edilen parametre tahmin vektörü,

D_i , Cook uzaklıđı,

p , Tahmin edici deđişken sayısı,

p' , Veri matrisinden tahmin edilen parametre sayısı

$(p'=p+1)$

dır.

$D_i \sim F_{p', n-p}$ dağılımına sahiptir. O halde D_i uzaklıđı, $F_{p', n-p, \alpha}$ tablo deđeri ile karşılaştırılır. Bu kritere göre i . gözlemin etkili gözlem olup olmadığına karar verilir. Cook uzaklıđının hesaplama formölüne ulaşmak için aşağıdaki özdeşlikten yararlanılır.

$$X'_{-i}X_{-i} = (X'X)^{-1} + \frac{(X'X)^{-1}x_i x'_i (X'X)^{-1}}{(1 - v_{ii})} \quad (0.23)$$

(0.23) özdeşliđini soldan x'_i ile sağdan x_i ile çarpılırsa ,

$$x'_i X'_{-i} X_{-i} x_i = x'_i (X'X)^{-1} x_i + \frac{\{x'_i (X'X)^{-1} x_i\} \{x'_i (X'X)^{-1} x_i\}}{(1 - v_{ii})}$$

$v_{ii} = x'_i (X'X)^{-1} x_i$ 'nin tanımından,

$$\begin{aligned} &= v_{ii} + \frac{(v_{ii} v_{ii})}{(1 - v_{ii})} \\ &= \frac{v_{ii} (1 - v_{ii}) + v_{ii}^2}{(1 - v_{ii})} \end{aligned}$$

$$\begin{aligned}
&= \frac{v_{ii}(1 - v_{ii} + v_{ii})}{(1 - v_{ii})} \\
&= \frac{v_{ii}}{(1 - v_{ii})}
\end{aligned} \tag{0.24}$$

Cook uzaklığı olan D_i 'ye geçişi sağlamak için aşağıdaki

$$\hat{\beta}_{-i} - \hat{\beta} = \frac{(X'X)^{-1}x_i\hat{e}_i}{(1 - v_{ii})} \tag{0.25}$$

ve

$$\hat{\sigma}_{-i}^2 = \frac{\hat{\sigma}^2(n - p' - r^2)}{(n - p' - 1)} \tag{0.26}$$

eşitliklerinden yararlanılır. Burada,

$\hat{\sigma}_{-i}^2$, Etkili olduğu düşünülen gözlemi işlem dışı bıraktıktan sonra bulunan varyansın tahmini,

dir. D_i test istatistiği aşağıdaki gibi tanımlanırsa;

$$\begin{aligned}
D_i &= \frac{\hat{e}_i x_i' (X'X)^{-1} (X'X) (X'X)^{-1} x_i \hat{e}_i}{(1 - v_{ii}) p' \hat{\sigma}^2 (1 - v_{ii})} \\
D_i &= \frac{\hat{e}_i \{x_i' (X'X)^{-1} x_i\} \hat{e}_i}{(1 - v_{ii}) p' \hat{\sigma}^2 (1 - v_{ii})} \\
D_i &= \frac{\hat{e}_i^2 v_{ii}}{p' \hat{\sigma}^2 (1 - v_{ii}) (1 - v_{ii})} \\
D_i &= \frac{r_i^2 v_{ii}}{p' (1 - v_{ii})}
\end{aligned} \tag{0.27}$$

$D_i \sim F_{p', n-p}$ dağılır. Etkili gözlem için test istatistiği olan

D_i 'nin değeri ile $F_{p', n-p, \alpha}$ karşılaştırılarak karar verilir.

Uç değeri, diğer gözlemlere göre artışı oldukça büyük olan gözlemdir. Y_i gözlemi regresyon modeline uyar fakat $\hat{\epsilon}_i$ oldukça büyük olabilir. $\hat{\epsilon}_i$ rasgele değişken olduğu için ilgili gözlemlerin artıkları büyük olabilir. Bu tip gözlemler modelden atılamaz. Gözlemler ölçülürken alet, teknisyen v.b. sebeplerden dolayı ölçme hatası yapılabilir. Bundan dolayı $\hat{\epsilon}_i$ büyük çıkabilir. Bu tip gözlemleri iyi incelemek gerekir. Ölçme hatası yapılmamış olabilir, modele ilave etmek gerekebilir. Şayet ölçme hatası yapılmış ise hatanın sebebi araştırılmalı ve giderilmelidir. Eğer, hata giderilemiyorsa gözlem modelden atılabilir. Y_i regresyon modeline uymuyorsa; gözlemde hatalı değilse, bu tip gözlemler önemlidir. Üzerinde çalışılması gerekir. Bu çalışma neticesinde gözlemlerin modele uygunluğuna karar verilebilir.

Uç değeri olduğu düşünülen i . gözlem modelden çıkarılır. i . gözlem modelden çıkarıldıktan sonra $\hat{\beta}_{-i}$ ve $\hat{\sigma}_{-i}^2$ hesaplanır. Örneği oluşturan ve uç değeri olduğu düşünülen gözlem değerlerine ilişkin aşağıda bazı hesaplama formülleri verilecektir. Bu formüller ile yeni bir test istatistiği olan t_i 'ye geçiş için bazı formüllerden yararlanılacaktır. $\hat{Y}_i$ 'nin varyansı,

$$\text{Var}(\hat{Y}_i) = \sigma^2 + \sigma_e^2 \quad (0.28)$$

ve uç değeri olduğu düşünülen gözlemi çıkardıktan sonra Y_i 'nin tahmini olan $\hat{Y}_{-i}$, aşağıdaki gibi,

$$\hat{Y}_i = X_i' \hat{\beta}_{-i} \quad (0.29)$$

ifade edilebilir. $\hat{Y}_i$ 'nin varyansı,

$$\text{Var}(\hat{Y}_i) = X_i'(X_{-i}'X_{-i})^{-1}X_i\hat{\sigma}_{-i}^2 \quad (0.30)$$

dir. Test istatistiği olan t_i için, student tipi artıklardan yararlanılmaktadır.

$Y_i - \hat{Y}_i$ ile ifade edilen artıkların varyansı,

$$S_{Y_i - \hat{Y}_i}^2 = \hat{\sigma}_{-i}^2 (1 + X_i'(X_{-i}'X_{-i})^{-1}X_i) \quad (0.31)$$

dir. Test istatistiği olan t_i ,

$$t_i = \frac{(Y_i - \hat{Y}_i) - E(Y_i - \hat{Y}_i)}{\hat{\sigma}_{-i}(1 + X_i'(X_{-i}'X_{-i})^{-1}X_i)^{1/2}} \quad (0.32)$$

formunda tanımlanmıştır. $E(Y_i - \hat{Y}_i) = 0$ olduğundan,

$$t_i = \frac{(Y_i - \hat{Y}_i)}{S_{Y_i - \hat{Y}_i}} \quad (0.33)$$

elde edilir. Student tipi artıklardan

$$r_i = \frac{\hat{e}_i}{\hat{\sigma}_{-i}(1 + X_i'(X_{-i}'X_{-i})^{-1}X_i)^{1/2}} \quad (0.34)$$

t_i test istatistiğine

$$t_i = r_i \left(\frac{n - p' - 1}{n - p' - r_i^2} \right)^{1/2} \quad (0.35)$$

ile geçiş yapmak mümkündür. Bu istatistik uç değeri belirlemek için kullanılır. (0.33)'deki t istatistiğinin $t_i \sim t_{n-p'-1}$ olduğu gösterilmiştir(14).

Uç değeri için hipotezler aşağıdaki gibi tanımlanır.

$$H_0 : i. \text{ gözlem uç değeri değildir.} \quad (0.36)$$

$$H_1 : i. \text{ gözlem uç değeridir.} \quad (0.37)$$

(0.36)'daki hipotezi test etmek için Bonferroni tarafından önerilen testten de yararlanılabilir(9,12,13,14). Her gözlem için birer t_i değeri bulunacağından her bir testin anlamlılık düzeyi α' dır. Burada $\alpha' = \alpha/n$ 'dir.

Bunlardan en az birinin yanlışlıkla uç değer vermesi olasılığı ise $n\alpha'$ olur(14).

Buraya kadar n gözlem içinde uç değer olduğu düşünülen gözlemlerin tek tek incelenmesi üzerinde duruldu. Bir kaç gözlemin oluşturduğu bir alt kümenin incelenmesi durumunda; uç değer ile ilgili bazı önemli durumlar ortaya çıkmaktadır(1).

ALT, n sayıda olan gözlemlerin herhangi bir alt kümesi olsun. λ_{ALT} ve Y_{ALT} sırasıyla λ ve Y 'nin ilgili alt vektörleri olup X_{ALT} da X 'in ilgili alt matrisidir. Yalnız ALT içindeki gözlemleri kullanarak $\hat{\beta}$ 'nin En Küçük Kareler tahmini,

$$\hat{\beta}_{ALT} = (X'_{ALT} X_{ALT})^{-1} X'_{ALT} Y_{ALT} \quad (0.38)$$

ve $\hat{\beta}_{ALT}$ 'in koşullu beklenen değeri,

$$E_{y/x}(\hat{\beta}_{ALT}) = \beta + (X'_{ALT} X_{ALT})^{-1} X'_{ALT} \lambda_{ALT} \quad (0.39)$$

dır. ALT herhangi bir uç noktayı içermezse $\lambda_{ALT} = 0$ ve $\hat{\beta}_{ALT}$ 'da β 'nin yansız tahminidir.

$(y_i, x'_i) \notin ALT$, yani (y_i, x'_i) , ALT'ın bir elemanı olmamak üzere; i . gözlem için tahmin edilen artık

$$\hat{e}_{pi} = y_i - x'_i \hat{\beta}_{ALT} \quad (0.40)$$

olarak tanımlanır. Burada $\hat{e}_{pi}$, p sayıda uç değer içinde i . gözlem için artık tahminidir.

Eğer ALT hiç bir uç noktayı içermezse (yani $\lambda_{ALT} = 0$ vektörü ise)

$$E_{y/x}(\hat{e}_{pi}) = \lambda_i \quad (0.41)$$

dir. Eğer ALT uç noktaları içerirse (yani, $\lambda_{ALT} \neq 0$ ise),bu durumda ,

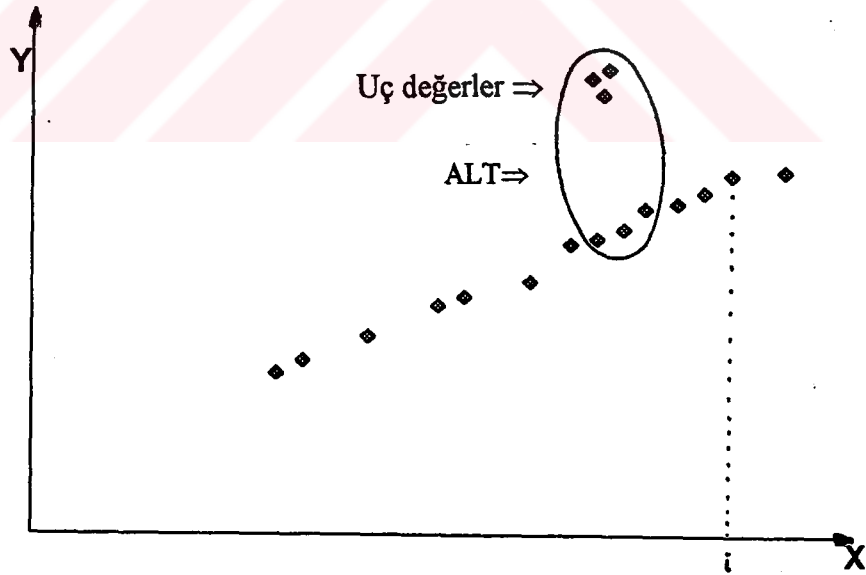
$$E_{y/x}(\hat{e}_{pi}) = \lambda_i - \sum_{(y_j, x'_j) \in ALT} h_{X_{ALT}}(i, j) \lambda_j \quad (0.42)$$

dır. Burada

$$h_{X_{ALT}}(i, j) = x'_i (X'_{ALT} X_{ALT})^{-1} x_j \quad (0.43)$$

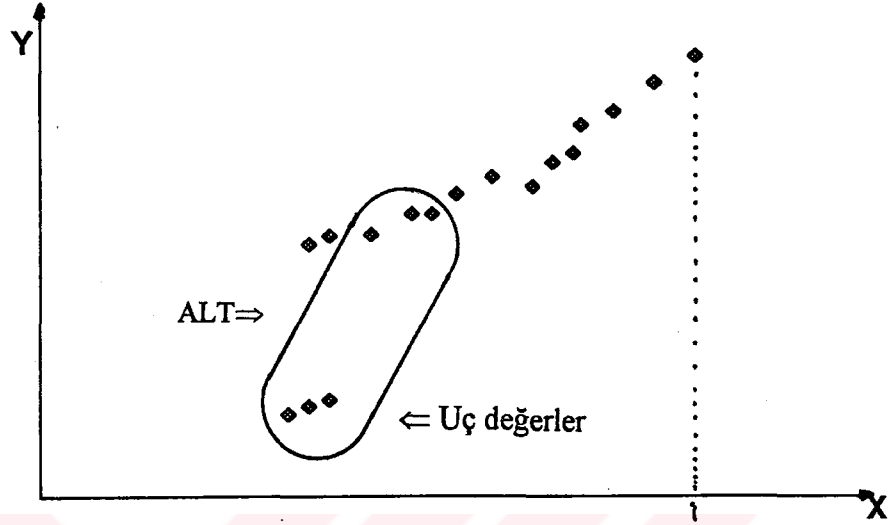
ve $(y_j, x'_j) \in ALT$ 'dir. $h_{X_{ALT}}(i, j)$ terimi regresyonda ALTUZAY (SUBSPACE) dizaynının ortak etkisinin etkileşimini saptamada rol oynar. Daha önce X 'in i . satırı x_i olmak üzere, $x_i (X'X)^{-1} X' \lambda$ 'nın sıfırdan önemli derecede farklı olması durumunda, uzaktaki i noktasının yutulabileceği (batırılıp yok olacağı) ifade edilmişti. Buna ilişkin farklı durumlar aşağıda verilmiştir($\lambda_i=0$).

