

# İçindekiler

Veri İmport .....	2
Normallik İncelemesi .....	4
Dönüşüm Yöntemleri.....	7
Doğrusallık .....	9
Model Kurma .....	10
Artık Değer İncelemesi .....	11
İkinci Artık İncelemesi.....	15
Üçüncü Artık İncelemesi.....	18
Dördüncü Artık İncelemesi .....	20
Beşinci Artık İncelemesi.....	23
Altıncı Artık İncelemesi .....	25
Yedinci Artık İncelemesi .....	28
Sekizinci Artık İncelemesi .....	30
Dokuzuncu Artık İncelemesi .....	33
Onuncu Artık İncelemesi .....	35
On Birinci Artık İncelemesi .....	38
On İkinci Artık İncelemesi .....	40
On Üçüncü Artık İncelemesi .....	43
On Dördüncü Artık İncelemesi .....	45
Değişen Varyanslılık.....	49
Öz ilişki Sorunu .....	50
Çoklu Bağlantı Sorunu .....	51
Model Anlamlılığı.....	54
Model 13 Katsayı Yorumları .....	56
Uyum Kestirimi .....	58
Ön Kestirim .....	58
Değişken Seçimi Yöntemi .....	59
İleriye Doğru Seçim Yöntemi .....	59
Geriye Doğru Çıkarma Yöntemi.....	60
Adımsal Regresyon Yöntemi.....	62
Ridge Regresyon .....	63

## Veri İmport

```
> veri <- read.csv("C:/veri.txt", sep="")  
> View(veri)
```

Kodu ile veri seti import edilmiştir.

	y	x1	x2	x3	x4
1	66.13559	8.835016	3.8963214	6.563156	1
2	115.03110	9.235936	3.0165488	5.977999	1
3	65.21587	9.404584	3.1102384	6.528822	1
4	51.38807	7.666818	1.6327196	5.348621	1
5	74.98727	9.947126	3.8271889	6.405233	1
6	56.73548	7.435135	1.0227485	3.504581	1
7	56.17426	7.531494	1.4806430	4.554256	1
8	65.72513	8.822960	2.7870391	5.558596	1
9	58.81402	8.145932	2.7082813	5.728513	1
10	72.96574	10.054776	3.1756723	6.091242	1
11	70.74528	9.914614	3.9590335	6.475117	1
12	61.17088	8.815822	2.7137033	6.535581	1
13	70.23022	9.547994	3.1465933	5.796811	1
14	62.28525	8.056788	2.6836445	4.694847	1
15	57.72190	8.277623	1.7524716	5.304629	1
16	64.61631	8.800388	2.7830172	6.014940	1
17	63.64425	8.608458	2.4119904	5.311534	1
18	58.40743	7.868240	1.7703302	4.746382	1
Showing 1 to 19 of 100 entries, 5 total columns					

Veri seti 100 gözlem ve 5 değişkenden oluşmaktadır. Y değişkeni bağımlı değişken olup x1, x2, x3 değişkenleri bağımsız değişkendir. X4 değişkeni ise nitel değişkendir.

```
attach(veri)  
veri[,5] <- as.factor(x4)  
attach(veri)  
x4
```

```
> x4
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2
[38] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3
[75] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
Levels: 1 2 3
```

“Attach” kodu ile veri seti etkin hale getirilmiştir. Daha sonra x4 değişkeni nitel değişken olarak R programına tanımlanmıştır. X4 değişkeni üç düzeyli bir değişkendir.

```
> glimpse(veri)
Observations: 100
Variables: 5
$ y <dbl> 66.13559, 115.03110, 65.21587, 51.38807, 74.98727, 56.73548, 56....
$ x1 <dbl> 8.835016, 9.235936, 9.404584, 7.666818, 9.947126, 7.435135, 7.53...
$ x2 <dbl> 3.8963214, 3.0165488, 3.1102384, 1.6327196, 3.8271889, 1.0227485...
$ x3 <dbl> 6.563156, 5.977999, 6.528822, 5.348621, 6.405233, 3.504581, 4.55...
$ x4 <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
```

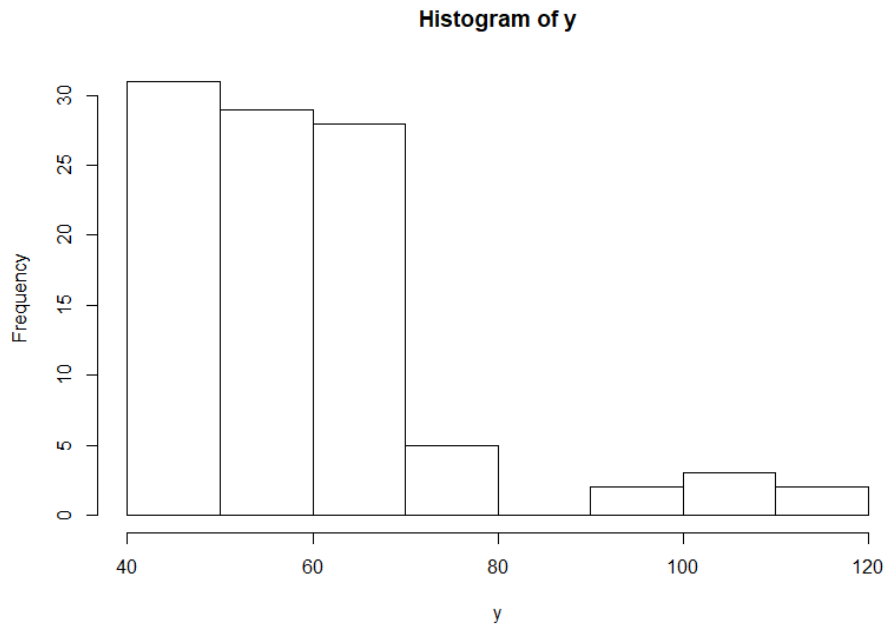
“Tidyverse” kütüphanesi yardımıyla “glimpse” kodu kullanılmaktadır. Bu kod yardımıyla veri setindeki değişkenler hakkında bilgi sahibi olmak mümkündür.

```
> summary(veri)
```

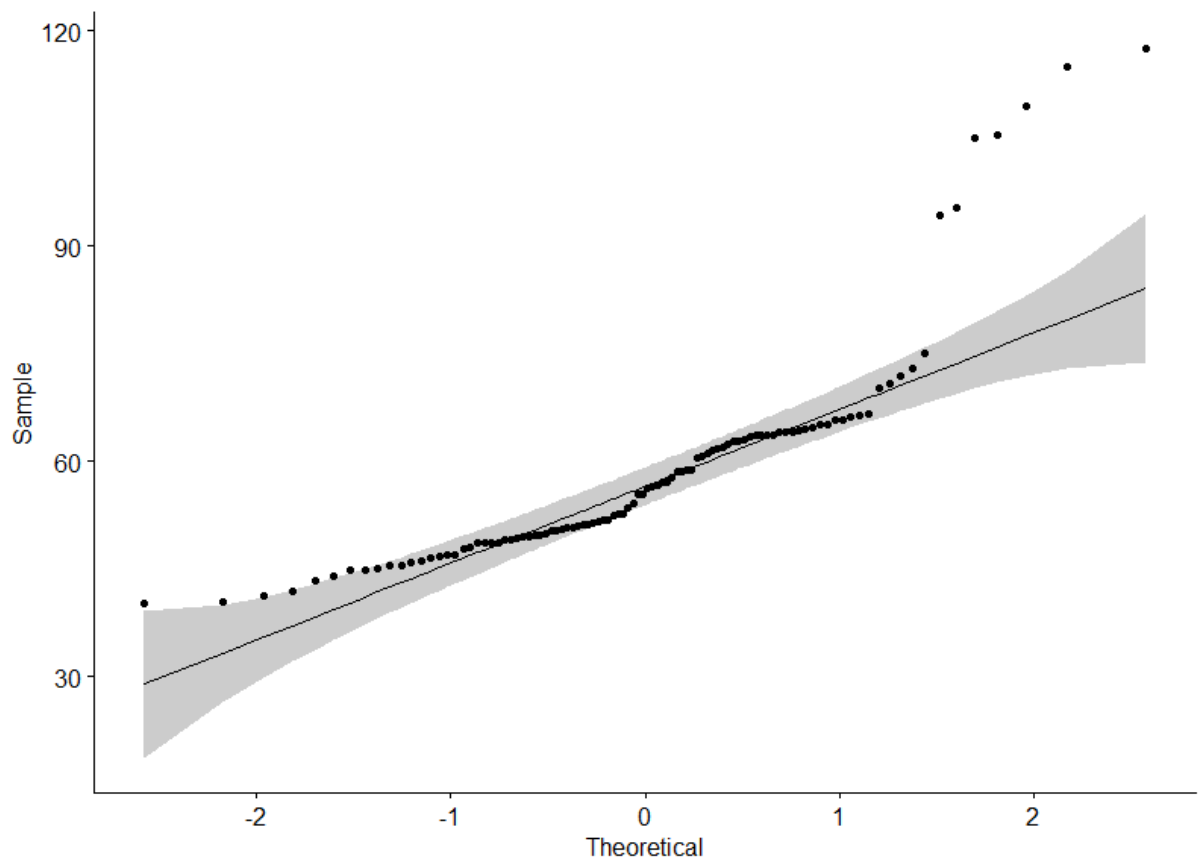
y	x1	x2	x3	x4
Min. : 40.27	Min. : 6.943	Min. : 0.2513	Min. : 3.505	1:25
1st Qu.: 49.31	1st Qu.: 8.393	1st Qu.: 2.4056	1st Qu.: 5.500	2:35
Median : 55.76	Median : 8.876	Median : 2.9067	Median : 5.982	3:40
Mean : 58.94	Mean : 9.013	Mean : 2.9931	Mean : 6.024	
3rd Qu.: 63.76	3rd Qu.: 9.582	3rd Qu.: 3.8021	3rd Qu.: 6.538	
Max. : 117.53	Max. : 10.749	Max. : 5.1124	Max. : 8.087	

“Summary” kodu yardımıyla veri setindeki değişkenlerin özetleyici istatistikleri yorumlanabilmektedir.

# Normallik İncelemesi



**"hist (y)"** kodu ile bağımsız değişkenin (y) histogram grafiği çizdirilmiştir. Grafik yorumlandığında sağa çarpık bir dağılım olduğu görülmektedir.



"ggpubr" kütüphanesi yardımıyla "ggqqplot(y)" kodu aktif hale getirilmiştir. Dağılım grafiğini daha ayrıntılı yorumlamamızı sağlayan bu kod grafikte aykırı değer olduğunu göstermektedir.

```
> library(nortest)
> lillie.test(y)

      Lilliefors (Kolmogorov-Smirnov) normality test

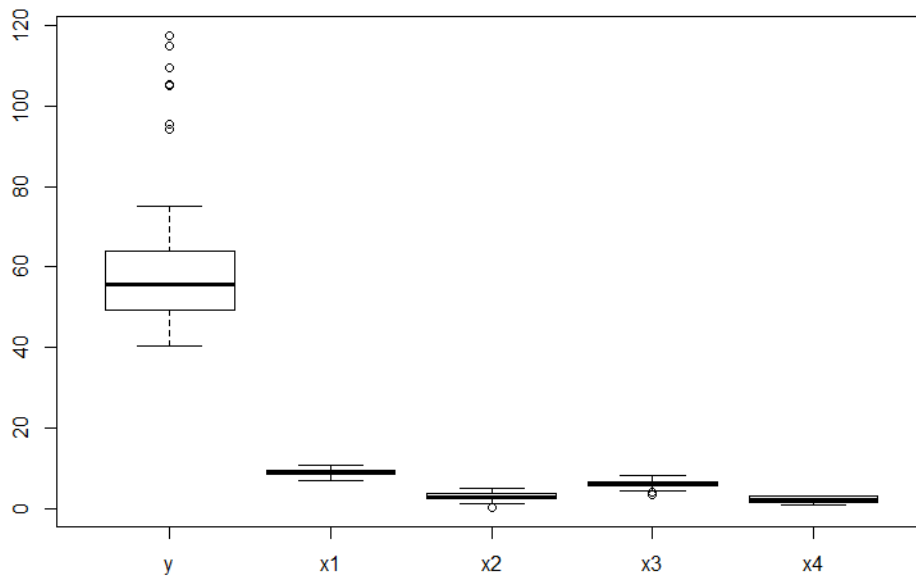
data:  y
D = 0.18908, p-value = 1.994e-09
```

"nortest" kütüphanesi yardımıyla normallik testi gerçekleştirilebilmektedir. Gözlem sayısı 50'den büyük olduğu için Kolmogorav-Smirnov testi kullanılmaktadır. Test için "lillie.test(y)" kodu kullanılmıştır.

Ho: Veri setindeki dağılım ile normal dağılım arasında fark yoktur.

Hs: Veri setindeki dağılım ile normal dağılım arasında fark vardır.

Test sonucunda ortaya çıkan P-value değeri =  $1.994e-09 < \alpha = 0.05$  olduğu için Ho hipotezi reddedilir. Veri setinin normal dağılım göstermediği 0.95 güven düzeyinde söylenebilir.



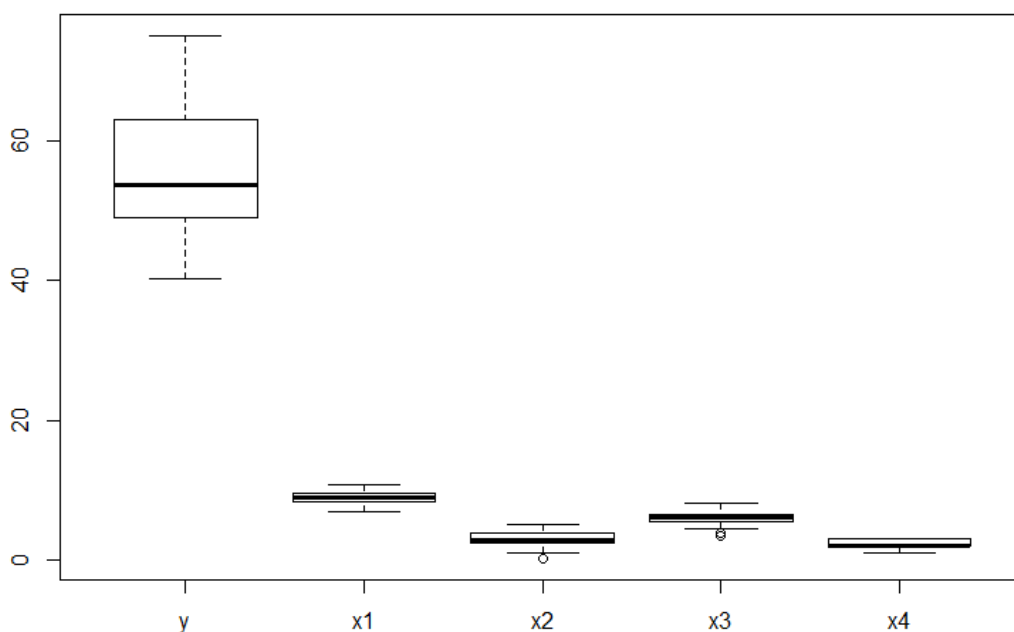
"**boxplot(veri)**" kodu ile bağımsız değişkenin aykırı değerlere sahip olduğu grafikten anlaşılmaktadır.

```
> boxplot(veri)
> a <- which(veri$y %in% boxplot.stats(veri$y)$out)
> a
[1]  2 20 37 44 70 93 96
```

Kodları ile aykırı değerlerin hangi indekste bulunduğu görülmektedir.

```
veri = veri[-c(2,20,37,44,70,93,96),]
attach(veri)
```

Tespit edilip indekslerine ulaşılan aykırı değerler veri setinden çıkarılmıştır. **Attach** komutu ile yenilenen veri seti etkin hale getirilmiştir.



Tekrar “**boxplot (veri)**” kodu kullanıldığında bağımsız değişkende aykırı değer bulunmadığı görülmektedir.

```
> boxplot(veri)
> lillie.test(y)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  y
D = 0.12503, p-value = 0.001073
```

Ho: Veri setindeki dağılım ile normal dağılım arasında fark yoktur.

Hs: Veri setindeki dağılım ile normal dağılım arasında fark vardır.

Test sonucunda ortaya çıkan P-value değeri =  $0.001073 < \alpha = 0.05$  olduğu için Ho hipotezi reddedilir. Veri setinin normal dağılım göstermediği 0.95 güven düzeyinde söylenebilir.

Aykırı değerler çıkarılmasına rağmen veri seti normal dağılım göstermediği için dönüşüm yöntemleri kullanılmalıdır.

## Dönüşüm Yöntemleri

```
> #Logaritmik Dönüşüm
> lny <- log(y)
> lillie.test(lny)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  lny
D = 0.10403, p-value = 0.01467
```

Logaritmik dönüşüm uygulandıktan sonra ortaya çıkan test sonucunda P-value değeri =  $0.01467 < \alpha = 0.05$  olduğu için Ho hipotezi reddedilir. Veri setinin normal dağılım göstermediği 0.05 anlamlılık düzeyinde söylenebilir.

```
> #Karekök Dönüşüm  
> karekok <- sqrt(y)  
> lillie.test(karekok)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: karekok  
D = 0.11427, p-value = 0.004411
```

Karekök dönüşümü uygulandıktan sonra ortaya çıkan test sonucunda P-value değeri =  $0.004411 < \alpha = 0.05$  olduğu için  $H_0$  hipotezi reddedilir. Veri setinin normal dağılım göstermediği 0.05 anlamlılık düzeyinde söylenebilir.

```
> #Ters Dönüşüm  
> bolum <- 1/y  
> lillie.test(bolum)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: bolum  
D = 0.11253, p-value = 0.005469
```

Ters dönüşüm uygulandıktan sonra ortaya çıkan test sonucunda P-value değeri =  $0.005469 < \alpha = 0.05$  olduğu için  $H_0$  hipotezi reddedilir. Veri setinin normal dağılım göstermediği 0.05 anlamlılık düzeyinde söylenebilir.

```
> #Küp Dönüşüm  
> y_kup <- sign(veri$y) * abs(veri$y)^(1/3)  
> lillie.test(y_kup)
```

Lilliefors (Kolmogorov-Smirnov) normality test

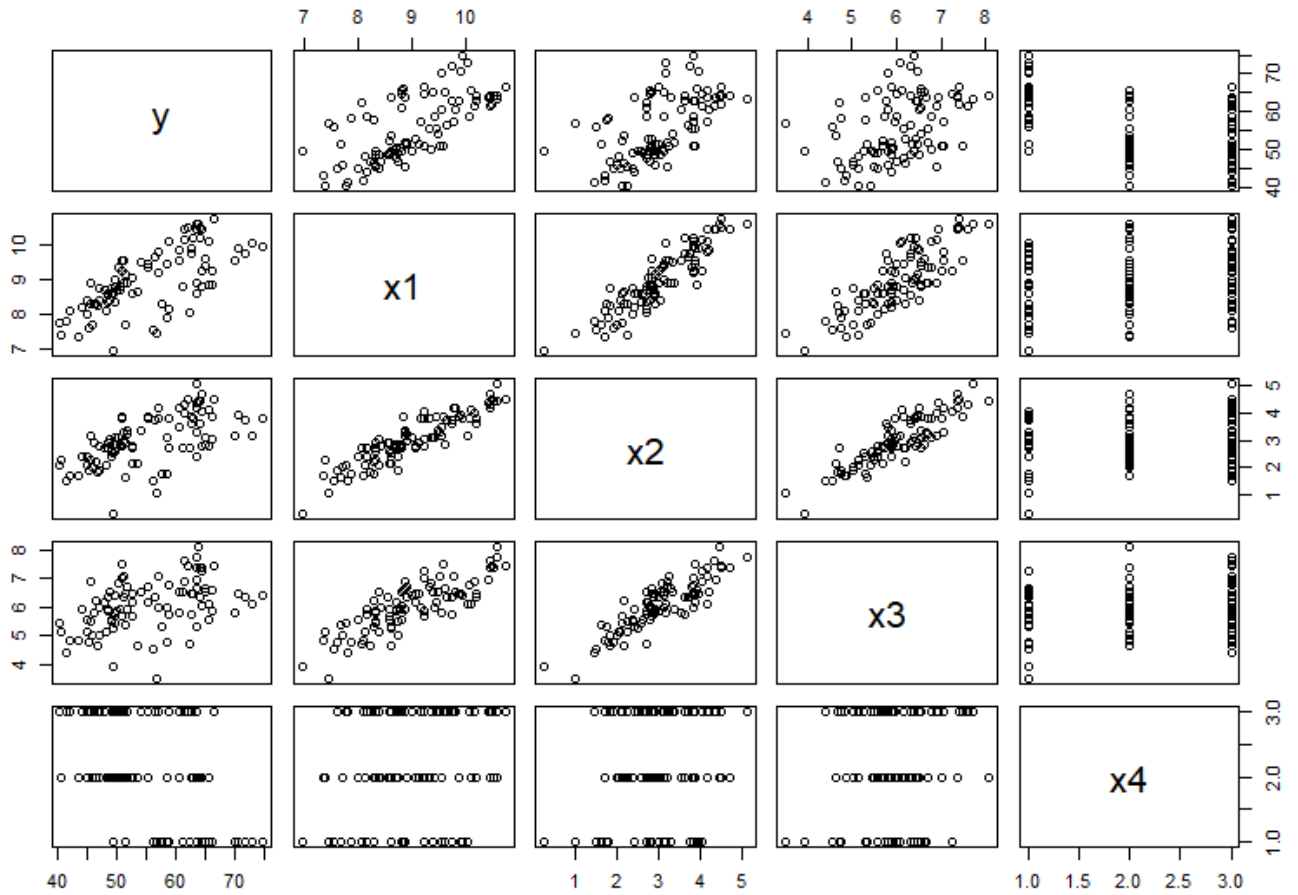
```
data: y_kup  
D = 0.11049, p-value = 0.006993
```



Küp dönüşümü uygulandıktan sonra ortaya çıkan test sonucunda P-value değeri =  $0.006993 < \alpha = 0.05$  olduğu için  $H_0$  hipotezi reddedilir. Veri setinin normal dağılım göstermediği 0.05 anlamlılık düzeyinde söylenebilir.