Bütün $(y_j, x'_j) \in ALT$ için $h_{X_{ALT}}(i, j) \gg 0$ ve $\lambda_j \geq 0$ olduğunda, $E_{y/x}(\hat{e}_{pi})$ negatif ve büyük çıkabilir.(Bakınız şekil 1)



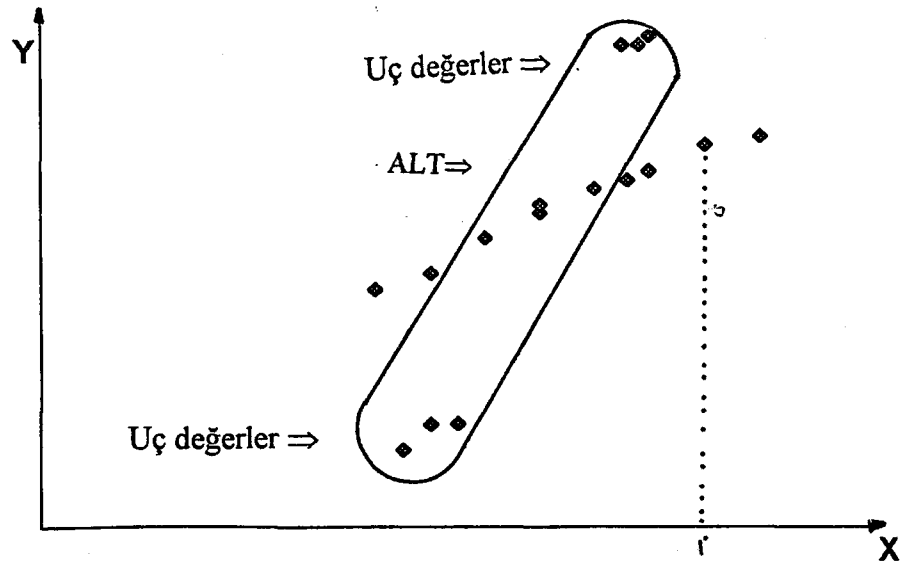
Şekil 1. $(y_j, x'_j) \in ALT$ için $h_{X_{ALT}}(i, j) \gg 0$ ve $\lambda_j \geq 0$ iken (y_j, x'_j) 'nin $E_{y/x}(\hat{e}_{pi})$ 'yi etkilemesi

Bütün $(y_j, x'_j) \in \text{ALT}$ için $hx_{\text{ALT}}(i,j) \ll 0$ ve $\lambda_j \leq 0$ olduğunda, $E_{y/x}(\hat{e}_{pi})$ negatif ve büyük olabilir. (Bakınız şekil 2)



Şekil 2. $(y_j, x'_j) \in \text{ALT}$ için $hx_{\text{ALT}}(i,j) \ll 0$ ve $\lambda_j \leq 0$ iken (y_j, x'_j) 'nin $E_{y/x}(\hat{e}_{pi})$ 'yi etkilemesi

Bütün j .'ler için $hx_{\text{ALT}}(i,j) \ll 0$ ve $\lambda_j < 0$ yada $hx_{\text{ALT}}(i,j) \gg 0$ ve $\lambda_j > 0$ olduğunda $E_{y/x}(\hat{e}_{pi})$ negatif ve büyük olabilir. (Bakınız şekil 3)



Şekil 3. j .'ler için $hx_{\text{ALT}}(i,j) \ll 0$ ve $\lambda_j < 0$ yada $hx_{\text{ALT}}(i,j) \gg 0$ ve $\lambda_j > 0$ iken j 'nin $E_{y/x}(\hat{e}_{pi})$ 'yi etkilemesi

$E_{y/x}(\hat{e}_{pi})$ 'nin pozitif ve büyük olduğu durumlarda da yukarıda verilen $h_{X_{ALT}(i,j)}$ (yada λ_j)'nin eşitsizliklerinin tersi olur.

O halde bu adımlar $E_{y/x}(\hat{e}_{pi})$ 'nin bozulmasını engelleyen(sınırlayan) ALT'ları seçmek için kullanılabilir. Eğer bu yapılırsa; $\hat{e}_{pi}$, λ_i 'nin makul(uygun) bir tahmin edicisi olacaktır. Daha sonra β 'nın tahmininde kullanılacak olan ALT'ı oluştururken (y_i, x'_i) , $i=1,2,...,n$, sorunlu gözlemlerin seçimine engel olunursa, bu özellik yani, λ_i 'nin makul tahminlerini veren $\hat{e}_{pi}$ 'ler elde edilecektir(15).

Regresyon modeli bozulmuş veride (y_i, x'_i) 'lerin varyans kovaryans matrisi,

$$\begin{bmatrix} \sigma_{yy} & \sigma_{yx} \\ \sigma_{xy} & \sigma_{xx} \end{bmatrix}$$

ile bağımsız dağıldıklarını fakat bozulmamış veride aynı varyans kovaryans matrisi ve

$$(\mu_y, \mu'_x)$$

ortalama vektörü ile benzer bağımsız dağıldıklarını varsayar. Artıklarla ilgili değişen varyans(heteroscedasticity)'in araştırılmasında, eşit varyans kovaryans matrisi varsayımı atılmalıdır.

Bu çalışmada, uç değerler varken bu uç değerlerden kurtulmanın yolları, uygun $\hat{e}_{pi}$ 'lere nasıl ulaşılacağı yani ALT'lar ile uygun gözlemler kümesinin belirlenmesi ve incelenmesinin üzerinde durulacaktır.

BÖLÜM 1

UÇ DEĞER İNCELEMEDE DAĞILIMA

BAĞLI OLMAYAN BİR YÖNTEM

Regresyon analizinde gözlemler serpmeye diyagramı üzerinde konumları itibariyle uzakta iseler bunlar giriş bölümünde özelliklerine göre etkili değer ve uç değer olarak isimlendirilmiştir. Bu çalışmanın amacı, sonuçta β 'nın tahmininde kullanılacak ALT'ı seçerken sorunlu olan gözlemlerin(uç değerlerin) seçimini engelleyerek, uygun(makul) $\hat{e}_{pi}$ 'leri veren ALT'ın bulunmasıdır. Bu konu ile ilgili olan bilgilere daha sonra yer verilecektir. Uç değerler belirlenirken budama yöntemlerinden yararlanılır. Önce yeniden örnekleme yöntemlerinden olan Bootstrap yöntemi ve bazı budama yöntemleri sırasıyla açıklanacaktır.

1.1. Yeniden Örnekleme Yöntemi

Yeniden örnekleme yöntemlerinin amacı λ_i 'yi $i=1,2,...,n$ tahmininde kullanılacak olan $\hat{e}_{pi}$ 'nin dağılımını tahmin etmektir. λ_i 'nin tahmini olarak kullanılacak olan $\hat{e}_{pi}$ 'nin dağılımının incelenmesinde, Sparks(1988) yeniden örnekleme yöntemlerinden Bootstrap'ı kullanmıştır(1). Diğer yeniden örnekleme yöntemleri kullanılabilirse de bu çalışmada yeniden örnekleme yöntemlerinden Bootstrap kullanılacaktır. Bootstrap Yığını(B.Y.), $(y_1, x'_1), ..., (y_n, x'_n)$ orijinal örnek gözlemlerinin bir kümesidir. Bootstrap Örneği(B.Ö.) ise, Bootstrap Yığınının yerine koyarak yöntemi ile

oluşturulan rasgele bir örnek(bir alt küme) olarak tanımlanmıştır(1). n_b , Bootstrap Örneğinin çapını gösterebilir. Eğer, tüm $(y_j, x'_j) \in$ Bootstrap Örneği için $\lambda_j \neq 0$ ise bu durumda bu Bootstrap Örneği kullanılarak hesaplanan $\hat{e}_{pi}$, genellikle λ_i 'nin sapmalı tahminini verecektir. n sayıda orijinal gözlem arasında p sayıda uç değer olduğu varsayılırsa, bu durumda en az bir uç değer

$$1 - \left(\frac{n-p}{n}\right)^{n_b} \quad (1.1)$$

olasılıkla $\hat{e}_{pi}$ 'yi bozacaktır. Bu duruma ilişkin bazı örnekler için tablo 1 verilmiştir.

Bu bozulmayı azaltan ve λ_i 'nin daha sağlam(robust) tahminlerini veren yöntemlerden bazıları sırasıyla açıklanacaktır.

1.2. Bazı Uç Değer Budama Yöntemleri

1.2.1. Bootstrap Örneğinin Çapını Küçük Seçmek

Bozulmanın azalmasını sağlayan yöntemlerden biri Bootstrap Örneğinin çapını küçük seçmektir. Diğer yandan n_b için çok küçük bir değer, tahmin edilen artıkları anlamsız duruma getirebilir. Regresyonda tahminin ortalama artığını, regresyonun kullanılmadığı duruma göre daha küçük yapmak için, n_b aşağıdaki eşitsizliği sağlayacak şekilde seçilebilir(1).

$$n_b > (k / \rho^2) + 2 \quad (1.2)$$

Burada

k , Tahmin edilen parametre sayısı,

dır.

Tablo 1. Bootstrap Örneği içindeki en az bir uç değerin $\hat{e}_{pi}$ 'yi bozma yüzdeleri

B.Ö. çapları n_b	B.Ö. içindeki uç değer yüzdeleri			
	5%	10%	15%	20%
4	18.55	34.39	47.80	59.04
5	22.62	40.95	55.63	67.23
6	26.49	46.86	62.29	73.79
7	30.17	52.17	67.94	79.03
8	33.66	56.95	72.75	83.22
9	36.98	61.26	76.84	86.66
10	40.13	65.13	80.31	89.26
11	43.12	68.62	83.27	91.41
12	45.96	71.76	85.78	93.13
13	48.67	74.58	87.91	94.50
14	51.23	77.12	89.72	95.60
15	53.67	79.41	91.26	96.48
16	55.99	81.47	92.57	97.19
17	58.19	83.32	93.69	97.75
18	60.28	84.99	94.64	98.20
19	62.26	86.49	95.44	98.56
20	64.15	87.84	96.12	98.85

Tanım (1.2)'deki ρ^2 'nin değeri

$$\rho^2 = \frac{\sigma'_{xy} \Sigma_{xx}^{-1} \sigma_{xy}}{\sigma_{yy}} \quad (1.3)$$

dir. Bu eşitlikte

σ_{xy} , X rasgele tahmin edici değişkenleri ile Y yanıt değişkeni arasındaki kovaryans vektörü,

σ_{yy} , Y yanıt değişkeninin standart sapması,

Σ_{xx}^{-1} tahmin edici değişkenlerin varyans kovaryans matrisinin tersidir.

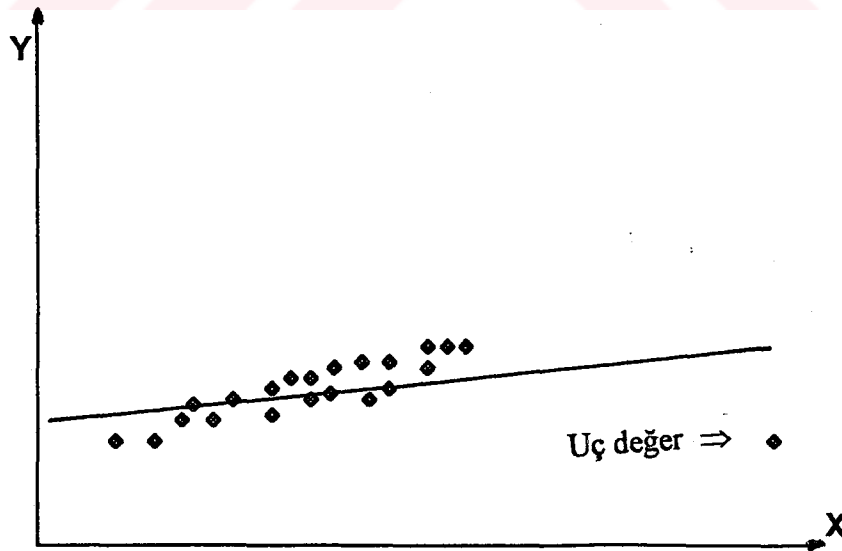
1.2.2. Bootstrap Yığınının Az Budanması

Bu yöntemin, Huber (1973), Koenher ve Bassett (1982), De Jongh ve De Wet(1985) tarafından kullanılan yöntemlerle benzerlikleri vardır(16,17,18). Bootstrap Örneğine dahil olan ve serpmeye diyağramının uzağındaki gözlemlere sıfır olasılık vermeye çalışılır. Bunun diğer yolu Budanmış Bootstrap Yığınının bir Bootstrap Örneği seçmektir. Burada Budanmış Bootstrap Yığını, β 'nın tahminini önemli derecede etkileyebilen herhangi bir gözlemi dışarda tutan Bootstrap Yığını olarak tanımlanır.

X ve Y arasındaki lineer ilişkiyi maksimize eden β vektörü,

$$\beta = \Sigma_{xx}^{-1} \sigma_{xy} \quad (1.4)$$

dir(1). Tahmin edici değişken X Bootstrap Örneğinin elemanı olmak üzere, herhangi bir uç değer, Σ_{xx} , σ_{xy} ve β 'nın tahminlerini önemli derecede etkileyebilir. Bu durum Şekil 4'te görülmektedir.



Şekil 4. $X \in$ Bootstrap Örneği için her hangi bir uç değer Σ_{xx} , σ_{xy} ve β 'yı etkilemesi

$F_{k-1, n-k}^{\alpha}$ değeri $k-1$ ve $n-k$ serbestlik dereceli F dağılımının sol tarafında α kadar olan alanın üst sınır değeri olmak üzere, budama kriterinde kullanılan $C_k(\alpha)$ değeri

$$C_k(\alpha) = (k-1)F_{k-1, n-k}^{\alpha} \quad (1.6)$$

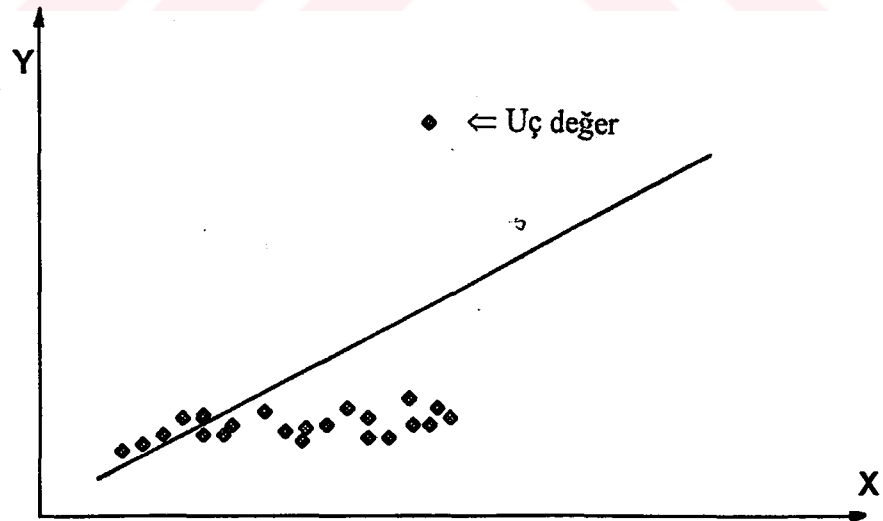
dır. X gözlemleri için test istatistiği

$$h_x(i, i) = \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.7)$$

olmak üzere, budama kriteri olan aşağıdaki eşitsizlik doğru ise Bootstrap Yığınınından gelen (y_i, x'_i) budanmalıdır.

$$h_x(i, i) > \frac{(n-1)C_k(\alpha)}{n(1-C_k(\alpha))} + \frac{1}{n} \quad (1.8)$$

Benzer şekilde serpmeye diyagramının uzağındaki Y değerleri σ_{xy} 'nin tahminini ve dolayısıyla β 'nin tahminini etkiler. (Bakınız şekil 5)



Şekil 5. Uç Y gözlem değerlerinin σ_{xy} ve β 'nin tahminini etkilemesi

Y gözlemleri için test istatistiği

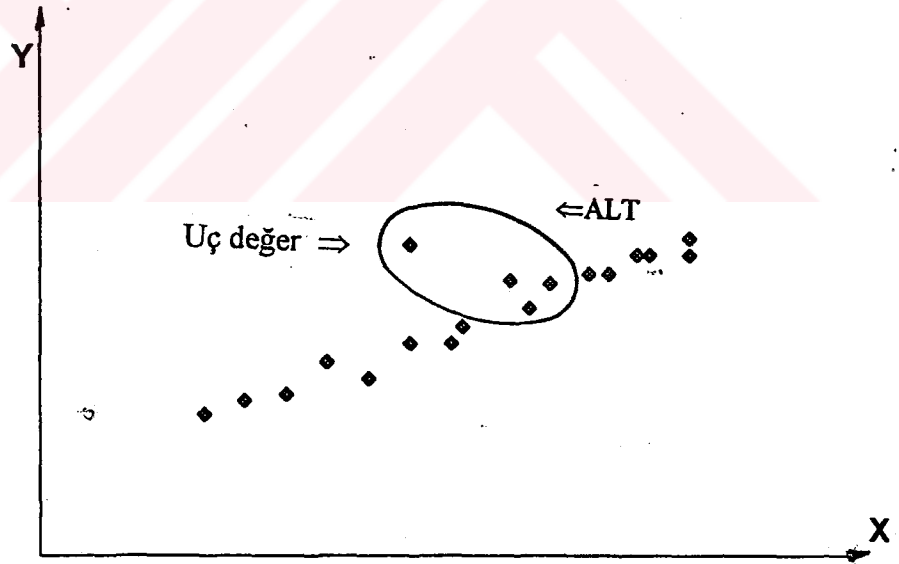
$$h_y(i,i) = \frac{(y_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (1.9)$$

olmak üzere, budama kriteri olan aşağıdaki eşitsizlik doğru ise,

$$h_y(i,i) > \frac{(n-1)C_k(\alpha)}{n(1-C_k(\alpha))} \quad (1.10)$$

Bootstrap Yığınının gelen (y_i, x'_i) 'ler budanmalıdır.

Bir gözlemin β 'nın Bootstrap Yığını tahminini fazla etkilememesi mümkün iken, bazı Bootstrap Örneklerinde β 'nın tahminini oldukça etkilemesi mümkün olabilir.(Bakınız şekil 6)



Şekil 6. Bootstrap Yığınını etkilemeyen fakat
Bootstrap Örneğini etkileyen gözlemler

(0.11), (0.12) ve (0.35)'deki r_i , σ ve t_i 'ye ilişkin tanımlardan t_i^2 test istatistiği için

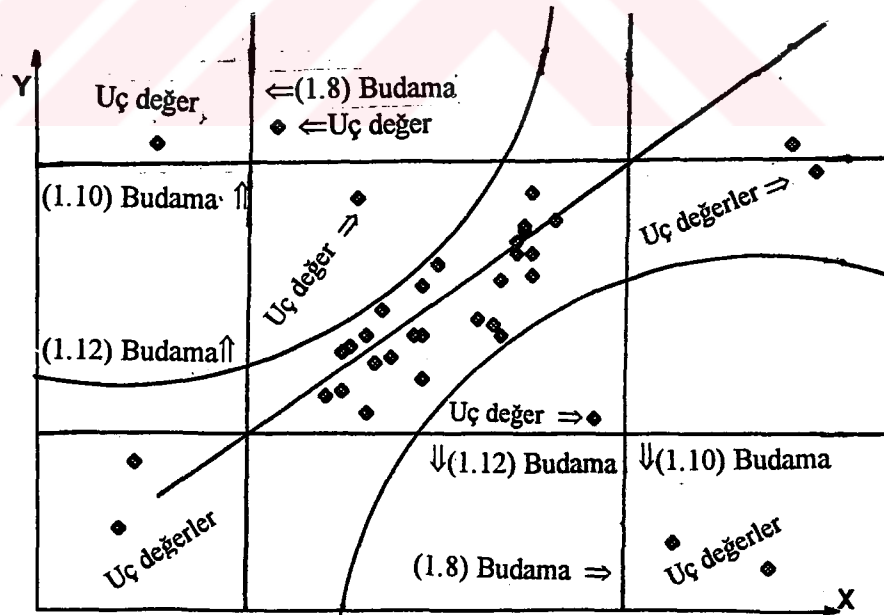
$$t_i^2 = r_i^2 \left(\frac{n - p' - 1}{n - p' - r_i^2} \right) \quad (1.11)$$

yazılabilir. Student tipi artıklar için t_i^2 test istatistiği ve $F_{k-1, n-k, \alpha}$ değeri F dağılımının sağ tarafında α kadar alanın alt sınır değeri olmak üzere,

$$t_i^2 > F_{k-1, n-k, \alpha} \quad (1.12)$$

ise Bootstrap Yığınının gelen (y_i, x'_i) budanmalıdır.

(1.8), (1.10) veya (1.12)'yi sağlayan herhangi bir gözlem Bootstrap Yığınının budanır. Böylece Budanmış Bootstrap Yığını(B.B.Y.) elde edilir.(Bakınız şekil 7)



Şekil 7. Budanmış Bootstrap Yığını

Eğer gözlemlerin %15'ten azı dışarıda kalıyorsa $\alpha=0.10$ ile bu budamalar serpmeye diyağramının uzağında kalan gözlemleri içermeyen bir Budanmış Bootstrap Yığınınını verir. Fakat bu uç noktaların yüzdesi %15'ten daha büyük ise, genelde daha fazla budamaya ihtiyaç duyulacaktır. Fazla budama ile ilgili açıklama aşağıda 1.2.3.'te verilecektir.

1.2.3. Bootstrap Yığınının Fazla Budanması

Bu yöntemde fazla budama, yukarıdaki 1.2.2.'de α 'nın değerini arttırmakla başarılabilir. Bununla beraber, uzaktaki bir çok gözlem dar bir bölgede kümelenmiş ise bu strateji iyi bir performans gösteremediğinden, bu yöntem pek önerilmemektedir. Sebebi ise, kümelenmiş gözlemlerde uç değeri belirlerken uç değeri olmayanın da uç değeri olarak belirlenebileceğinden, hata yapılmaması için fazla önerilmemektedir. Fazla budama yöntemi, daha çok dar bir bölgede değil, geniş bir bölgeye yayılmış gözlemler için önerilmektedir.

Fazla budama yöntemlerinden üçü aşağıdaki gibidir.

Birincisi, $\alpha = 0.10$ olarak 1.2.2.'deki yöntemler ile Bootstrap Yığınının ardışık olarak budanmasıdır. Her budama adımından sonra bozulma kontrol edilmelidir. Bozulmanın ölçüsü Kısım 1.5'te anlatılacaktır. Eğer sonuçta Budanmış Bootstrap Yığını bozulmuş ise, bu durumda Budanmış Bootstrap Yığınınını budamaya devam edilir, aksi halde budama durdurulur ve böylece Budanmış Bootstrap Yığını elde edilir.

Bu yaklaşım hesaplama yönüyle ayrıntılı olabilir. Buna rağmen, eğer akılcıca kullanılabilirse, bozulmamış Budanmış Bootstrap Yığınının bu yaklaşımla ulaşılabilir, fakat garanti edilemez. Gözlemlerin yayılması dar bir bölgede değilse, geniş bir bölgede ise, bozulmamış Budanmış Bootstrap Yığınının yaklaşım dar bir bölgeye göre daha iyi kontrol edilebilir.

İkincisi; En Küçük Medyan Kareleri kullanan Atkinson'un yaklaşımıdır(6). Ancak bu yaklaşım birinci yaklaşıma göre daha pahalı olabilir. Regresyonda verinin analizinden elde edilecek bulgulara (çıkarsamalara) her bir gözlemin etkisini saptamak amacıyla, tek bir gözlemin veriden çıkartılmasına dayalı olan bir çok yöntem vardır. Birden fazla gözlemin veriden çıkartılması ile ilgili yöntemler de çok olmasına rağmen, bu yöntemler veriden çıkartılacak olan gözlemlerin belirlenmesindeki zorluk nedeniyle çok az kullanılırlar. Örneğin 30 gözlemden oluşan bir veride 4 gözlemin çıkartılması düşünülürse olası durumların sayısı $\binom{30}{4} = 27405$ kadardır. Bazı örnekler tek bir gözlemin çıkartılmasına götürebilir. Diğer yandan, gözlemlerin birkaçı veriden çıkartılmadıkça bu gözlemlerin etkileri anlaşılamaz. Bu etki gizleme(maskeleme)[masking] olarak bilinir. Eğer, gizleme(maskeleme) varsa tek bir gözlemin veriden çıkartılması ile ilgili belirleme (teshiş etme) yöntemleri, etkili gözlemler ve uç değerlerin açığa çıkartılmasında başarılı olamaz(6).

Atkinson(1986), çalışmasında, gizleme mevcut iken etkili gözlemler ve uç değerlerin incelenmesi için iki evreli(aşamalı) bir yöntem tanımlar(6).

Araştırmacı(exploratory) nitelikte olan birinci aşama kareler regresyonunun en küçük medyanını kullanır. Bu aşama verideki bozulmanın yaklaşık %50 sini ayıklar. İkinci aşamada en küçük kareler regresyonunun belirleme (teshis etme) yöntemleri sağlam(robust) yöntemin bulgularını doğrulamada kullanılır.

Kareler regresyonunun en küçük medyanı için kullanılan sayısal yöntem, bir rassal araştırma formu kullanır. Regresyon modelinin sırası p ise, mevcut n sayıdaki gözlemden alınan p durumlu örnekler dikkate alınır. Bu regresyon modellerinin her biri için işlemler yapılır. Uç değerlerin doğrudan incelenmesi için bu yöntemi kullanan Hawkins, Bradu ve Kass(1984) bu örnekleri elemanların setleri olarak adlandırırlar(4). Örneklemme uç değerlerin yapısı ortaya çıkana kadar tekrar edilir. Uç değerlere oldukça karşı koyar nitelikte olsa da, kareler regresyonunun en küçük medyanının sonuçları zayıf özelliklere sahiptir. Bu yöntem, uç değerlerin varlığında sağlam yöntemin bunlara direnmesi(karşı koyması) ile uç değerlerin belirlenmesinde en küçük karelerin etkinliğinin birarada kullanılmasını önerir(6). Aşağıda araştırmacı analiz ve doğrulayıcı analiz ile ilgili bilgi verilecektir.

Araştırmacı analiz(birinci aşama); bağımsız değişkenlerin fonksiyonlarının p sayıda taşıyıcılarının $(x_1, x_1^2, x_2, x_2^2, \dots)$ gibi bilinen değerlerin $n \times p$ boyutlu matrisi X olmak üzere, gözlemlerin çoğunluğunun

$$E(Y) = X\beta \quad (1.12)$$

standart lineer regresyon modeline uygun olduğunu varsayalım. Bununla beraber, gözlemlerin bir kısmının bu modele uygunluk göstermediği şüphesi mevcuttur. Bu tür gözlemlerin oranı oldukça fazla olabilir, ancak bunların oranının 1/2 den küçük olması gerekir. Kareler regresyonunun en küçük medyanının amacı sadece “iyi gözlemlere” bir doğru uydurmaktır. Böylece “kötü gözlemler” uç değer olarak açığa çıkartılır. Rousseeuw(1984), kareler regresyonunun en küçük medyanının özelliklerini tanımlamış ve bir bağımsız değişkenli regresyon için bir algoritma vermiştir(19). $p > 2$ durumları için rassal araştırma (random search), alt setlerde kullanılır. J ile gösterilen böyle bir alt set düşünelim. Böylece X_J , X 'in p farklı satırını kapsar. X_J 'ye karşı gelen yanıtların(Y 'nin) p sayıda değerleri Y_J olmak üzere, p noktaya göre uygun regresyon denklemi,

$$\hat{\beta}_J = X_J^{-1}Y_J \quad (1.13)$$

katsayılarına sahiptir. Eğer, $i \in J$ ise,

$$r_{ij} = y_i - x_i' \hat{\beta}_J \quad (1.14)$$

artıklarının toplamı ve beklenen değeri sıfırdır. Diğer durumda

$$d'_{ij} = x_i' X_J^{-1} \quad (1.15)$$

olmak üzere,

$$\text{Var}(r_{ij}) = \sigma^2 (1 + d'_{ij} d'_{ij}) \quad (1.16)$$

dir.

En küçük kareler regresyonda, denklemler artık kareler toplamını minumum, dolaylı olarak artıkların karelerinin ortalama değerini minumum yapacak şekilde tahmin edilir. Basit bir örnekle, ortalamanın alternatifi

medyandır. O halde, aynı amaca ulaşmak için, sıfır olmayan artıkların karelerinin medyanını kullanabiliriz. Doğaldır ki, artıkların mutlak değerlerinin herhangi bir pozitif gücü aynı sonuca götürecektir. Genel olarak n-p sayıda sıfır olmayan artık olacaktır. Fakat bazı durumlarda, örneğin gözlem sonuçlarının en yakın tamsayıya yuvarlandığı durumlarda olduğu gibi, J'nin bazı seçimleri sıfır olmaması gerektiği halde sıfır değerli olan artıklara götürür. Belirli J için bu tür artıklar Z_J kadar olsun. Buna göre T alt seti, J alt seti için sıfır olmayan artıkların n-p- Z_J seti N_J olmak üzere,

$$\text{Med}(r_{iT}^2) = \text{Min}_J \text{Med}_{N_J} r_{iJ}^2 \quad (1.17)$$

ile araştırılır.

Z'nin yaklaşık çözümü, J gözlemden oluşan örneklerin tekrar çekimleri ile bulunur. Bu işlemde her bir örnek yerine koymadan yöntemi ile seçilecektir. Eğer, bozulmanın belirli bir düzeyinin varlığında sadece iyi gözlemleri kapsayan en az bir regresyonun belirli bir olasılığına sahip olması gerekiyorsa, örneklerin gereken sayısı hesaplanabilir. Bununla beraber, az sayıda sayısal örneklerin sonucu olarak, büyük artıkların bir durağan hali ortaya çıkana kadar örnekleme devam edilir.

Bu yöntemin amacı, analizin ikinci aşaması için uç değerlerin belirlenmesini sağlamak ise de, varyansın (σ^2 'nin) yaklaşık tahmini artık karelerin en küçük medyanının standardizasyonunu sağlamasını gerektirir.

Atkinson(1986)'nın ikinci yaklaşımı doğrulayıcı(confirmatory) analizdir. Regresyon karelerin en küçük medyanı uç değerlere karşı iyi bir belirleyici ise de, doğrusal modelin parametrelerinin etkin tahminlerini sağlamazlar. Analizin doğrulayıcı(ikinci) aşamasında en küçük kareler regresyonunun etkin belirleme(teşhis etme) yöntemleri kullanılır. Bu yöntemler, Atkinson(1985), Cook ve Weisberg(1982) ve Belsley, Kuh ve Welsch(1980)'in kitaplarında ayrıntılı olarak açıklanmıştır. İki amacın biri kötü gözlemlerin belirlenmesi diğeri de bunların çıkartılmasının veriden elde edilecek kararı anlamlı şekilde değiştirip değiştirmeyeceğinin belirlenmesidir.