Dönüşüm yapılmasına rağmen veri seti normal dağılım göstermediği için normal dağıldığı varsayılarak işlemlere devam edilmiştir.

## Doğrusallık



"pairs(veri)" kodu ile doğrusallık grafiği çizdirilmiştir. Grafiğe bakıldığında bağımsız değişkenler arasında (x1-x2 ve x2-x3) ilişki olduğu görülmektedir bu nedenle çoklu bağlantı sorunu olduğu söylenebilir.

## Model Kurma

```
> df <- veri
> model <- lm( y ~. , data=df)
> model

Call:
lm(formula = y ~ ., data = df)

Coefficients:
(Intercept)          x1          x2          x3          x42          x43
      12.516       7.383       3.779      -4.181     -11.423     -14.064

> summary(model)

Call:
lm(formula = y ~ ., data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-3.5801 -0.8805 -0.0229  0.9463  3.8146

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.5165     2.7791   4.504 2.06e-05 ***
x1           7.3826     0.4036  18.293 < 2e-16 ***
x2           3.7786     0.4766   7.927 6.95e-12 ***
x3          -4.1811     0.3649 -11.460 < 2e-16 ***
x42         -11.4230     0.3957 -28.868 < 2e-16 ***
x43         -14.0639     0.3868 -36.363 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.426 on 87 degrees of freedom
Multiple R-squared:  0.9724,    Adjusted R-squared:  0.9708
F-statistic: 612.6 on 5 and 87 DF,  p-value: < 2.2e-16
```

Bağımlı değişken y olmak üzere x1 ,x2 ,x3, x42 ve x43 bağımsız değişkenleri girilmiştir. X41 değişkeninin kılavuz değişken olduğu modelde olmadığından görülmektedir.

Ho: Model istatistiksel olarak anlamlı değildir.

Hs: Model istatistiksel olarak anlamlıdır.

P-value değeri, alfa 0.05 değerinden çok küçük bir değere sahip olduğu için Ho hipotezi reddedilir. Modelin istatistiksel olarak anlamlı olduğu 0.95 güven düzeyinde söylenebilir. Bağımsız değişkenlerin hepsi anlamlıdır. Ayrıca bağımlı değişkenin %97.24 'ünün bağımsız değişkenlerce açıklandığı görülmektedir. Varsayım bozulumlarından söz etmek mümkün değildir.

# Artık Değer İncelemesi

```
#Artık Değer İncelemesi
influence.measures(model)
artik <- ls.diag(model)
artik
```

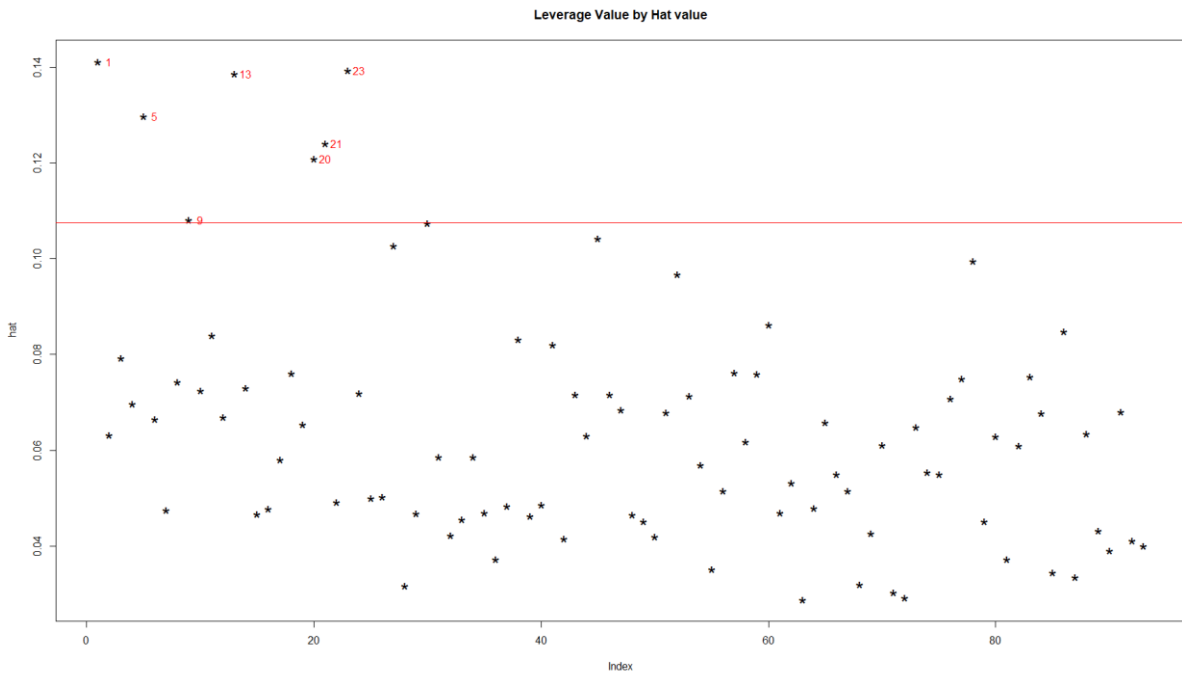
Kodları ile oluşturulan modele ait artık incelemesi yapılmıştır.

→ Gözlem Uzaklığı

"  $\text{Hat} > (2 \cdot 5) / 93$  " ise uç değer vardır. Bu değer 0.1075268 çıkmıştır. Uç değerler 1, 5, 9, 13, 20, 21, 23 indekslerine sahiptir.

```
> artik <- ls.diag(model)
> artik
$std.dev
[1] 1.426183

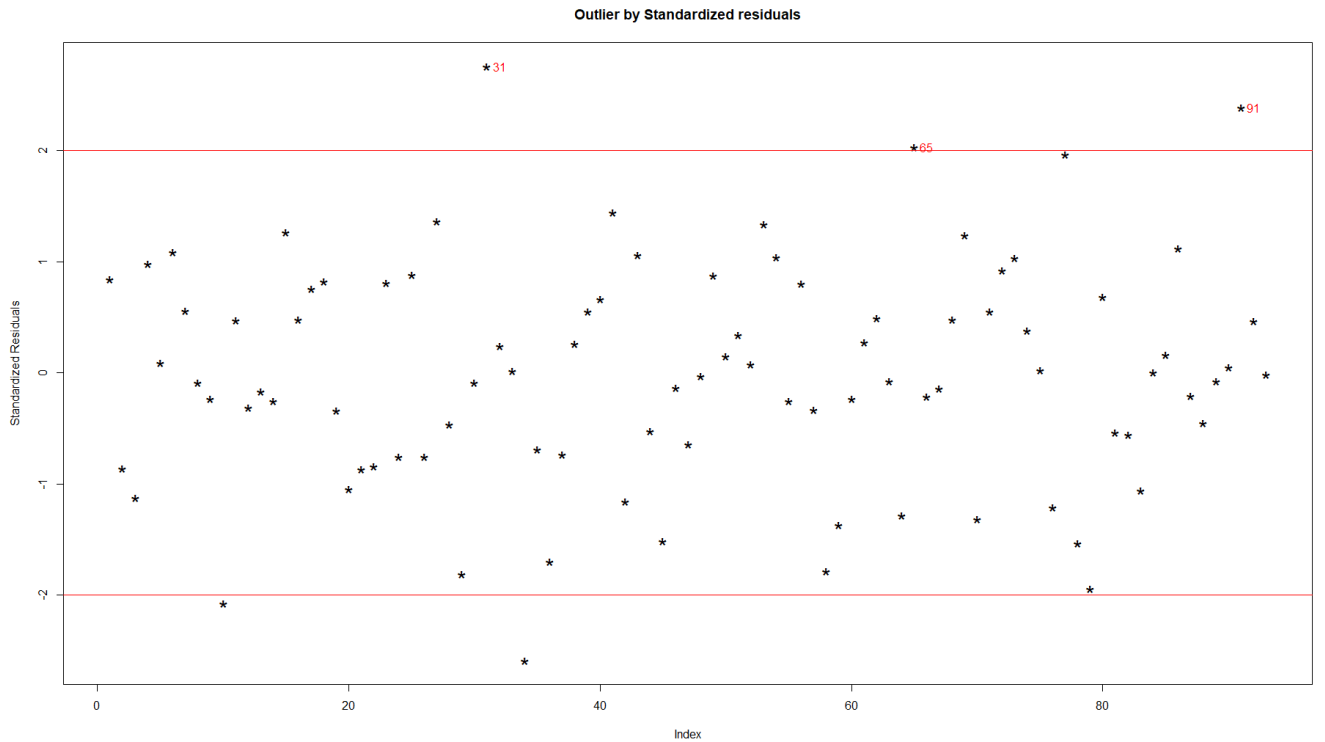
$hat
[1] 0.14112258 0.06322788 0.07925563 0.06972919 0.12979663 0.06650276 0.04758986 0.07425767
[9] 0.10816877 0.07244578 0.08396510 0.06691911 0.13861557 0.07304273 0.04670964 0.04787735
[17] 0.05804075 0.07608047 0.06550142 0.12094678 0.12405236 0.04927858 0.13928383 0.07191372
[25] 0.05008071 0.05036696 0.10269367 0.03169909 0.04682191 0.10743888 0.05858490 0.04230740
[33] 0.04563799 0.05868575 0.04705046 0.03734578 0.04844612 0.08321916 0.04637425 0.04863738
[41] 0.08213618 0.04155420 0.07168691 0.06312966 0.10422868 0.07173499 0.06852098 0.04657105
[49] 0.04517035 0.04200543 0.06793761 0.09673436 0.07133382 0.05703826 0.03524333 0.05161693
[57] 0.07622793 0.06180454 0.07596170 0.08618055 0.04700294 0.05325475 0.02888182 0.04803064
[65] 0.06582338 0.05502484 0.05151685 0.03208503 0.04275435 0.06109050 0.03037629 0.02927415
[73] 0.06493774 0.05539468 0.05506712 0.07087304 0.07495083 0.09948454 0.04524360 0.06299680
[81] 0.03724247 0.06099400 0.07541609 0.06773085 0.03452325 0.08478395 0.03356727 0.06352665
[89] 0.04332671 0.03903727 0.06801078 0.04113289 0.04011196
```



### → Standartlaştırılmış Artık

"(-2,+2)" aralığı dışında kalan değerlere standartlaştırılmış artıklara göre aykırı değer denir. Standartlaştırılmış artıklara göre aykırı değerler 10, 31, 34, 65, 91 indekslerine sahiptir.