Şimdi I seti ile gösterilen m gözlemin veriden çıkartılacağına karar verildiğini kabul edelim. Böylece n-m gözlem kalır.

$$H = X(X'X)^{-1}X' \quad (1.18)$$

köşegen elemanları h_i ile gösterilen ve geri kalan gözlemler için 'hat' matrisi olarak adlandırılan $(n-m)(n-m)$ boyutlu matris olsun. σ^2 'nin artık ortalama kareler tahmini S^2 olsun. Y_i uygun gözlem bulunmadığı zaman, Y_i 'nin X_i 'deki tahmini olan $Y_{i\cdot}$ ile Y_i gözlem değeri arasındaki farkın(artığın) önemliliği için yapılacak t testi

$$r_i^* = \frac{r_i}{S_{(i)}(1-h_i)^{1/2}} \quad (1.19)$$

ile verilen r_i^* çıkarma artığını kullanır. (1.19)'da, r_i bilinen en küçük kareler artığı, yani $r_{ij} = y_i - x_i' \beta_j$ 'dir. Ayrıca,

$$(n - p - m - 1)S_{(i)}^2 = (n - p - m)S^2 - \frac{r_i^2}{(1 - h_i)} \quad (1.20)$$

de veriden çıkartılacak i gözlemi ile ilgili σ^2 'nin tahminini verir(6). Veriden elde edilecek bulgulara i . gözlemin etkisini saptamak için, Cook istatistiği kullanılır(20). Bu istatistik, β için güven bölgesine i . gözlemin çıkartılması sonucu $\beta - \beta_{(i)}$ değişmesini kullanır. İşaretleme amaçları için, Atkinson tarafından önerilen düzeltilmiş Cook istatistiği olan

$$C_i = \left(\frac{(n - p - m)}{p\{h_i / (1 - h_i)\}} \right)^{1/2} |r_i^*| \quad (1.21)$$

kullanılır. (1.19) ve (1.21)'deki artıklar ve hat matris elemanları, m gözlemi çıkarılmış model içindir.

Uygun modelde kapsanmayan x_i noktasındaki tahminin varyansı

$$\begin{aligned} \text{Var}(\hat{Y}_i) &= \sigma^2 (1 + x_i' (X'X)^{-1} x_i) \\ \text{Var}(\hat{Y}_i) &= \sigma^2 (1 + d_i) \end{aligned} \quad (1.22)$$

dir. Y_i ile $\hat{Y}_i$ 'in farkının önemliliği için t testi

$$r_i^o = \frac{r_i}{S(1 + d_i)} \quad (1.23)$$

istatistiğini kullanır.

Parametre tahminlerinde bir gözlemin çıkarılmasının etkisi ile eklenmesinin etkisi arasındaki ilişki aşağıda açıklanmıştır. Eğer $A = X'X$ olduğunu kabul edersek, bir gözlemin çıkarılması aşağıdaki

$$(A - xx')^{-1} = A^{-1} + A^{-1}x(1 - x'A^{-1}x)^{-1}x'A^{-1} \quad (1.24)$$

ve

$$\hat{\beta}_{(i)} - \hat{\beta} = \frac{-(X'X)^{-1}x_i r_i}{(1 - h_i)} \quad (1.25)$$

formüllerdeki gibi tanımlanmış olur. Bir i. gözlemin eklenmesi ise

$$(A + xx')^{-1} = A^{-1} - A^{-1}x(1 + x'A^{-1}x)^{-1}x'A^{-1} \quad (1.26)$$

ve

$$\hat{\beta}_{(i)}^o - \hat{\beta} = \frac{(X'X)^{-1}x_i r_i}{(1 + d_i)} \quad (1.27)$$

formülleri ile tanımlanır. Cook istatistiğine geçiş aşağıdaki gibidir,

$$D_i = \frac{(\hat{\beta}_{(i)}^o - \hat{\beta})' X' X (\hat{\beta}_{(i)}^o - \hat{\beta})}{pS^2}$$

$$D_i = \frac{r_i^2 d_i^2}{pS^2 (1 + d_i)^2} \quad (1.28)$$

ve değiştirilmiş Cook istatistiği,

$$C_i^* = \left\{ \left(\frac{n - p - m}{p} \right) \left(\frac{d_i}{1 + d_i} \right) \right\}^{1/2} |r_i^o| \quad (1.29)$$

gibi tanımlanır(6). Parametre tahmini, m etkili gözlem çıkarılırsa n-m gözlem üzerinden yapılır. (1.19) ve (1.25) Atkinson'un etkili gözlemler üzerine yaptığı çalışmaların sonucudur.

Üçüncü budama kuralı, iyi regresyon modelinin sağlam(robust) ölçüsünü gerektirir. Artık kareler ortalaması kriteri buna bir örnektir. Eğer, (y_i, x_i') 'ler benzer, bağımsız ve normal dağılmışlarsa, bu durumda

$$E(A.K.O.) = \sigma^2 \quad (1.30)$$

ve

$$\text{Var}(A.K.O.) = \frac{2\sigma^4}{n - k} \quad (1.31)$$

dır.

A.K.O._j, j Bootstrap Örneği için Artık Kareler Ortalamasını gösterebilir ve Bootstrap Yığınının tüm gözlemleri kullanarak hesaplanan A.K.O.'yu da $A.K.O. = \sigma^2_{BY}$ ile gösterelim. Bu durumda,

$$A.K.O._j \leq \left\{1 - \left(\frac{2}{n_b - k}\right)^{1/2}\right\} A.K.O._{B.Y.} \quad (1.32)$$

ise j. Bootstrap Örneği iyi (bozulmamış) olarak sınıflandırılacaktır.

Bu budama stratejisi, $(y_i, x'_i) \in B.\ddot{O}.$ için iyi regresyon modeli üreten örneklerin sayısı, $(y_i, x'_i) \in B.\ddot{O}.$ için mümkün örneklerin sayısına oranının Bootstrap tahminine dayanır. Bu oranı $\hat{P}_g(i)$ ile gösterelim. $B.\ddot{O}._j$, j. Bootstrap Örneği olmak üzere, $(y_i, x'_i) \in B.\ddot{O}._j$ ise $n_s(i,j) = 1$ ve aksi halde $n_s(i,j) = 0$ tanımını verelim. Benzer şekilde, $(y_i, x'_i) \in B.\ddot{O}._j$ ve $B.\ddot{O}._j$ iyi bir regresyon modeli veriyorsa $n_g(i,j) = 1$ ve aksi halde $n_g(i,j) = 0$ olsun. $P_g(i)$ 'nin Bootstrap tahmini

$$\hat{P}_g(i) = \frac{\sum_{j=1}^B n_g(i,j)}{\sum_{j=1}^B n_s(i,j)} \quad (1.33)$$

ile elde edilir. Eğer bu koşullu olasılık yeteri kadar büyükse, o zaman i. gözlem önemli (değersiz olmayan) olarak sınıflandırılır ve bu durumda Budanmış Bootstrap Yığına alınır.

(y_i, x'_i) 'yi Budanmış Bootstrap Yığına dahil etmek için $P_g(i)$ ne kadar büyük olmalıdır? Bir çok simülasyon çalışmasının bulunabileceği iki kural aşağıda verilmiştir.

(I) Eğer $\hat{P}_g(i); \hat{P}_g(1), \dots, \hat{P}_g(n)$ için 100 δ 'ıncı yüzdelik değerinden büyük ise, (y_i, x'_i) 'yi Budanmış Bootstrap Yığınınına dahil edilir. Bu değer yani 100 δ 'ıncı yüzdelik değeri, $P(\delta)$ ile gösterilir. $0 < \delta < 1$ olan δ değeri, veride gözlenen bozulmanın düzeyine göre seçilir(1).

(II) Tek tek tüm i 'ler için,

$$H_0 : P_g(i) = P(\delta) \quad (1.34)$$

$$H_1 : P_g(i) < P(\delta) \quad (1.35)$$

hipotezleri tanımlansın. Yukarıda H_0 'ın red edildiği durum ile ilgili olan her bir gözlem Bootstrap Yığınınından budanmalıdır. Bir çok pratik problemde, sadece hafif(az) budamaya ihtiyaç duyulur. Fazla budama gerektiren bir çok simülasyon örneği çalışılmıştır. Bu çalışmalar, serpmeye diyağramının uzağındaki gözlemlerin sayısının uygun (gerçekte uzak olmayan) gözlemlerin sayısından küçük olduğunda birinci ve ikinci budama kurallarının iyi çalıştığını göstermiştir. Üçüncü budama kuralı, serpmeye diyağramının uzağındaki gözlemlerin sayısının uzakta olmayan gözlemlerin sayısından büyük olduğu durumda bile iyi sonuçlar vermiştir. Zorluk δ 'yı seçerken ortaya çıkar. Bir çok Bootstrap Örneğinde bozulma seviyesi ölçülebilirse, o zaman δ 'nın seçimi yapılabilir, aksi durumda büyük bir δ seçilir($\delta = 0.80$)(1).

1.3. Bootstrap Tahminleri

Budanmış Bootstrap Yığını(BBY), yığından seçilen gözlemlerden uç değer olduğuna karar verilenler atıldıktan sonra geriye kalan gözlemlerin kümesi olarak düşünülebilir. Örnek çapları n_b olan Bootstrap Örnekleri(BÖ), Budanmış Bootstrap Yığınınından çekilir. Bootstrap Örneğindeki (y_i, x'_i) 'lerin tekrar sayısı w_i olsun. Ayrıca köşegen(diyagonal) üzerindeki i . elemanı w_i olan köşegen matris D_w ile gösterilsin. Bilinen Ağırlıklı En Küçük Kareleri kullanarak doğrusal regresyon modelinin katsayılarının Bootstrap Örneği kullanılarak bulunan tahminleri,

$$\hat{\beta}_{BO} = (X'D_w X)^{-1} X'D_w Y \quad (2.1)$$

ile elde edilir. Eğer $w_i = 0$ ise $i=1, \dots, n$ için o zaman tahmin edilen $\hat{e}_{pi}$ artığı hesaplanır. B Bootstrap Örneklerinin Budanmış Bootstrap Yığınınından rasgele çekildiğini varsayalım. B, Bootstrap Örneklerinin sayısını göstermektedir.

Dağılım parametresi τ_i^2 ve konum parametresi λ_i olan $F_i(\lambda_i, \tau_i^2)$ 'nin dağılımı, β için sapmasız tahmin kullanarak hesaplanan $\hat{e}_{pi}$ 'nin dağılımı olarak tanımlanır. Konum parametresi olan λ_i , serpmeye diyagramının uzağındaki gözlemleri gösterirken, dağılım parametresi olan τ_i^2 aşağıdaki yorumlarda kullanılabilir.

(I) X değişkenleri saptanmış olduğunda, i . nci gözlemin etkisinin yorumlanmasında kullanılabilir. ALT, Bootstrap Örneğine eşit olmak üzere, X'lerin saptanmış olduğu regresyon modellerinde,

$$\tau_i^2 = \sigma_{BY}^2 (1 + hx_{ALT}(i, j)) \quad (2.2)$$

dir. $\hat{e}_{pi}$ 'nin Bootstrap dağılımı, doğrudan $hx_{ALT}(i, i)$ ile orantılı olarak saçılır(21). Ayrıca bu dağılım Budanmış Bootstrap Yığınının tasarım uzayı bakımından (y_i, x'_i) 'nin etkisini gösterir. X'lerin saptanmış olmadığı regresyonda (x'lerin rasgele değişken olduğu durum) koşullu tasarım uzayı bakımından benzer yorumlar vardır.

(II) X'lerin saptanmış olmadığı regresyon modelleri durumunda değişen varyans(heteroscastic) artık yapısının yorumlanmasında kullanılabilir. X'lerin saptanmış olduğu durumda

$$\frac{\hat{e}_{pi}}{(1 + hx_{ALT}(i, j))^{1/2}} \quad (2.3)$$

Bootstrap dağılımının yayılması(saçılması) değişen varyans durumunu göstermede kullanılabilir.

$F_i(\lambda_i, \tau_i^2)$ 'nin Bootstrap tahmini, $\hat{F}_i(\lambda_i, \tau_i^2)$, $i = 1, \dots, n$ ile gösterilir. $\hat{e}_{pi}$ 'nin B Bootstrap Örneklerinin maksimum tahmininin deneysel dağılımı $F_i(\lambda_i, \tau_i^2)$ 'nin Bootstrap tahmini olarak alınacaktır. Eğer Budanmış Bootstrap Yığını sadece serpmeye diyağramının uzağında olmayan gözlemleri içeriyorsa, $\hat{F}_i(\lambda_i, \tau_i^2)$ 'nin $F_i(\lambda_i, \tau_i^2)$ için uygun bir tahmin olduğu kabul edilebilir. $\hat{F}_i(\lambda_i, \tau_i^2)$ 'nin konum parametresi μ_i olmak üzere, alternatif hipotezin doğruluğu altında, yani $H_1: \lambda_i = \mu_i$ altında; $\hat{F}_i(\lambda_i, \tau_i^2)$ 'yi $\hat{e}_{pi}$ 'nin tahmin edilen dağılımı olarak kabul edilebilir. $H_0: \lambda_i = 0$ iken $F_i(\lambda_i, \tau_i^2)$ 'nin tahmini, $F_i(0, \tau_i^2)$ ile gösterilen $\hat{e}_{pi} - \mu_i$ 'nin Bootstrap dağılımıdır. X'lerin

saptanmış olduğu regresyonda $H_0: \lambda_i=0$ hipotezini test ederken, bu dağılım standart tahmin edilmiş artığın, yani (2.3)'ün Bootstrap dağılımına daha uygun olabilir. $\hat{e}_{pi}$ 'nin Bootstrap dağılımının $\alpha 100$ yüzdelerlik değeri $l_i(\alpha)$ ile gösterilsin. $F_i(\lambda_i, \tau_i^2)$ kullanılarak λ_i için, $i=1, \dots, n$, $(1-\alpha)100\%$ güven aralığı

$$l_i(\alpha/2) < \lambda_i < l_i(1-\alpha/2) \quad (2.4)$$

dir(23). Eğer sıfır değeri bu aralığa düşmez ise, o zaman (y_i, x_i) serpmeye diyagramının uzağındaki gözlem olarak sınıflandırılabilir. $F_i(0, \tau_i^2)$ ve $F_i(\lambda_i, \tau_i^2)$ kullanılarak $H_0: \lambda_i=0$ test edilirse, yukarıdaki güven aralığının daima aynı olmayacağı nedeniyle bu karar oluşmuştur.

$$F_i(\lambda_i, \tau_i^2) = \begin{cases} \hat{F}_i(0, \tau_i^2), H_0 \text{ altında} \\ \hat{F}_i(\lambda_i, \tau_i^2), H_1 \text{ altında} \end{cases} \quad (2.5)$$

verildiğinde,

$$H_1: \lambda_i = \mu_i \text{ hipotezine karşı } H_0: \lambda_i = 0 \quad (2.6)$$

hipotezinin testi sonraki bölümde açıklanacaktır.