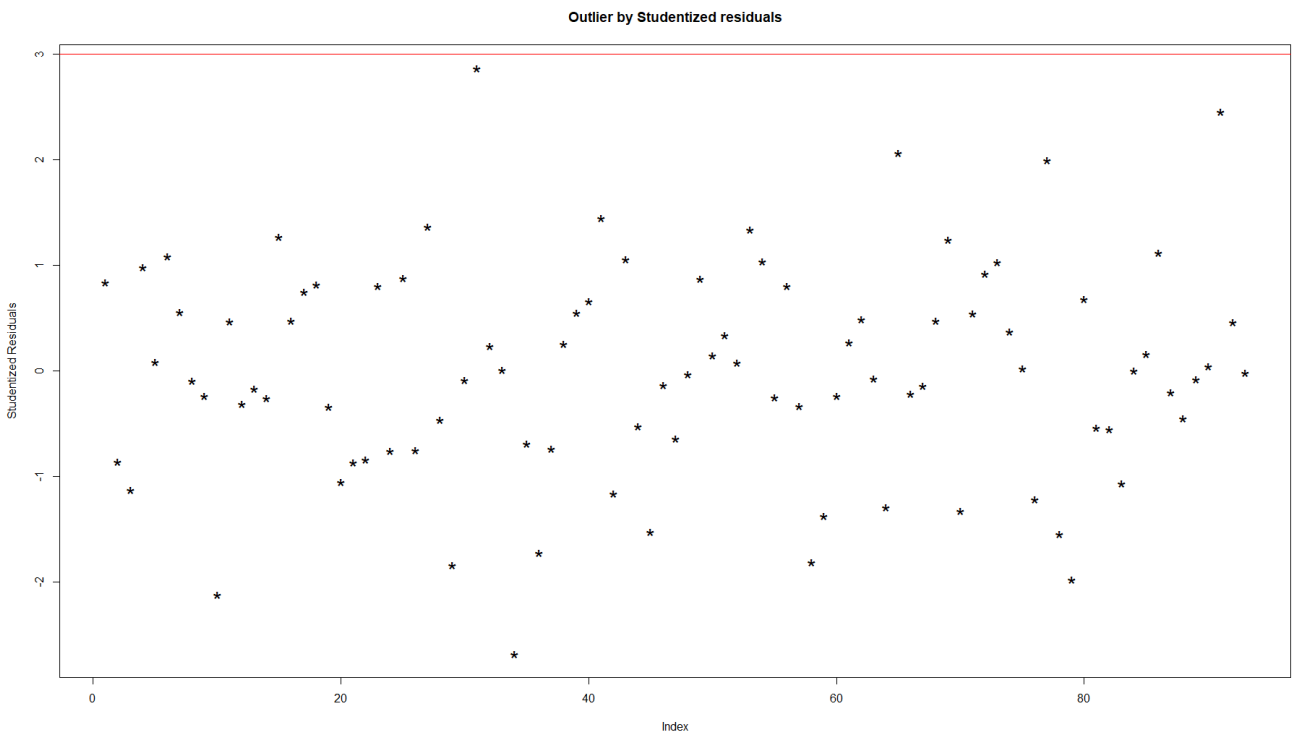
```
$std.res
[1] 0.841500810 -0.858946743 -1.122259116 0.984756041 0.087781595 1.090578224 0.561960846
[8] -0.089425844 -0.232162056 -2.077372857 0.470726482 -0.310680611 -0.167798780 -0.253248685
[15] 1.266211898 0.480752136 0.757224910 0.825577914 -0.336368423 -1.047905466 -0.865249337
[22] -0.837031810 0.807790551 -0.753997036 0.882442442 -0.752113797 1.363980913 -0.462396492
[29] -1.812466450 -0.085472289 2.756660817 0.241025161 0.015987481 -2.587322624 -0.690981219
[36] -1.700271242 -0.734624870 0.263607421 0.555610982 0.667883978 1.444709653 -1.154533699
[43] 1.062077787 -0.523695828 -1.512241799 -0.134842409 -0.639677957 -0.030007305 0.878041969
[50] 0.149451443 0.340312406 0.078688871 1.337990478 1.041968018 -0.251477302 0.806429690
[57] -0.329241707 -1.783509089 -1.365715142 -0.235461336 0.276451306 0.491864297 -0.072934667
[64] -1.281900665 2.032936638 -0.215007327 -0.140328426 0.479945371 1.239469495 -1.318411310
[71] 0.550940131 0.921654703 1.036479492 0.379705852 0.026861063 -1.212949770 1.962958390
[78] -1.531139591 -1.941846048 0.685472508 -0.538538842 -0.553782025 -1.059106040 0.007485844
[85] 0.166198130 1.123319592 -0.203408363 -0.451380561 -0.074698834 0.050180459 2.387230999
[92] 0.469630951 -0.016401599
```



### → Student Tipi Artıklar

"(-3, +3)" aralığı dışında kalan değerlere student tipi aykırı değer denir. Koda bakıldığında bu aralığın dışında kalan değer görülmemektedir. Bu nedenle student tipi artık değer olmadığı söylenebilir.

```
$stud.res
[1] 0.84007645 -0.85764028 -1.12395591 0.98458286 0.08727951 1.09178087 0.55973867
[8] -0.08891450 -0.23089547 -2.11861269 0.46861048 -0.30906142 -0.16685863 -0.25188188
[15] 1.27067666 0.47861737 0.75535373 0.82405376 -0.33464736 -1.04850369 -0.86398771
[22] -0.83557869 0.80616358 -0.75211261 0.88130928 -0.75022176 1.37085597 -0.46029732
[29] -1.83703593 -0.08498322 2.86893073 0.23971600 0.01589536 -2.67748138 -0.68889150
[36] -1.71927758 -0.73266664 0.26219279 0.55339126 0.66574337 1.45392895 -1.15677509
[43] 1.06286911 -0.52150002 -1.52368472 -0.13407922 -0.63749195 -0.02983451 0.87687506
[50] 0.14860912 0.33857636 0.07823811 1.34418029 1.04248756 -0.25011878 0.80479522
[57] -0.32754817 -1.80656282 -1.37263710 -0.23417883 0.27497872 0.48971069 -0.07251651
[64] -1.28672196 2.07100820 -0.21382490 -0.13953540 0.47781206 1.24335225 -1.32410650
[71] 0.54872272 0.92084903 1.03692746 0.37783053 0.02670635 -1.21628674 1.99635405
[78] -1.54324947 -1.97390421 0.68336951 -0.53632955 -0.55156316 -1.05985625 0.00744270
[85] 0.16526644 1.12503361 -0.20228408 -0.44930534 -0.07427067 0.04989195 2.45524811
[92] 0.46751710 -0.01630709
```

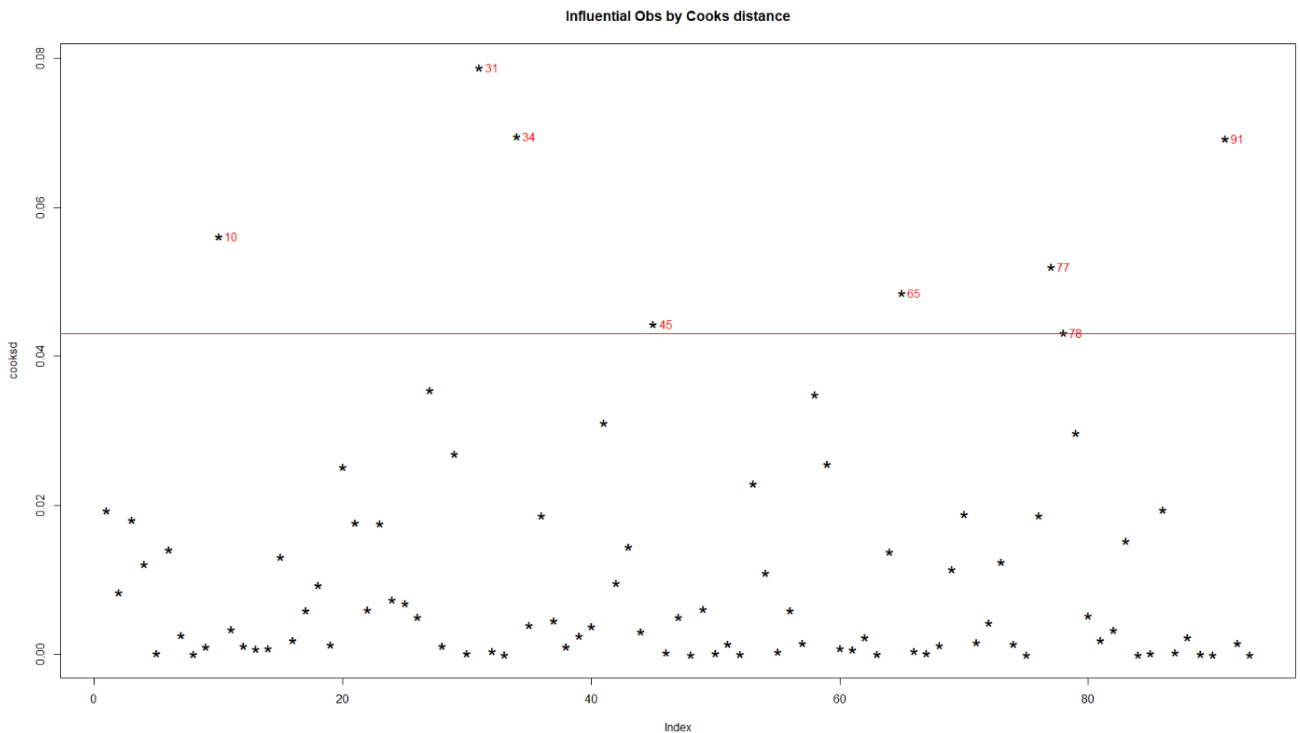


### → Cook Uzaklığı

"n = 93" olduğu için " $DI > 4/n = 93$ " ise etkin gözlem vardır. Etkin değerler 10, 31, 34, 45, 65, 77, 78, 91 indekslerine sahiptir.

\$cooks

```
[1] 1.939201e-02 8.299576e-03 1.806867e-02 1.211466e-02 1.915571e-04 1.412177e-02 2.629973e-03
[8] 1.069119e-04 1.089558e-03 5.617608e-02 3.385110e-03 1.153740e-03 7.551641e-04 8.422876e-04
[15] 1.309311e-02 1.936994e-03 5.888429e-03 9.354141e-03 1.321756e-03 2.518095e-02 1.767088e-02
[22] 6.052539e-03 1.759896e-02 7.341950e-03 6.842350e-03 5.000412e-03 3.548692e-02 1.166579e-03
[29] 2.689452e-02 1.465625e-04 7.881700e-02 4.277241e-04 2.037146e-06 6.955813e-02 3.928938e-03
[36] 1.869203e-02 4.579369e-03 1.051288e-03 2.502012e-03 3.800799e-03 3.112907e-02 9.631843e-03
[43] 1.451800e-02 3.080073e-03 4.434874e-02 2.341859e-04 5.016745e-03 7.330446e-06 6.078646e-03
[50] 1.632267e-04 1.406921e-03 1.105199e-04 2.291874e-02 1.094535e-02 3.850396e-04 5.899160e-03
[57] 1.490829e-03 3.492420e-02 2.555487e-02 8.714387e-04 6.282313e-04 2.268112e-03 2.636752e-05
[64] 1.381825e-02 4.853417e-02 4.486355e-04 1.782624e-04 1.272617e-03 1.143608e-02 1.884953e-02
[71] 1.584853e-03 4.269460e-03 1.243446e-02 1.409162e-03 7.007879e-06 1.870425e-02 5.203344e-02
[78] 4.316609e-02 2.978125e-02 5.265095e-03 1.869839e-03 3.320055e-03 1.524914e-02 6.785401e-07
[85] 1.646156e-04 1.948253e-02 2.395140e-04 2.303536e-03 4.211807e-05 1.704869e-05 6.931137e-02
[92] 1.576859e-03 1.873589e-06
```



İnceleme sonucunda artık değerleri veri setinden çıkarıyoruz. Yeni oluşan veri setinin normal dağılım gösterip göstermediği test edilecektir.

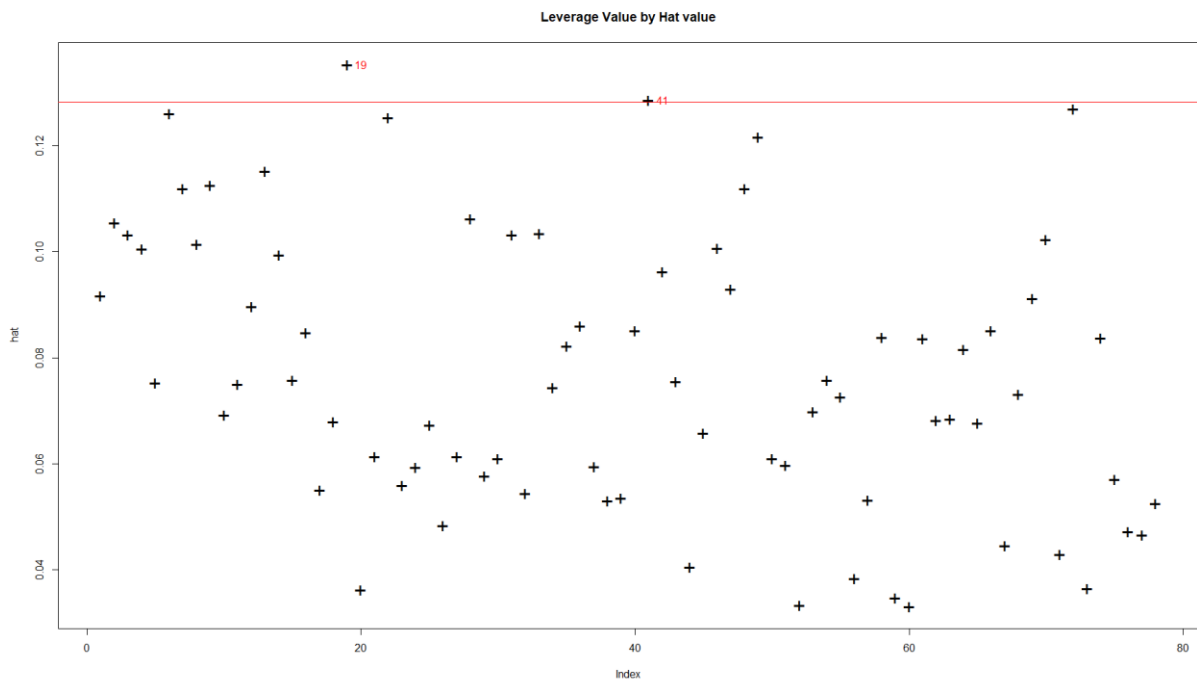
```
df1 = df[-c(31,65,91,10,34,45,77,78,1,5,9,13,20,21,23),]  
attach(df1)  
lillie.test(df1$y)
```

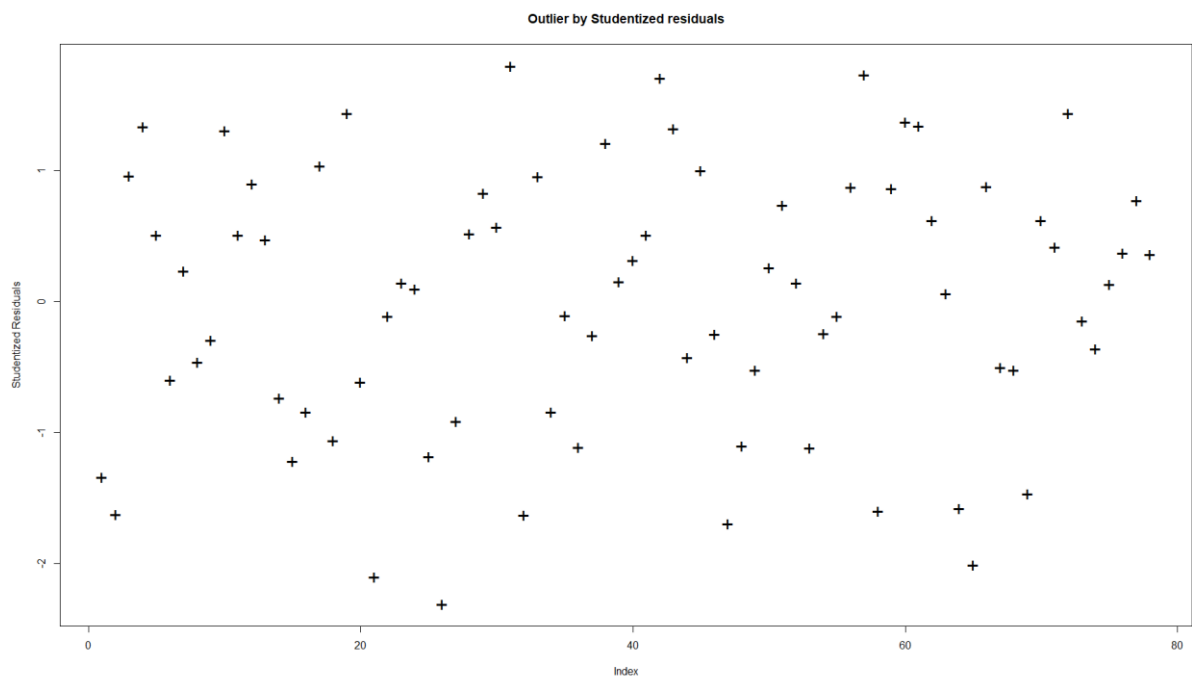
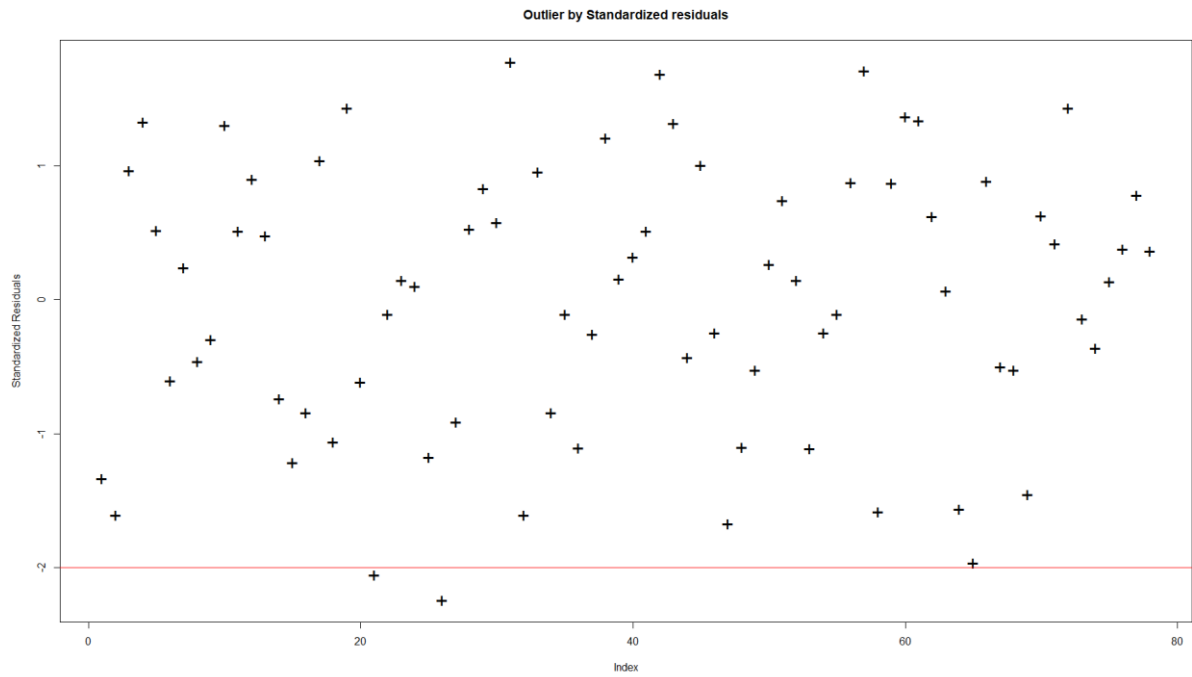
```
> lillie.test(df1$y)  
  
Lilliefors (Kolmogorov-Smirnov) normality test  
  
data: df1$y  
D = 0.13258, p-value = 0.001698
```

Test sonucunda p-value değeri  $0.001698 < \alpha = 0.05$  olduğu için veri setinin normal dağılım göstermediği 0.05 anlamlılık düzeyinde söylenebiliriz.