Bağımsız değişkenler rasgele değişken iken

$$sdp = \sigma_{BY} \sqrt{1 + \frac{1}{n_b} + \frac{k(n_b + 1)}{n_b(n_b - k - 2)}} \quad (2.7)$$

ve bağımsız değişkenler saptanmış iken

$$sdp = \sigma_{BY} \sqrt{(1 + h_{X_{ALT}}(i, i))} \quad (2.8)$$

olmak üzere, serpmeye diyagramının uzağındaki (y_i, x_i) gözlemler için bir kriter

$$P(\hat{e}_{pi} > 2sdp/H_1) \text{ veya } P(\hat{e}_{pi} < -2sdp/H_1)$$

aracılığıyla ölçülebilir. Eğer bu olasılıkların biri diğerine yakın ise, o zaman (y_i, x'_i) , serpmeye diyaqramından aşırı uzak gözlem olarak sınıflandırılır.

$P(\hat{e}_{pi} > c/H_1)$ 'in tahminleri aşağıdaki gibi bulunabilir. B Bootstrap Örnekleri içindeki i . gözlem için hesaplanan tahmin edilmiş $\hat{e}_{pi}$ artığının sayısı $n_s(i)$ olsun

$$I_c(i, j) = \begin{cases} \hat{e}_{pi} \text{ artığı } j. \text{ B. Ö. } i \text{ için} \\ 1, \text{ hesaplandı ve } \hat{e}_{pi} > c_i \\ 0, \text{ diğer durumlar} \end{cases} \quad (2.9)$$

yi tanımlayalım.

Ayrıca $n_c(i) = \sum_{j=1}^n I_c(i, j)$ olsun. Bunlara göre $P(\hat{e}_{pi} > c/H_1)$ 'nin tahmini

$$P(\hat{e}_{pi} > c/H_1) = \frac{n_c(i)}{n_s(i)} \quad (2.10)$$

dir.

$$\hat{P}(\hat{e}_{pi} > c/H_0) = \hat{P}(\hat{e}_{pi} > c + \mu_i / H_1) \quad (2.11)$$

olduğu açıktır.

1.4. Şartlı Hipotez Testi

(2.5)'deki şart, birinci tip ve ikinci tip hataların her ikisinin de kontrol edebilen bir test ortaya koymaktadır. (2.6)'daki hipotezin testi için α anlamlılık düzeyinde en iyi kritik bölge, $\hat{e}_{pi} > cr$, gücü maksimum yapar. Diğer bir ifade ile $\hat{P}(\hat{e}_{pi} > cr/H_0) = \alpha$ ve $\hat{P}(\hat{e}_{pi} > cr/H_1)$ maksimum olacak

şekilde seçilirse, o zaman, $\hat{e}_{pi} > cr$ en iyi kritik bölge olarak tanımlanır(22).
 $cr \geq \sigma_{BY}$ ve $\hat{P}(\hat{e}_{pi} > cr \geq 0.60)$ koşullarını sağlayan bazı $\alpha \leq 0.10$ için, en iyi kritik bölge varsa, i gözlemi uzaktaki gözlem olarak sınıflandırılır.

Eğer tahmin edici bağımsız değişkenler saptanmış ve yanılığalar homojen ise, o zaman

$$H_0: \lambda_i = 0, i=1, \dots, n \quad (3.1)$$

için koşullu hipotez testi oluşturmada $\hat{e}_{pi} / (1 + h_{X_{ALT}}(i, i))^{1/2}$ 'nin dağılımı $F_i(\lambda_i, \tau_i^2)$ 'dir.

1.5. Bozulmanın Ölçülmesi

Bozulma, Artık Kareler Ortalaması(A.K.O.) ve belirleme katsayısı(R^2) istatistikleri kullanılarak ölçülür. Eğer, uç değerler Budanmış Bootstrap Yığını içinde varsa, bu durumda bu Budanmış Bootstrap Yığınının rasgele olarak seçilecek küçük çaplı Bootstrap Örnekleri(örneğin $n_b \ll n$), genellikle uç noktaları kapsamazlar. Bundan dolayı, bu istatistiklerin olasılık yoğunluk fonksiyonlarının Bootstrap tahmini genel olarak çoklu(multi)-model olacaktır. Bu, Budanmış Bootstrap Yığınının seçilen B Bootstrap Örneklerinin kullanımıyla histogram oluşturarak gözlenebilir.

Burada yukarıda tanımlanan yaklaşıma bir alternatif verilecektir. R^2_{BBY} ve $A.K.O._{BBY}$ ile Budanmış Bootstrap Yığınındaki tüm gözlemler kullanılarak hesaplanan R^2 ve A.K.O.'yu gösterelim. $\bar{R}^2_B$ ve $\overline{A.K.O.}_B$ ile bu

istatistiklerin B Bootstrap Örnekleri değerlerinin ortalaması alınarak bulunan A.K.O. ve R^2 'nin Bootstrap tahminlerini gösterelim. Sparks(1988)'ın önerdiği bozulmanın ölçüsü, her bir Budanmış Bootstrap Yığını için

$$|R_{BBY}^2 - \bar{R}_B^2| \quad (4.1)$$

ve

$$\frac{|A.K.O_{BBY} - A.K.O_B|}{A.K.O_{BY}} \quad (4.2)$$

dir. Eğer bunların her ikisinde sıfıra çok yakın ise, o zaman Budanmış Bootstrap Yığını, için bozulmanın olmadığı söylenebilir. Eğer, veri bozulmamış ise, bu ölçülerin ikisinde sıfıra çok yakın olur. Diğer yandan verideki herhangi bir bozulma yukarıdaki istatistiklerin değerlerini genel olarak arttıracaktır.

BÖLÜM 2

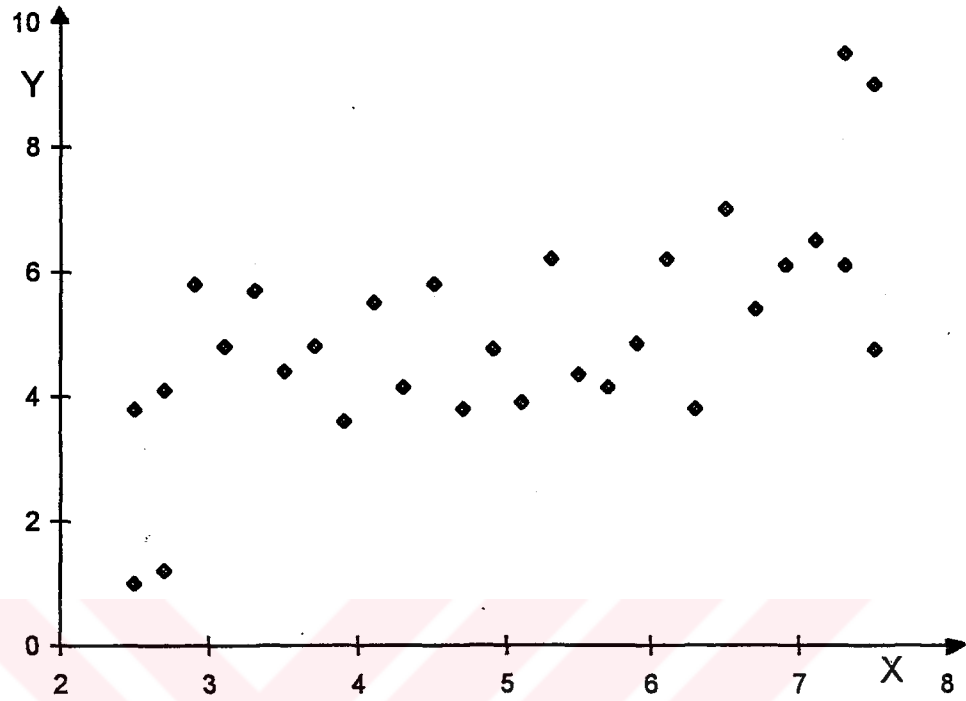
UYGULAMA

Regresyon analizinde uç değerlerin önemi büyüktür. Bu bölümde veri uç değerler içeriyorken tek bağımsız değişkenli regresyon modeli için bir uygulamaya yer verilecektir. Bu çalışmada uç değer incelemede Bootstrap Yığınının budanması temel alınacaktır. Uygulamada uç değer belirleme ve uç değer inceleme için kullanılacak olan veri bir özel televizyon kanalına aittir(24). Televizyon kanalının yöneticisi, haber ratingini etkileyen faktörler üzerine bir araştırma yaptırıyor. Araştırmada kısaltılmış(özet) haber gösterim ratingleri ile ilgilenilmektedir. Ayrıca haberin hazırlanışı, sunuluşu, genel düzeni ve haberlerden hemen önce yer alan özet haber gösteriminden bir etkinin kaldığı düşünülmektedir. Haber gösterim ratingi, haberlerden hemen önceki gösterimin ratingine kısmen bağlıdır. Y , haber gösterim ratingi ve X de haberlerden hemen önceki özet haber gösterim ratingidir. Özet haber gösterim ratinginin, haber gösterim ratingini etkilemesinin ölçülmesi amacıyla geçmiş iki yıl için rating ölçme yerlerinden 30 yer rasgele seçilmiştir. Rasgele seçilen bu 30 yerde haber öncesi özet haberler ve haberler için ratingler ölçülmüştür. Bu ratingler tablo 2’de verilmiştir.

Tablo 2. Uygulama için Veri Tablosu

Sıra No	X_i	Y_i
1	2.50	1.00
2	2.50	3.80
3	2.70	1.20
4	2.70	4.10
5	2.90	5.80
6	3.10	4.80
7	3.30	5.70
8	3.50	4.40
9	3.70	4.80
10	3.90	3.60
11	4.10	5.50
12	4.30	4.15
13	4.50	5.80
14	4.70	3.80
15	4.90	4.75
16	5.10	3.90
17	5.30	6.20
18	5.50	4.35
19	5.70	4.15
20	5.90	4.85
21	6.10	6.20
22	6.30	3.80
23	6.50	7.00
24	6.70	5.40
25	6.90	6.10
26	7.10	6.50
27	7.30	6.10
28	7.30	9.50
29	7.50	4.75
30	7.50	9.00
Toplam	150.00	151.00

Bu veriye ilişkin serpmeye diyagramı aşağıdadır.



Şekil 8. Veri için Serpmeye Diyagramı

Tablo 2'deki veriden doğrusal regresyon modeli için bazı regresyon analizi sonuçları aşağıdaki tabloda verilmiştir. Burada k tahmin edilen parametre sayısıdır.

Tablo 3. Veri için Regresyon Analizi Tablosu

n	k	$\hat{\beta}_0$	$\hat{\beta}_1$	s_e	$\sqrt{V(\hat{\beta}_1)}$	R^2
30	2	1.7065	0.6654	1.4019	0.1552	0.3963

Tablo 4. Tablo 2'deki Verinin Budanmasına İlişkin İşlem Sonuçları

Sıra No	X_i	Y_i	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$\frac{(X_i - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}$	$\frac{(Y_i - \bar{Y})^2}{\Sigma(Y_i - \bar{Y})^2}$
1	2.50	1.00	6.2500	16.2409	0.0766	0.1782
2	2.50	3.80	6.2500	1.5129	0.0766	0.0166
3	2.70	1.20	5.2900	14.6689	0.0648	0.1609
4	2.70	4.10	5.2900	0.8649	0.0648	0.0095
5	2.90	5.80	4.4100	0.5929	0.0541	0.0065
6	3.10	4.80	3.6100	0.0529	0.0443	0.0006
7	3.30	5.70	2.8900	0.4489	0.0354	0.0049
8	3.50	4.40	2.2500	0.3969	0.0276	0.0044
9	3.70	4.80	1.6900	0.0529	0.0207	0.0006
10	3.90	3.60	1.2100	2.0449	0.0148	0.0224
11	4.10	5.50	0.8100	0.2209	0.0099	0.0024
12	4.30	4.15	0.4900	0.7744	0.0060	0.0085
13	4.50	5.80	0.2500	0.5929	0.0031	0.0065
14	4.70	3.80	0.0900	1.5129	0.0011	0.0166
15	4.90	4.75	0.0100	0.0784	0.0001	0.0009
16	5.10	3.90	0.0100	1.2769	0.0001	0.0140
17	5.30	6.20	0.0900	1.3689	0.0011	0.0150
18	5.50	4.35	0.2500	0.4624	0.0031	0.0051
19	5.70	4.15	0.4900	0.7744	0.0060	0.0085
20	5.90	4.85	0.8100	0.0324	0.0099	0.0004
21	6.10	6.20	1.2100	1.3689	0.0148	0.0150
22	6.30	3.80	1.6900	1.5129	0.0207	0.0166
23	6.50	7.00	2.2500	3.8809	0.0276	0.0426
24	6.70	5.40	2.8900	0.1369	0.0354	0.0015
25	6.90	6.10	3.6100	1.1449	0.0443	0.0126
26	7.10	6.50	4.4100	2.1609	0.0541	0.0237
27	7.30	6.10	5.2900	1.1449	0.0648	0.0126
28	7.30	9.50	5.2900	19.9809	0.0648	0.2192
29	7.50	4.75	6.2500	0.0784	0.0766	0.0009
30	7.50	9.00	6.2500	15.7609	0.0766	0.1729
Toplam	150.00	151.00	81.5800	91.1420	1.0000	1.0000

Şekil 8'deki serpm diyagramı incelendiğinde bazı gözlemlerin diğer gözlemlerden farklı konumlarda olduğu görülmektedir. Bu farklı gözlemler 1, 3, 28 ve 30 nolu gözlemlerdir. Diğer 26 gözlem 4 gözleme göre daha homojen bir yapıya sahip gibi gözükmemektedir. Bu 26 gözlem içerisinde bazı gözlemler diğer gözlemlerinde etkilerinde kalmaktadır. Etkilenen bu gözlemler ancak test edildikten sonra homojenliği bozup bozmadığına karar verilebilecektir. Regresyon modelinin doğru tahminlere götürmesi için gözlemlerin yoğun olduğu bölgenin uzağında ya gözlem olmamalı yada uzak olan gözlemler iyi incelenmelidir.

Bu çalışmada özel televizyon kanalının verisinin modele uygunluğunun tespiti için birinci bölümde önerilen budama kriterlerinden faydalanılacaktır.

Tablo 2'deki özel televizyon kanalına ait olan veri Bootstrap Yığını olarak alınacaktır. Budama işlemi gerçekleştirildikten sonraki yığın ise Budanmış Bootstrap Yığını olarak adlandırılacaktır. Bootstrap Yığınının budanması ile ilgili işlemler aşağıda tablo 5'te verilmiştir.