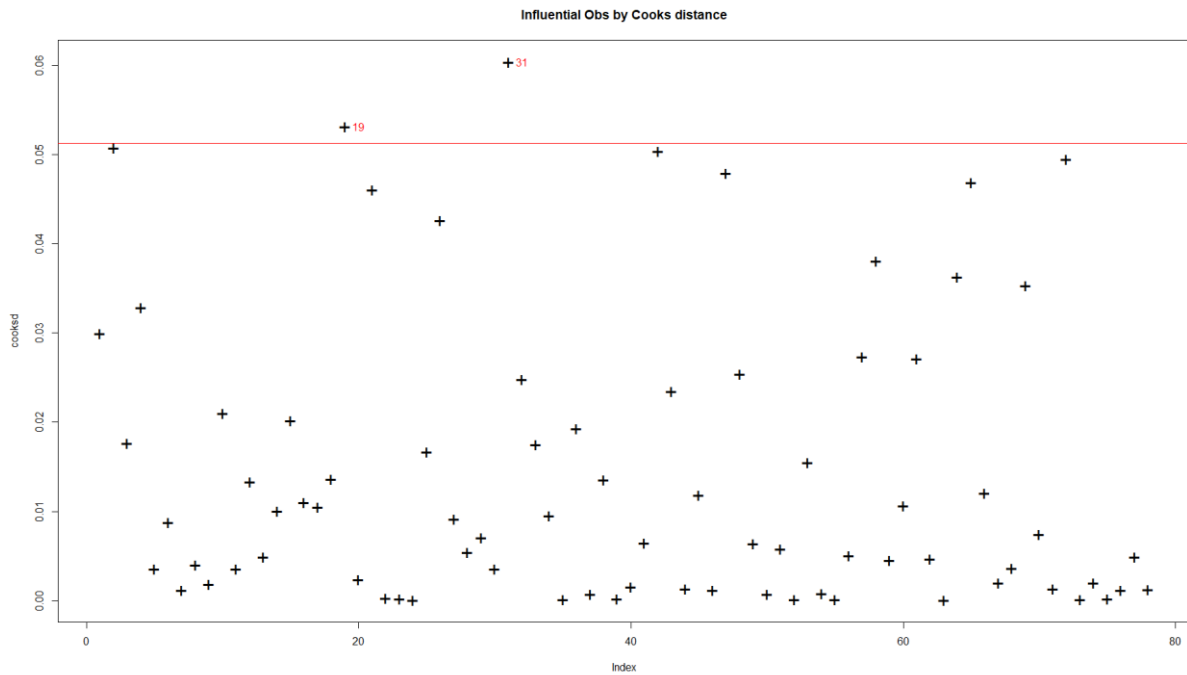
Artık incelemesine devam edilecektir. Buradan sonraki işlemler artık değer kalmayana kadar grafiklerle gösterilecektir.

## İkinci Artık İncelemesi









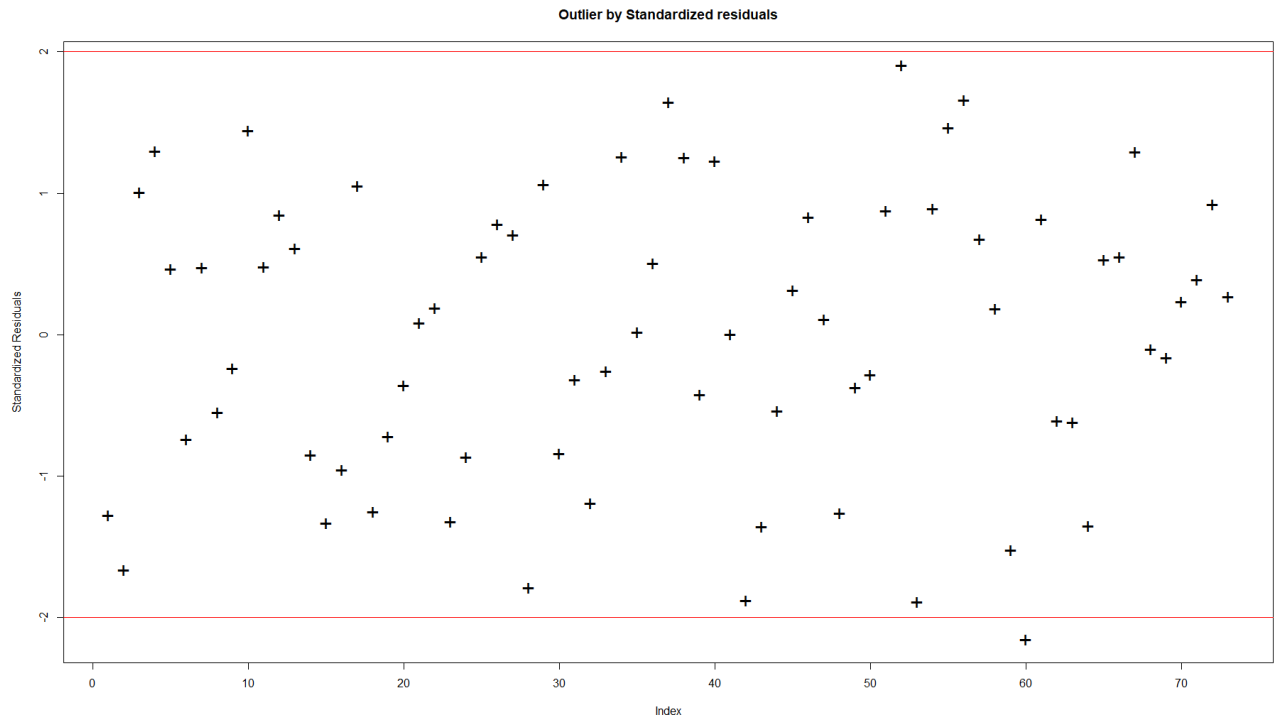
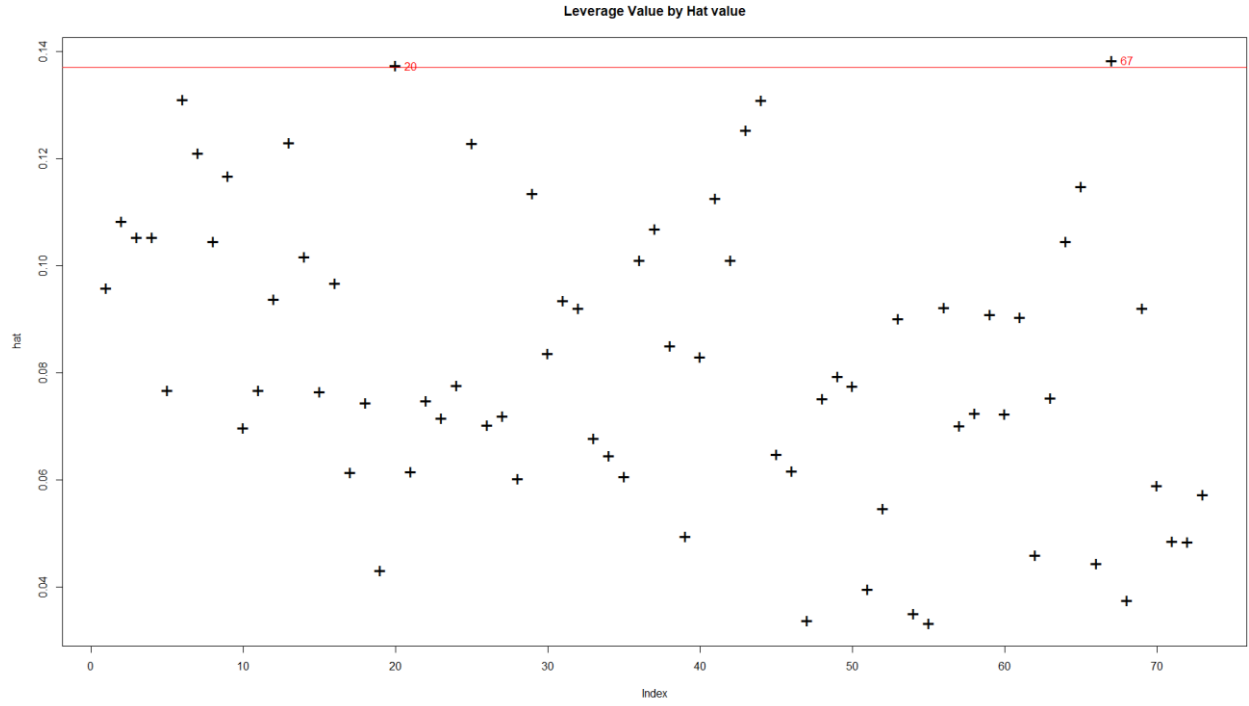
Grafiğe göre 19 ve 31 indekslerine sahip gözlemler çıkarılmalıdır.