Bootstrap yığını kullanılarak, yani $n=30$ gözlem kullanılarak elde edilen

$$\hat{Y}_i = 1.7065 + 0.6654X_i$$

için artıklar, artık kareler, v_{ii} , r_i , ve t_i^2 'ler aşağıda verilmiştir.

Tablo 5. n=30, k=2 iken Artıklar, Artık Kareler, v_{ii} , r_i ve t_i^2 'lere ilişkin işlem Sonuçları

Sıra No	$\hat{Y}_i$	$(Y_i - \hat{Y}_i)$	$(Y_i - \hat{Y}_i)^2$	v_{ii}	r_i	t_i^2
1	3.3698	-2.3698	5.61572	0.1099	-1.7901	3.4892
2	3.3698	0.4303	0.18512	0.1099	0.3250	0.1022
3	3.5028	-2.3028	5.30293	0.0982	-1.7298	3.2305
4	3.5028	0.5972	0.35664	0.0982	0.4486	0.1954
5	3.6359	2.1641	4.68346	0.0874	1.6160	2.7771
6	3.7689	1.0311	1.06311	0.0776	0.7658	0.5776
7	3.9020	1.7980	3.23284	0.0688	1.3291	1.8181
8	4.0351	0.3650	0.13319	0.0609	0.2686	0.0698
9	4.1681	0.6319	0.39928	0.0540	0.4634	0.2087
10	4.3012	-0.7012	0.49164	0.0482	-0.5127	0.2558
11	4.4342	1.0658	1.13587	0.0433	0.7773	0.5954
12	4.5673	-0.4173	0.17413	0.0393	-0.3037	0.0892
13	4.7004	1.0997	1.20923	0.0364	0.7991	0.6301
14	4.8334	-1.0334	1.06794	0.0344	-0.7502	0.5538
15	4.9665	-0.2165	0.04686	0.0335	-0.1571	0.0238
16	5.0995	-1.1995	1.43887	0.0335	-0.8704	0.7508
17	5.2326	0.9674	0.93588	0.0344	0.7023	0.4841
18	5.3657	-1.0157	1.03154	0.0364	-0.7381	0.5357
19	5.4987	-1.3487	1.81902	0.0393	-0.9816	0.9622
20	5.6318	-0.7818	0.61116	0.0433	-0.5701	0.3171
21	5.7648	0.4352	0.18937	0.0482	0.3182	0.0980
22	5.8979	-2.0979	4.40114	0.0540	-1.5387	2.4938
23	6.0310	0.9691	0.93906	0.0609	0.7133	0.4997
24	6.1640	-0.7640	0.58371	0.0688	-0.5648	0.3111
25	6.2971	-0.1971	0.03884	0.0776	-0.1464	0.0207
26	6.4301	0.0699	0.00488	0.0874	0.0522	0.0026
27	6.5632	-0.4632	0.21454	0.0982	-0.3480	0.1173
28	6.5632	2.9368	8.62485	0.0982	2.2065	5.6828
29	6.6963	-1.9463	3.78789	0.1099	-1.4716	2.2633
30	6.6963	2.3038	5.30726	0.1099	1.7419	3.2814
Toplam	150.9900	0.0100	55.0260	2.0000		

$n=30$, $k=2$ ve $\alpha=0.15$ iken Bootstrap yığını oluşturulan gözlemler için X , Y ve t_i^2 budama kriterlerine göre budamaya ilişkin işlemler gerçekleştirilecektir. Bootstrap yığınının budanması için α 'nın değerinin küçük alınması veya α 'nın değerinin büyük alınması durumunda istenilen budamaya ulaşılabilir. Bu çalışmada kullanılan tablo değerleri "Statistical Graphics System" paket programından hesaplanmıştır.

$n=30$, $k=2$ ve $\alpha=0.15$ iken

$$F_{1,28}^{0.15} = 1 / F_{28,1,0.15}$$

$$F_{1,28}^{0.15} = 1/27.4459$$

$$F_{1,28}^{0.15} = 0.0364$$

tablo değeri bulunur. (1.6)'daki tanımdan budama kriterinde kullanılan $C_k(\alpha)$ değeri,

$$C_k(\alpha) = (k-1) F_{k-1, n-k}^{\alpha}$$

dan hesaplanır. Burada $k=2$, $\alpha=0.15$ iken $C_k(\alpha)$ değeri,

$$C_2(0.15) = (2-1) 0.0364$$

$$C_2(0.15) = 0.0364$$

dür. Yukarıdaki kritik değer formüllerini kullanarak X ve Y için budama kritik değerleri hesaplanır. (1.8)'deki tanımdan X bağımsız değişkeninin gözlemleri için budama kriteri,

$$h_x(i,i) > \frac{(n-1)C_k(\alpha)}{n(1-C_k(\alpha))} + \frac{1}{n}$$

dir.

Bu eşitsizliğin sağ tarafının değeri

$$\begin{aligned}
 \frac{(n-1)C_k(\alpha)}{n(1-C_k(\alpha))} + \frac{1}{n} &= \frac{(30-1)C_2(0.15)}{30(1-C_2(0.15))} + \frac{1}{30} \\
 &= \frac{(29)0.0364}{30(1-0.0364)} + \frac{1}{30} \\
 &= \frac{1.0556}{28.908} + \frac{1}{30} \\
 &= 0.0365 + 0.0333 \\
 &= 0.0698
 \end{aligned}$$

hesaplanır. (1.7)'deki tanımdan X gözlemleri için test istatistiği,

$$h_x(i,i) = \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

dir. X gözlem değerlerinin budanabilmesi için budama kriterinin değeri 0.0698'dir. Tablo 4'ün altıncı sütunundaki hesaplanmış $h_x(i,i)$ değerlerinden 0.0698'den büyük olan gözlemler budanır. Büyük olan gözlemler sırasıyla 1, 2, 29 ve 30'dur. Bu gözlemler Bootstrap yığınınından budanır. Diğer bir yaklaşımla X gözlem değerleri için test istatistiği ve budama kritik değer eşitsizliklerinden budama için alt ve üst sınırlar bulunabilir. (1.7) ve (1.8)'deki tanımlar gereği

$$\frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{(n-1)C_k(\alpha)}{n(1-C_k(\alpha))} + \frac{1}{n}$$

yazılabilir. Bu yaklaşımla

$$(x_i - \bar{x})^2 = \left(\frac{(n-1)C_k(\alpha)}{n(1-C_k(\alpha))} + \frac{1}{n} \right) \sum_{i=1}^n (x_i - \bar{x})^2$$

eşitliği kullanılarak alt ve üst sınır değerler bulunabilir.

$$(x_i - \bar{x})^2 = (0.0698) \cdot 8158$$

$$(x_i - \bar{x})^2 = 5.6943$$

$$(x_i - \bar{x}) = \mp 2.3863$$

dür. X gözlem değerlerinin ortalaması 5.00 dir. Üst sınır değeri

$$x_u = 5.00 + 2.3863$$

$$x_u = 7.3863$$

ve alt sınır değeri

$$x_a = 5.00 - 2.3863$$

$$x_a = 2.6137$$

dir.

(2.6137, 7.3863) aralığının dışına düşen bağımsız değişken X'in gözlem değerleri budanır. X budama kriterine göre Bootstrap yığınının 1, 2, 29 ve 30 nolu gözlemler budanır.

(1.10)'daki tanımdan Y bağımlı değişkeninin gözlemleri için budama kriteri,

$$h_y(i,i) > \frac{(n-1)C_k(\alpha)}{n(1-C_k(\alpha))}$$

dır. Bu eşitsizliğin sağ tarafı

$$\begin{aligned} \frac{(n-1)C_k(\alpha)}{n(1-C_k(\alpha))} &= \frac{(30-1)C_2(0.15)}{30(1-C_2(0.15))} \\ &= \frac{(29)0.0364}{30(1-0.0364)} \\ &= \frac{1.0556}{28.9080} \\ &= 0.0365 \end{aligned}$$

hesaplanır. (1.10)'daki tanımdaki Y gözlemleri için test istatistiği,

$$h_y(i,i) = \frac{(y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

dir. Y gözlem değerlerinin budanabilmesi için budama kriterinin değeri 0.0365'dir. Tablo 4'ün yedinci sütunundaki hesaplanmış $h_y(i,i)$ değerlerinden 0.0365'den büyük olan gözlemler budanır. Büyük olan gözlemler sırasıyla 1, 3, 23, 28 ve 30'dur. Bu gözlemler Bootstrap yığınınından budanır. Diğer bir yaklaşımla Y gözlem değerleri için test istatistiği ve budama kritik değer eşitsizliklerinden budama için alt ve üst sınırlar bulunabilir. (1.9) ve (1.10)'daki tanımlar gereği

$$\frac{(y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{(n-1)C_k(\alpha)}{n(1-C_k(\alpha))}$$

yazılabilir. Bu yaklaşımla

$$(y_i - \bar{y})^2 = \frac{(n-1)C_k(\alpha)}{n(1-C_k(\alpha))} \sum_{i=1}^n (y_i - \bar{y})^2$$

eşitliği kullanılarak alt ve üst sınır değerler bulunabilir.

$$(y_i - \bar{y})^2 = (0.0365) \cdot 91.1420$$

$$(y_i - \bar{y})^2 = 3.3267$$

$$(y_i - \bar{y}) = \pm 1.8239$$

dir. Y gözlem değerlerinin ortalaması 5.0333'dür. Üst sınır değeri

$$Y_u = 5.0333 + 1.8239$$

$$Y_u = 6.8572$$

ve alt sınır değeri

$$Y_a = 5.0333 - 1.8239$$

$$Y_a = 3.2094$$

dür.

(3.2094; 6.8572) aralığının dışına düşen bağımlı değişken Y'nin gözlem değerleri budanır. Y budama kriterine göre Bootstrap yığınınından 1,

3, 23, 28, ve 30 nolu gözlemler budanır. X ve Y için budama kriterlerine ilişkin hesaplamalar tablo 4'te ayrıntılı olarak verilmiştir.

Student tipi artıklar için tanım (1.12)'ye göre budama kriteri,

$$t_i^2 > F_{k-1, n-k-1, \alpha}$$

dır. $F_{k-1, n-k-1, \alpha} = F_{1, 27, 0.15} = 2.1954$ 'dür. Test istatistiğinin değeri,

$$r_i = \frac{\hat{e}_i}{\hat{\sigma} \sqrt{(1 - v_{ii})}}, \quad \hat{\sigma} = \sqrt{\frac{(Y_i - \hat{Y}_i)^2}{n - p - 1}} \text{ ve}$$

$$t_i^2 = r_i^2 \left(\frac{n - p' - 1}{n - p' - r_i^2} \right)$$

den hesaplanır. Artıkları en büyük olan gözlemde başlayarak budama kriteri uygulanır. t_i^2 için alt sınır değeri 2.1954'dür. Bu değerden küçük olan t_i^2 değerleri için budama işlemi gerçekleştirilemez. 2.1954'den büyük t_i^2 değerlerine sahip gözlemler budanır. Tablo 5'in yedinci sütunundaki hesaplanmış t_i^2 değerleri 2.1954 ile karşılaştırılarak budanacak gözlemler belirlenir. Bu gözlemler sırasıyla 1, 3, 5, 22, 28, 29 ve 30 nolu gözlemlerdir. Diğer bir yaklaşımla student tipi artıklar yani r_i 'ler için alt ve üst sınır değerleri bulunabilir.

$$t_i^2 = F_{k-1, n-k-1, \alpha}$$

eşitliği kullanılarak r_i 'ler için alt ve üst sınır değerleri bulunabilir.

$$F_{k-1, n-k-1, \alpha} = 2.1954$$

$$2.1954 = r_i^2 \left(\frac{n - p' - 1}{n - p' - r_i^2} \right)$$

$$r_i^2 = 2.1055$$

$$r_i = \pm 1.4510$$

dir. r_i 'ler için alt sınır değeri

$$r_a = -1.4510$$

ve üst sınır değeri,

$$r_u = 1.4510$$

dür.

$(-1.4510, 1.4510)$ aralığının dışında kalan r_i student tipi artıklara sahip gözlemler Bootstrap yığınının budanır. 1, 3, 5, 22, 28, 29 ve 30 nolu gözlemler geriye kalan diğer gözlemlere göre artıkları daha büyüktür. Artığı en büyük olan gözlem 28 nolu gözlemdir. Tablo 5'te t^2_i 'ye ilişkin budama için hesaplamalar ayrıntılı olarak verilmiştir. Tablo 5'in altıncı sütunundaki student tipi artıklar r_i 'ler için hesaplamalar verilmiştir. Bu değerlerden aralığın dışına düşen gözlemler belirlenebilir. Buna göre 1, 3, 5, 22, 28, 29 ve 30 nolu gözlemler budanır.

$n = 30$, $k = 2$ ve $\alpha = 0.15$ anlamlılık düzeyine göre X , Y ve t^2_i budama kriterlerine göre 1, 2, 3, 5, 22, 23, 28, 29 ve 30 nolu gözlemler Bootstrap Yığınının budanır. Bu durumda Budanmış Bootstrap Yığını

{4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,

21, 24, 25, 26, 27}

nolu gözlemlerden oluşur. Bu budamanın yeterli olup olmadığına karar verebilmek için gözlemlerin serpmeye diyagramı üzerindeki dağılışı ve budama için bozulma kriterlerinin simülasyon çalışmasına göre karar verilir.