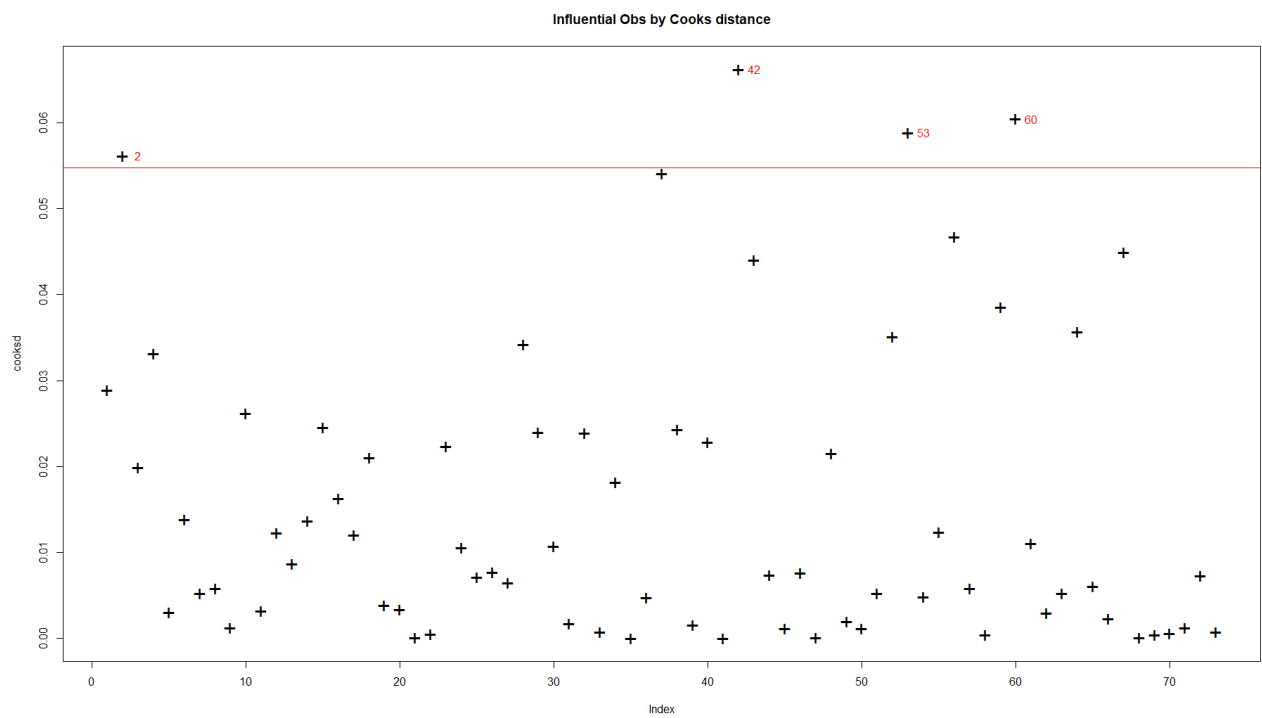
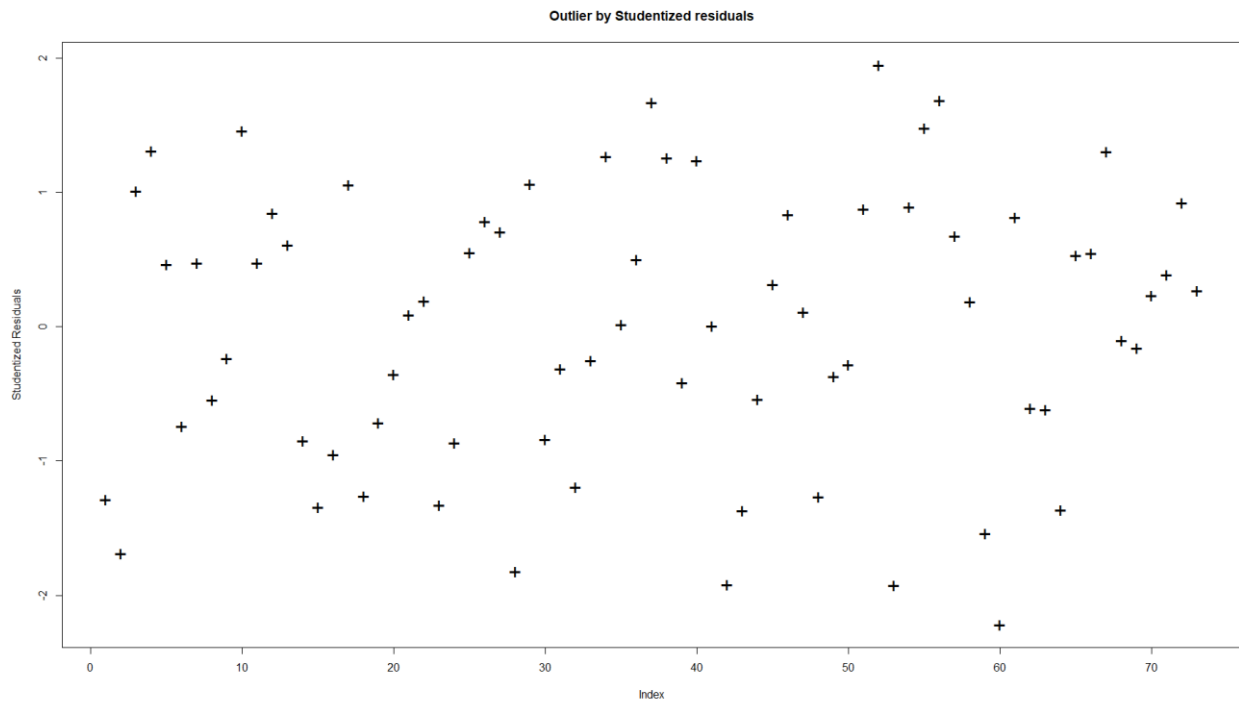
```
df2 = df1[-c(19,31,41,21,26),]  
attach(df2)  
lillie.test(df$y)
```

```
> lillie.test(df2$y)  
  
Lilliefors (Kolmogorov-Smirnov) normality test  
  
data: df2$y  
D = 0.13147, p-value = 0.003211
```

Değerler veri setinden çıkarılmıştır. Yeni veri setinin p-value değeri  $0.003211 < \alpha = 0.05$  olduğu için veri setinin normal dağılım göstermediği 0.05 anlamlılık ile söylenmektedir. Artık incelemesine devam edilmektedir.

## Üçüncü Artık İncelemesi





```
df3 = df2[-c(60,2,42,53,20,67),]  
attach(df3)  
lillie.test(df3$y)
```

```
> lillie.test(df3$y)
```

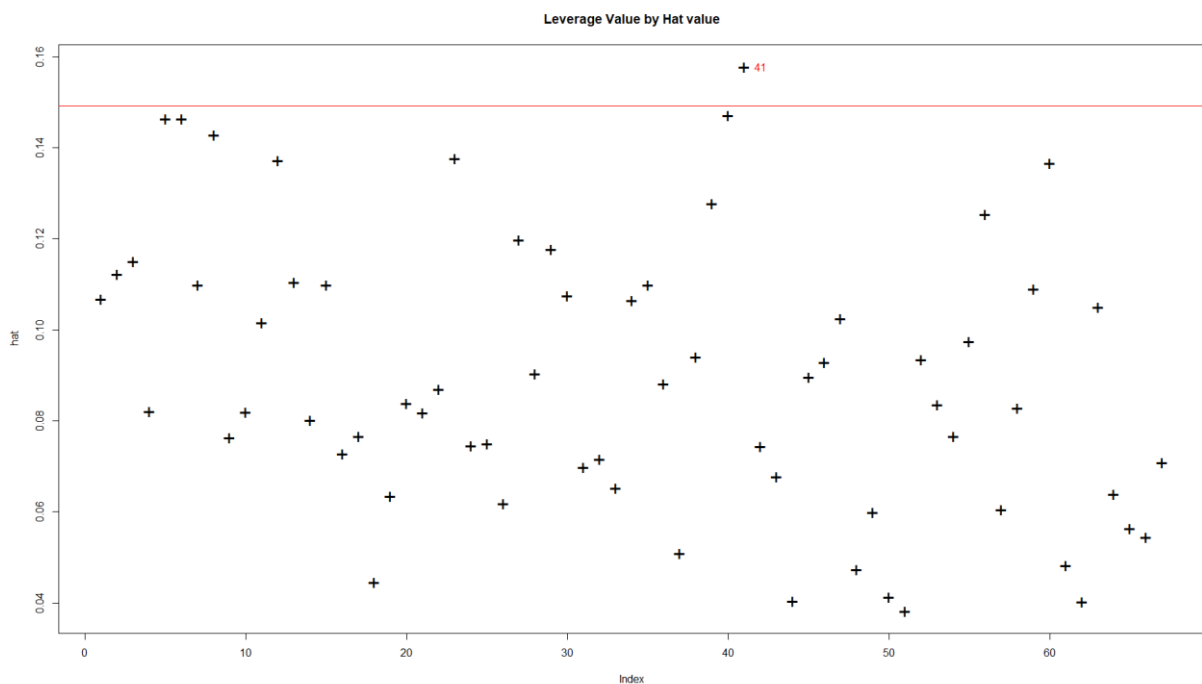
Lilliefors (Kolmogorov-Smirnov) normality test