Tablo 6. n=21 ve k=2 için Budanmış Booststrap Yığınının İlişkin Bazı İşlemlerin Sonuçları

Sıra No	X_i	Y_i	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$\frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2}$	$\frac{(Y_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$
4	2.70	4.10	5.1379	0.8228	0.1366	0.0487
6	3.10	4.80	3.4846	0.0429	0.0926	0.0025
7	3.30	5.70	2.7779	0.4801	0.0738	0.0284
8	3.50	4.40	2.1512	0.3686	0.0572	0.0218
9	3.70	4.80	1.6045	0.0429	0.0426	0.0025
10	3.90	3.60	1.1378	1.9799	0.0302	0.1171
11	4.10	5.50	0.7512	0.2430	0.0200	0.0144
12	4.30	4.15	0.4445	0.7346	0.0118	0.0435
13	4.50	5.80	0.2178	0.6287	0.0058	0.0372
14	4.70	3.80	0.0711	1.4571	0.0019	0.0862
15	4.90	4.75	0.0044	0.0661	0.0001	0.0039
16	5.10	3.90	0.0178	1.2257	0.0005	0.0725
17	5.30	6.20	0.1111	1.4230	0.0030	0.0842
18	5.50	4.35	0.2844	0.4318	0.0076	0.0255
19	5.70	4.15	0.5377	0.7346	0.0143	0.0435
20	5.90	4.85	0.8710	0.0247	0.0232	0.0015
21	6.10	6.20	1.2844	1.4230	0.0341	0.0842
24	6.70	5.40	3.0043	0.1544	0.0798	0.0091
25	6.90	6.10	3.7376	1.1944	0.0993	0.0707
26	7.10	6.50	4.5510	2.2288	0.1210	0.1319
27	7.30	6.10	5.4443	1.1944	0.1447	0.0707
Toplam	104.30	105.15	37.6267	16.9014	1.0000	1.0000

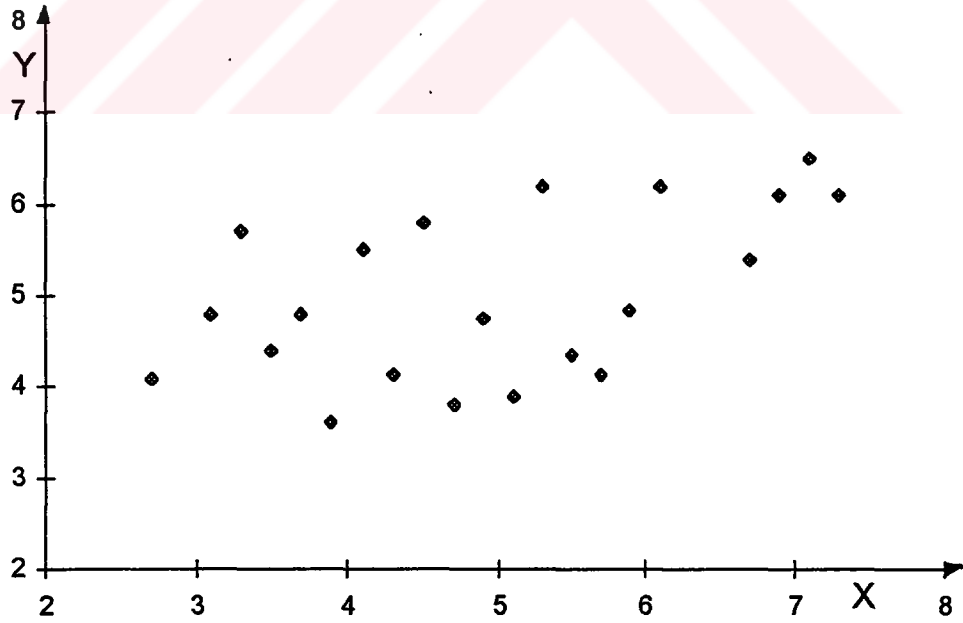
n=21 birimlik Budanmış Bootstrap Yığını için doğrusal regresyon modelinin analiz sonuçlarından bazıları aşağıdaki tabloda verilmiştir.

Tablo 7. n=21, k=2 için Budanmış Bootstrap Yığına İlişkin

Bazı Regresyon Analizi Sonuçları

n	k	$\hat{\beta}_0$	$\hat{\beta}_1$	S_e	$\sqrt{V(\hat{\beta}_1)}$	R^2
21	2	3.2832	0.3471	0.8068	0.1315	0.2682

Budanmış Bootstrap Yığına ilişkin serpmе diyağramı aşağıda verilmiştir.



Şekil 9. Budanmış Bootstrap Yığını için Serpmе Diyağramı

Budanmış Bootstrap yığını kullanılarak yani $n=21$ gözlem kullanılarak elde edilen $Y_i=3.2832+0.3432X_i$ için artıklar, artık kareler, v_{ii} , r_{ii} , ve t_i^2 'ler aşağıda verilmiştir.

Tablo 8. Budanmış Bootstrap Yığını için $n=21$, $k=2$ iken Artıklar, Artık Kareler, v_{ii} , r_i ve t_i^2 'lere ilişkin işlem Sonuçları

Sıra No	$\hat{Y}_i$	$(Y_i - \hat{Y}_i)$	$(Y_i - \hat{Y}_i)^2$	v_{ii}	r_i	t_i^2
4	4.2204	-0.1204	0.0145	0.1841	-0.1652	0.0259
6	4.3592	0.4408	0.1943	0.1402	0.5892	0.3349
7	4.4286	1.2714	1.6164	0.1214	1.6812	3.1455
8	4.4981	-0.0981	0.0096	0.1048	-0.1284	0.0156
9	4.5675	0.2325	0.0541	0.0903	0.3022	0.0869
10	4.6369	-1.0369	1.0751	0.0779	-1.3383	1.8733
11	4.7063	0.7937	0.6299	0.0676	1.0188	1.0401
12	4.7757	-0.6257	0.3915	0.0594	-0.7997	0.6270
13	4.8451	0.9549	0.9117	0.0534	1.2164	1.5203
14	4.9146	-1.1146	1.2423	0.0495	-1.4170	2.1270
15	4.9840	-0.2340	0.0547	0.0477	-0.2972	0.0841
16	5.0534	-1.1534	1.3303	0.0481	-1.4652	2.2928
17	5.1228	1.0772	1.1603	0.0506	1.3702	1.9736
18	5.1922	-0.8422	0.7094	0.0552	-1.0740	1.1633
19	5.2617	-1.1117	1.2358	0.0619	-1.4225	2.1455
20	5.3311	-0.4811	0.2314	0.0708	-0.6186	0.3700
21	5.4005	0.7995	0.6392	0.0818	1.0342	1.0737
24	5.6087	-0.2087	0.0436	0.1275	-0.2770	0.0730
25	5.6782	0.4218	0.1779	0.1470	0.5661	0.3088
26	5.7476	0.7524	0.5661	0.1686	1.0227	1.0485
27	5.8170	0.2830	0.0801	0.1923	0.3903	0.1455
Toplam	105.1495	0.0005	12.3684	2.0000		

Şekil 9'daki serpm diyagramı incelendiğinde veri merkezine uzak konumda olan gözlem görünmemektedir. Bu şekilde görüldüğü gibi gözlemler serpm diyagramı üzerinde homojen olarak saçılmıştır. Bu durumu pekiştirmek için simülasyon çalışması sonuçları bozulma kriterlerini sağlar nitelikte olmalıdır. Simülasyon çalışmasına ilişkin değerler tablo 9'da verilmiştir.

Budanmış Bootstrap Yığını içerisinde uç değer kalıp kalmadığının kontrolü önemlidir. Bu işlem bozulmanın ölçüsü olarak adlandırılmaktadır. Bozulmanın ölçüsü için Bootstrap Örneklerinden yararlanılacaktır. Bootstrap Örnekleri basit tesadüfi örneklemeye göre yerine koyarak yöntemiyle çekildi. $n_b=7$ orta büyüklükte ALT'lar çekilerek bozulmanın kontrolü için Bootstrap tahminlerinden yararlanıldı. Bootstrap tahminlerinden yararlanılarak bozulma olup olmadığının tesbiti için (4.1) ve (4.2) 'de verilen bozulma kriterleri

$$|R_{BBY}^2 - \bar{R}_B^2|$$

ve

$$\frac{|A.K.O_{BBY} - A.K.O_B|}{A.K.O_{BY}}$$

için bazı sonuçlar tablo 9'da verilmiştir.

Tablo 9. Bootstrap Örneğinin çapı $n_b=7$ için Bootstrap Tahminlerine İlişkin

Simülasyon Sonuçları Tablosu

ΣB	ΣAKO_B	ΣR_B^2	$\overline{A.K.O}_B$	$\bar{R}_B^2$	$\frac{ A.K.O_{BBY} - A.K.O_B }{A.K.O_{BY}}$	$ R_{BBY}^2 - \bar{R}_B^2 $
50	29.7448	17.2016	0.5949	0.3440	0.0285	0.0758
100	64.4695	34.3143	0.6447	0.3431	0.0081	0.0749
150	99.1846	50.8118	0.6612	0.3387	0.0052	0.0705
200	131.6400	66.2787	0.6582	0.3314	0.0037	0.0632
300	196.1461	96.4407	0.6538	0.3215	0.0015	0.0533

Burada B, Budanmış Bootstrap yığınının çekilen n_b çaplı Bootstrap örneklerin sayısıdır. B Bootstrap Örneklerinin sayıları 100, 200 ve 300 için hesaplanan bozulma kriterleri tablo 9'dan alınmıştır. Bozulma kriteri,

$$|R_{BBY}^2 - \bar{R}_B^2|$$

için 100, 200 ve 300 Bootstrap örneğine ilişkin simülasyon sonuçları sırasıyla 0.0749, 0.0632 ve 0.0533 dir. B Bootstrap Örneklerinin sayısı arttıkça bozulma kriterinin değeri sıfıra yaklaşmaktadır. Bu kritere göre budanmış Bootstrap yığını için bozulmanın olmadığı söylenebilir. Diğer bozulma kriteri,

$$\frac{|A.K.O_{BBY} - A.K.O_B|}{A.K.O_{BY}}$$

için 100, 200 ve 300 Bootstrap örneğine ilişkin simülasyon sonuçları sırasıyla 0.0081, 0.0037 ve 0.0015 dir. B Bootstrap Örneklerinin sayısı arttıkça bozulma kriterinin değeri sıfıra yaklaşmaktadır. Bu kritere göre de budanmış Bootstrap yığını için bozulmanın olmadığı söylenebilir. Her iki kriter içinde B Bootstrap Örnekleri ile veride bozulma kontrol edilmiştir. Bu durumda Budanmış Bootstrap Yığınının bozulmanın olmadığını söyleyebiliriz. Simülasyon sayısını ne kadar çok arttırılırsa veri için bozulmayı kontrol etmek de o kadar kolay olmaktadır.

Simülasyon çalışması sonucuna göre budanabilecek gözlem gözükmemektedir. Bu durumda Budanmış Bootstrap Yığını aşağıdaki gibi elde edilir.

$$B.B.Y.=\{ 2, 4, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, \\ 18, 19, 20, 21, 24, 25, 26, 27\}$$

Budanmış Bootstrap Yığını içerisinde uç değer kalıp kalmadığının kontrolü önemlidir. Bu amaçla yapılan simülasyon çalışması neticesinde budama işlemine son verilmiştir. Veride $\alpha=0.15$ 'e göre yeterli budama yapılmıştır.

Verinin budanmasına ilişkin α 'nın bazı değerleri için yapılabilecek budamalara kısaca değinilecektir. $\alpha=0.10$ iken X gözlem değerleri için budama kriterinin değeri 0.0492 hesaplanır. Daha önce Tablo 4'te $h_x(i,i)$ değerleri verilmişti. Bu $h_x(i,i)$ değerleri 0.0492 değeri ile karşılaştırıldığında 1, 2, 3, 4, 5, 26, 27, 28, 29 ve 30 nolu gözlemler budanır. X gözlemleri için alt ve üst sınır değerleri,

$$X_a=2.9966 \text{ ve } X_u=7.0034$$

dür. Y gözlem değerleri için budama kriterinin değeri 0.0158 hesaplanır. Tablo 4'ün yedinci sütunundaki hesaplanmış $h_y(i,i)$ değerleri ile 0.0158 karşılaştırıldığında 1, 2, 3, 10, 14, 22, 23, 26, 28, ve 30 nolu gözlemler budanır. Y gözlemleri için alt ve üst sınır değerleri,

$$Y_a=3.8333 \text{ ve } Y_u=6.2333$$

dir. Student tipi artıklara göre t^2_i değeri 2.9013 tablo değerinden büyük olan gözlemler budanır. Tablo 5'in yedinci sütunundaki hesaplanmış t^2_i değerleri 2.9013 ile karşılaştırılarak budanacak gözlemler belirlenir. Bu kritere göre 1, 3, 5, 7, 22, 28, 29 ve 30'nolu gözlemler budanır. Student tipi artıklar r_i 'ler için alt ve üst sınır değerleri

$$r_a=-1.0545 \text{ ve } r_u=1.0545$$

dir. Her üç budama kriterine göre 1, 2, 3, 4, 5, 10, 14, 22, 23, 26, 27, 28, 29 ve 30 nolu gözlemler Bootstrap Yığınınından budanır. Budanmış Bootstrap Yığını

$$\{ 6, 7, 8, 9, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 24, 25 \}$$

nolu gözlemlerden oluşur. Budanan gözlemlerin oranı 0.46'dır. Uç değer olarak belirlenen gözlemlerin oranı 0.15 büyük olduğu durumlarda genellikle fazla budamaya ihtiyaç duyulur. Buda α 'nın değerini arttırmakla başarılabilir.

$\alpha=0.20$ iken X gözlem değerleri için budama kriterinin değeri 0.1009 hesaplandı. Tablo 4'de altıncı sütunda $h_x(i,i)$ değerleri verilmişti. Bu $h_x(i,i)$ değerleri 0.1009 değeri ile karşılaştırıldığında X'in gözlem değerleri için budama yapılamadı. X gözlemleri için alt ve üst sınır değerleri,

$$X_a=2.1310 \text{ ve } X_u=7.8690$$

dır. Y gözlem değerleri için budama kriterinin değeri 0.0676 hesaplandı. Tablo 4'ün yedinci sütununda hesaplanmış $h_y(i,i)$ değerleri verildi. Bu $h_y(i,i)$ değerleri 0.0676 değeri ile karşılaştırılır. Bu karşılaştırmaya göre 1, 3, 28, ve 30 nolu gözlemler budanır. Y gözlemleri için alt ve üst sınır değerleri,

$$Y_a=2.5511 \text{ ve } Y_u=7.5155$$

dir. Student tipi artıklara göre t^2_i değeri 1.7259 tablo değerinden büyük olan gözlemler budanır. Tablo 5'in yedinci sütunundaki hesaplanmış t^2_i değerleri 1.7259 ile karşılaştırıldığında 1, 3, 5, 7, 22, 28, 29 ve 30'nolu gözlemler budanır. Student tipi artıklar r_i 'ler için alt ve üst sınır değerleri

$$r_a=-1.2970 \text{ ve } r_u=1.2970$$

dir. Her üç budama kriterine göre 1, 3, 5, 7, 22, 28, 29 ve 30'nolu gözlemler Bootstrap Yığınınından budanır. Budanmış Bootstrap Yığını

$$\{ 2, 4, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, \\ 21, 23, 24, 25, 26, 27 \}$$

nolu gözlemlerden oluşur.

$\alpha=0.25$ iken X gözlem değerleri için budama kriterinin değeri 0.1450 hesaplandı. Tablo 4'ün altıncı sütunundaki hesaplanmış $h_x(i,i)$ değerleri verilmişti. Bu $h_x(i,i)$ değerleri 0.1450 ile karşılaştırıldığında X'in gözlem değerleri için budama yapılamadı. X gözlemleri için alt ve üst sınır değerleri,

$$X_a=1.5607 \text{ ve } X_u=8.4393$$

dür. Y gözlem değerleri için budama kriterinin değeri 0.1117 hesaplandı. Tablo 4'ün yedinci sütunundaki hesaplanmış $h_y(i,i)$ değerleri daha önce verilmişti. Hesaplanmış $h_y(i,i)$ değerleri ile 0.1117 karşılaştırıldığında 1, 3, 28, ve 30 nolu gözlemler budanır. Y gözlemleri için alt ve üst sınır değerleri,

$$Y_a=1.8424 \text{ ve } Y_u=8.2240$$

dür. Student tipi artıklara göre t^2_i değeri 1.3822 tablo değerinden büyük olan gözlemler budanır. Tablo 5'in yedinci sütunundaki hesaplanmış t^2_i değerleri daha önce verilmişti. Hesaplanmış t^2_i değerleri 1.3822 ile karşılaştırıldığında 1, 3, 5, 7, 22, 28, 29 ve 30'nolu gözlemler budanır. Student tipi artıklar r_i 'ler için alt ve üst sınır değerleri

$$r_a=-1.1121 \text{ ve } r_u=1.1121$$

dir. Her üç budama kriterine göre 1, 3, 5, 7, 22, 28, 29 ve 30'nolu gözlemler

Bootstrap Yığınınından budanır. Budanmış Bootstrap Yığını

{ 2, 4, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,
21, 23, 24, 25, 26, 27}

nolu gözlemlerden oluşur.

Bu çalışmada α 'nın değerleri 0.10, 0.15, 0.20 ve 0.25 alınarak budamalar gerçekleştirildi. Bilinen artıklara göre uç değer inceleme bazen iyi sonuç vermeyebilir. Bu yöntem dar bir bölgede kümelenmiş gözlemler için iyi sonuç vermeyebilir. Bunun sebebi ise uç değer olmayan gözlemlerde uç değer olarak budanabilir. Uç değer olmayan gözlemlerin de uç değer olarak budanmasına engel olmak için α 'nın değerini arttırmakla bunu başarabiliriz.

BÖLÜM 3

SONUÇ VE TARTIŞMA

Bu çalışmanın giriş kesiminde; etkili değerler ve uç değerler için yapılan çalışmalara yer verilmiştir. Ayrıca bu bölümde etkili değer ve uç değerlerin belirlenmesinde kullanılan test istatistikleri de tanıtılmıştır. Aynı bölümde ALT içinde uç değerlerin bulunması durumunda tahminlerin özellikleri ile ilgili bilgilere de yer verilmiştir.

Birinci bölümde yeniden örnekleme yöntemlerinden Bootstrap yöntemi kısaca tanıtılmıştır. Veri içindeki uç değerlerin Bootstrap Örneğine girmesine engel olmak için Bootstrap Örneğinin çapını küçük seçmek, Bootstrap yığınının az budanması ve Bootstrap yığınının fazla budanması yöntemlerinden biri seçilebilir. Bootstrap yığınının budanmasında bağımsız değişken gözlemleri, bağımlı değişken gözlemleri ve student tipi artıklar ile ilgili budama kriterleri ve test istatistikleri tanıtıldı. Ayrıca Bootstrap tahmini ve bozulmanın ölçüsü ile ilgili bilgilere de yer verilmiştir.

İkinci bölümde ise, gerçek veri kullanılarak bir uygulamaya yer verilmiştir. Uygulamada bir özel televizyon kanalının yaptırmış olduğu araştırmadan yararlanılmıştır. Bu özel televizyon kanalının yöneticisi ana haber bülteni öncesi kısaltılmış(özet) haber gösterim ratinginin ana haber ratingini etkilediğini düşünmektedir. Bu amaçla daha önceki rating ölçme yerlerinden rasgele 30 yer seçilmiş ve veri derlenmiştir.

$\alpha=0.15$ iken Bootstrap yığınının budama işlemleri gerçekleştirilmiştir. Budama sonucunda serpmeye diyagramının uzağındaki gözlemlerden kurtularak daha iyi tahminlere götüren modellere ulaşmak

mümkün olabilmektedir. Budanmış Bootstrap Yığını için bozulmanın tespiti çok önemlidir. Bu doğrultuda Budanmış Bootstrap Yığınının B Bootstrap Örnekleri seçilerek, bozulma kriterleri kontrol edilmiştir. Bu kontroller sonucu veride kötü(modele uymayan) gözlemlere raslanılmamıştır. Bu çalışmada haberlerden hemen önce yayınlanan özet haberlerin ratingi, haberlerin ratingini etkilediği düşünülmektedir. Budama öncesi yapılan t testinin sonucunda özet haber ratinginin ana haber ratingini etkilediği sonucuna varılmıştır. Budama işlemi gerçekleştirildikten sonra yapılan t testine göre yine özet haber gösterim ratingi, ana haber gösterim ratingini etkilediği sonucuna varılmıştır. Ancak bu sonuca varılırken tahminlerdeki değişim önemlidir. Bootstrap Yığınındaki gözlemler için hesaplanan artıklar ile Budanmış Bootstrap Yığınındaki gözlemler için hesaplanan artıklar arasındaki farklar önemlidir. Gözlemler için artıklar ne kadar küçük olursa, modele o kadar iyi uyduğunu gösterir. Bu durumda Bootstrap Yığını doğrusal regresyon modeli

$$\hat{Y}_i = 1.7065 + 0.6654X_i$$

ile Budanmış Bootstrap Yığını doğrusal regresyon modeli

$$\hat{Y}_i = 3.2832 + 0.3471X_i$$

dir. Budanmış Bootstrap Yığını regresyon modeli için hesaplanan hata kareler ortalaması, Bootstrap Yığını regresyon modeli için hesaplanan hata kareler ortalamasına göre daha küçüktür. Bundan dolayı Budanmış Bootstrap Yığını tahminlerinin daha tutarlı olduğu söylenebilir.

Bu çalışmada $\alpha=0.15$ 'e göre budama yapıldı. İkinci bölümde α 'nın değişik değerleri için budamalar yapılmıştı. Bu budamalardan elde edilen bazı sonuçlara kısaca değinilecektir.

$\alpha=0.10$ için bağımsız değişken X'in budama kriterine göre 1, 2, 3, 4, 5, 26, 27, 28, 29 ve 30 nolu gözlemler budanır. Bağımlı değişken Y'nin budama kriterine göre 1, 2, 3, 10, 14, 22, 23, 26, 28, ve 30 nolu gözlemler budanır. Student tipi artıklar t_i^2 göre 1, 3, 5, 7, 22, 28, 29 ve 30'nolu gözlemler budanır. Her üç budama kriterine göre 1, 2, 3, 4, 5, 10, 14, 22, 23, 26, 27, 28, 29 ve 30 nolu gözlemler Bootstrap Yığınınından budanır.

Budanmış Bootstrap Yığını

{ 6, 7, 8, 9, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 24, 25 }

nolu gözlemlerden oluşur.

$\alpha=0.20$ için bağımsız değişken X'in budama kriterine göre gözlem budanamadı. Bağımlı değişken Y'nin budama kriterine göre 1, 3, 28, ve 30 nolu gözlemler budanır. Student tipi artıklar t_i^2 'ye göre 1, 3, 5, 7, 22, 28, 29 ve 30'nolu gözlemler budanır. Her üç budama kriterine göre 1, 3, 5, 7, 22, 28, 29 ve 30'nolu gözlemler Bootstrap Yığınınından budanır. Budanmış Bootstrap Yığını

{ 2, 4, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,
21, 23, 24, 25, 26, 27 }

nolu gözlemlerden oluşur.

$\alpha=0.25$ için bağımsız değişken X'in budama kriterine göre budama yapılamadı. Bağımlı değişken Y'nin budama kriterine göre 1, 3, 28, ve 30 nolu gözlemler budanır. Student tipi artıklar t_i^2 'ye göre 1, 3, 5, 7, 22, 28, 29 ve 30'nolu gözlemler budanır. Her üç budama kriterine göre 1, 3, 5, 7, 22, 28, 29 ve 30'nolu gözlemler Bootstrap Yığınınından budanır. Budanmış Bootstrap Yığını

$$\{ 2, 4, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, \\ 21, 23, 24, 25, 26, 27 \}$$

nolu gözlemlerden oluşur.

Bu çalışmada uç değerlerin Bootstrap yığınınından budanması için α 'nın değeri 0.15 alınarak gerçekleştirilmiştir. Bağımsız değişken X'in gözlem değerleri 1, 2, 29 ve 30 nolu gözlemler Bootstrap Yığınınından budanır. Bağımlı değişken Y'nin gözlem değerlerinden 1, 3, 23, 28 ve 30 nolu gözlemler Bootstrap Yığınınından budanır. Student tipi artıklara göre ise 1, 3, 5, 22, 28, 29 ve 30 nolu gözlemler Bootstrap Yığınınından budanır. Her üç budama kriteri için 1, 2, 3, 5, 22, 23, 28, 29 ve 30 nolu gözlemler Bootstrap Yığınınından budanır. $\alpha=0.15$ için Budanmış Bootstrap yığını

$$\{ 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, \\ 18, 19, 20, 21, 24, 25, 26, 27 \}$$

nolu gözlemlerden oluşur. Budanmış Bootstrap Yığınında bozulmanın kontrolü için ikinci bölümde verilen simülasyon tablosundaki değerler bozulma kriterlerine ilişkin şartları sağlar. Yani simülasyon sayısı artıkça bozulma kriterlerinin sıfıra yaklaştığı görülmektedir. Bozulma kriteri,

$|R_{BBY}^2 - \bar{R}_B^2|$ için 100, 200 ve 300 Bootstrap örneğine ilişkin simülasyon sonuçları sırasıyla 0.0749, 0.0632 ve 0.0533 dir. B Bootstrap Örneklerinin sayısı arttıkça bozulma kriterinin değeri sıfıra yaklaşmaktadır. Bu kritere göre budanmış Bootstrap yığını için bozulmanın olmadığı söylenebilir.

Diğer bozulma kriteri, $\frac{|A.K.O_{BBY} - A.K.O_B|}{A.K.O_{BY}}$ için 100, 200 ve 300 Bootstrap

örneğine ilişkin simülasyon sonuçları sırasıyla 0.0081, 0.0037 ve 0.0015 dir. B Bootstrap Örneklerinin sayısı arttıkça bozulma kriterinin değeri sıfıra yaklaşmaktadır. Bu kritere göre de budanmış Bootstrap yığını için bozulmanın olmadığı söylenebilir. Her iki kriter içinde B Bootstrap Örnekleri ile verideki bozulma kontrol edilmiştir. Bu durumda Budanmış Bootstrap Yığnında bozulmanın olmadığını söyleyebiliriz. Simülasyon sayısını ne kadar çok arttırılırsa veri için bozulmayı kontrol etmek de o kadar kolay olmaktadır. Simülasyon çalışması sonucuna göre budanabilecek gözlem gözükmemektedir. Bu durumda Budanmış Bootstrap Yığını elde edilir. Budanmış Bootstrap Yığını içerisinde uç değer kalıp kalmadığının kontrolü önemlidir. Bu amaçla yapılan simülasyon çalışması neticesinde budama işlemine son verilmiştir. Veride $\alpha=0.15$ 'e göre yeterli budama yapılmıştır.

Sonuç olarak, yapılan tespit neticesinde gözlemlerden ne kadarının kullanılabilir olduğuna karar verilebilmektedir. Budanmış Bootstrap Yığını belirli bir bölge içindeki özellikleri itibariyle homojen yapıya sahip gözlemleri birarada tutar. Yapı itibariyle homojenliği bozan gözlemler

Sparks(1988)'ın önerdiği yöntemler vasıtasıyla ayıklanır. Bu da kullanılabilir gözlemler kümesi olarak adlandırılır.



KAYNAKLAR

1. Sparks, S.Ross., "A Distribution Free Method of Detecting Outliers in Regression", *Commun. in Statist., Theory and Method*, 17(3), 705-727 (1988)
2. Johnson, W. and Geisser, S., "A Predictive View of Detection and Characterization of Influential Observations in Regression Analysis", *Journal of American Statistical Association*, 78, 137-144,(1983).
3. Beckman, R.J, and Cook, R.D., "Outliers". *Technometrics*, 25, 119-149, (1983)
4. Hawkins, D.M., Bradu, D. and Kass, G.V., "Location of Several Outliers in Multiple Regression Data using Elemental Sets", *Technometrics*, 26,3, 197-208, (1984).
5. Barnett, V. and Lewis, T., *Outliers in Statistical Data*, Chichester, England: John Wiley, (1985)
6. Atkinson, A.C., "Masking unmasked", *Biometrika*, 73, 3, 533-541, (1986)
7. Andrews, D.F. and Pregibon D., "Finding the Outliers that Matter", *Journal of Royal Statistical Society*, Ser. B,40, 85-93 (1978).
8. Belsley, D.A., Kuh, E. and Welsch, R.E., *Regression Diagnostics: Identifying Influential Data and Source of Collinearity*, New York:John Wiley , (1980).
9. Cook, R.D. and Weisberg, S., *Residuals and Influential in Regression* , London : Chapman and Hall, (1982).
10. Gray, J.B. and Ling, R.F., "K-Clustering as a Detection Tool for Influential Subsets in Regression", *Technometrics*, 26, 4, 305-320, (1984).
11. Draper, N.R.and John, J.A., "Testing for three or fewer Outliers in two Two-way Tables", *Technometrics*, 12, 394-398, (1980)

12. Doornbos, R., "Testing for Single Outlier in a Linear Model", *Biometrics*, 37, 705-711, (1981).
13. Draper R. N., Fuchs C. "Percentage points and power calculations for outlier test in a regression situation", *Commun. in Statist., Theory and Method*, 14(10),2475-2487, (1985).
14. Weisberg, S., *Applied Linear Regression*. New York: John Wiley and Sons, (1980).
15. Efron, B., "Bootstrap Method: Another Look at the Jackknife", *Annals of Statistics*, 7, 1-26, (1979).
16. Huber, P.J., "Robust Regression: Asymptotic Conjectures and Monte Carlo", *Annals of Statistics*, 1, 799-821, (1973).
17. De Jongh, P.J. and De Wett, T., "A Monte Carlo Comparison of Regression Trimmed Mean", *Commun. in Statist., Theory and Method*, 14, 10, 2457-2477, (1985).
18. Kroenker, R., and Basset, G., "Tests of Linear Hypotheses and L_1 Estimation", *Econometrica*, 50, 1577-1583, (1982).
19. Rousseeuw, P.J. "Least median of squares regression", *Journal of American Statist. Assoc*, 79, 871-880, (1984).
20. Cook, R. D., "Detection of influential observations in linear regression", *Technometrics*, 19, 15-18 (1977)
21. Hoaglin, D.C., Welsch, R., "The Hat Matrix in Regression and ANOVA", *The American Statistician*, 32, 17-22, (1978).
22. Hogg, R. V. and Craig, A. T., *Introduction to Mathematical Statistics*, 4th ed., New York: Macmillan Publishing Co., (1978).
23. Efron, B., "Bootstrapping Regression Models", *Biometrika*, 68, 589-599, (1981).
24. Chatterjee, S. and Price, B., *Regression Analysis by Example*. Detection and Correction of Model Violations: Simple Linear Regression. 2th ed., New York: John Wiley and Sons, (1991).

ÖZGEÇMİŞ

01.04.1966 tarihinde Konya Seydişehir’de doğdu. İlkokulu Seydişehir İhsan Tekin İlkokulunda, Orta öğrenimini ise, Seydişehir Mahmut Esat Ortaokulunda ve Seydişehir Endüstri Meslek Lisesinde tamamladı. Gazi Üniversitesi, Fen-Edebiyat Fakültesi İstatistik Bölümünden mezun oldu. 1992 yılı Güz yarıyılında Gazi Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Programına kaydoldu. Şu anda D.İ.E. Ticaret ve Hizmet İstatistikleri Şubesinde görev yapıyor.