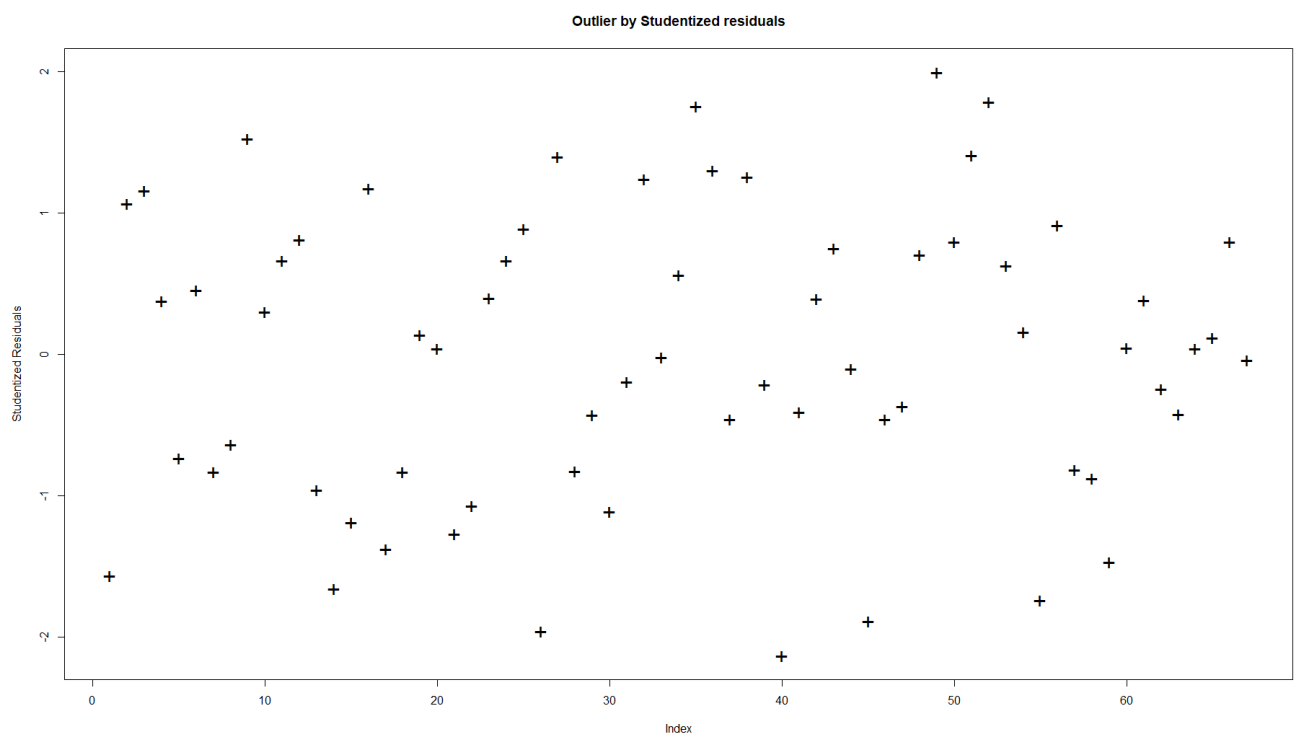
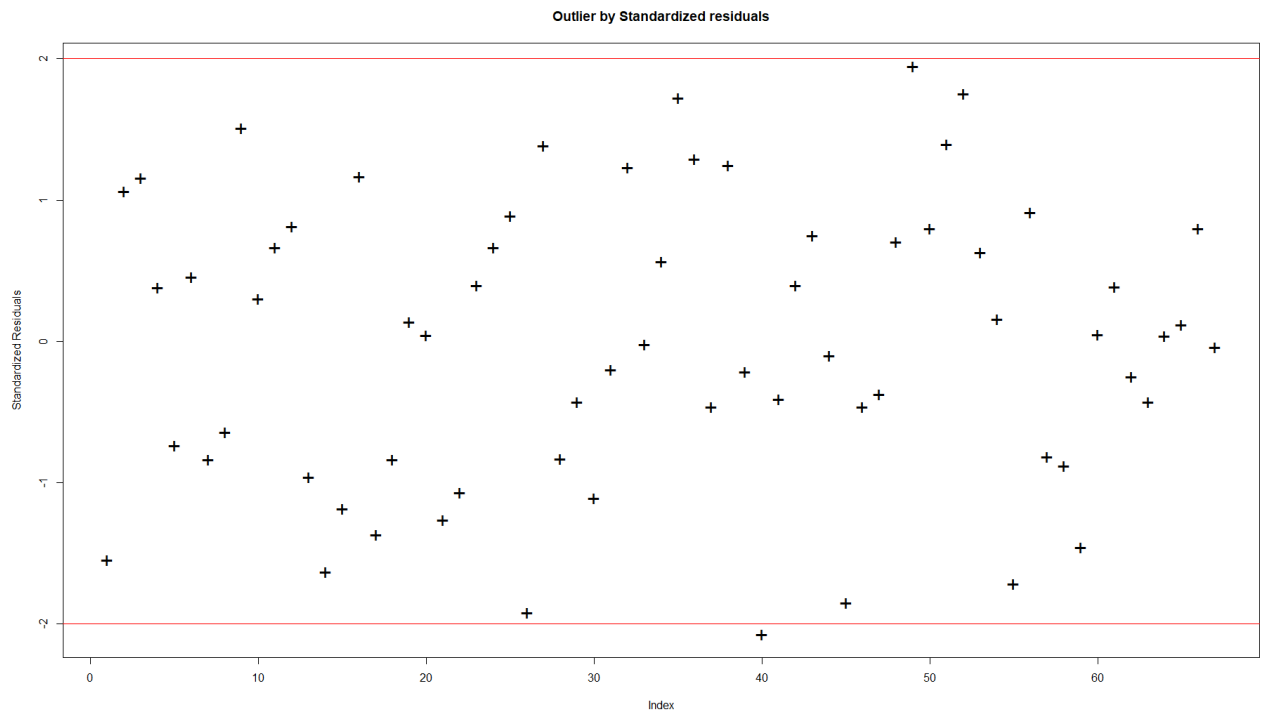
data: df3\$y

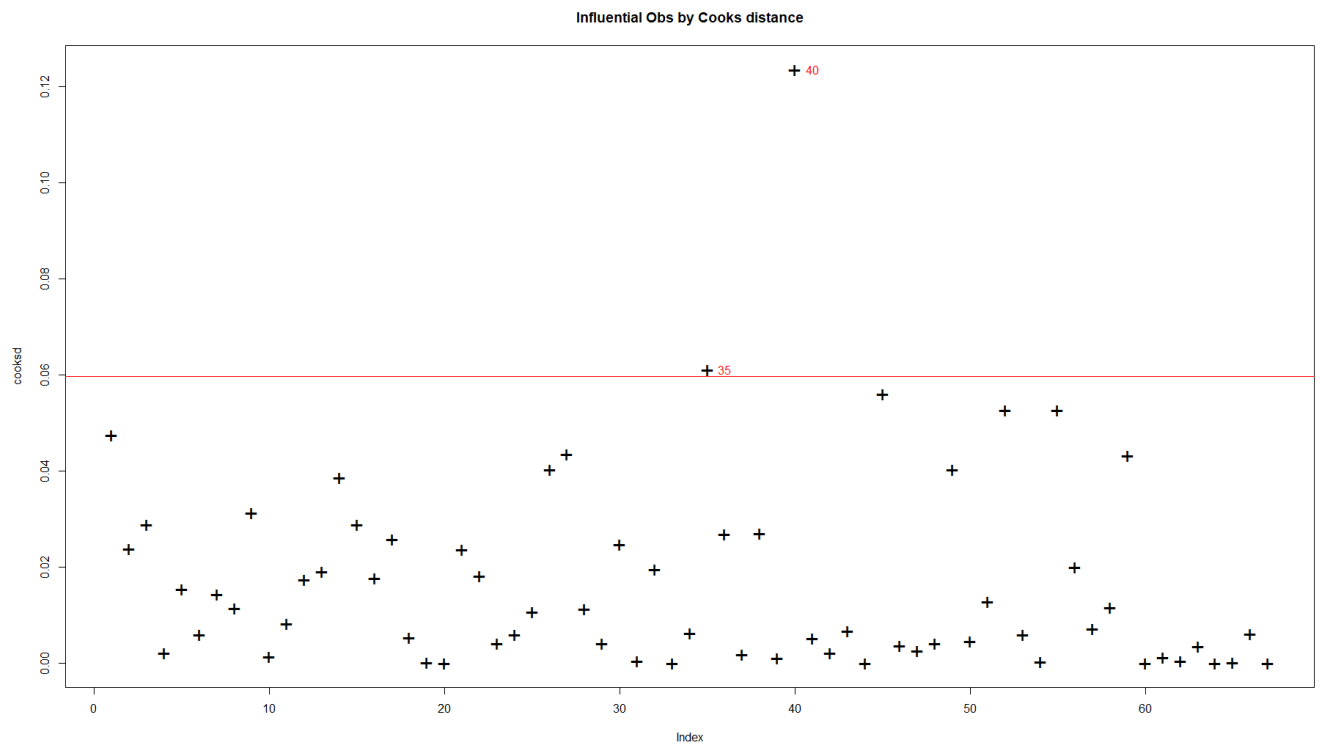
D = 0.13382, p-value = 0.004551

Değerler veri setinden çıkarılmıştır. Yeni veri setinin p-value değeri  $0.004551 < \alpha = 0.05$  olduğu için veri setinin normal dağılım göstermediği 0.05 anlamlılık ile söylenmektedir. Artık incelemesine devam edilmektedir.

### Dördüncü Artık İncelemesi





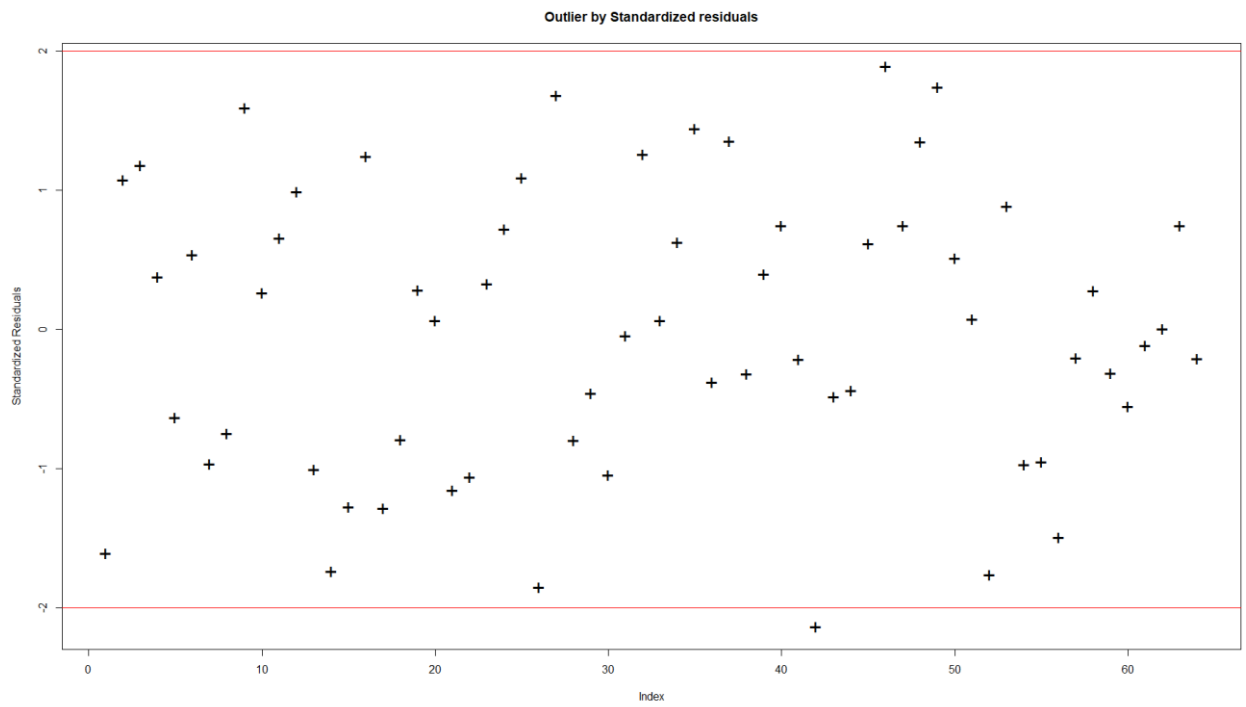
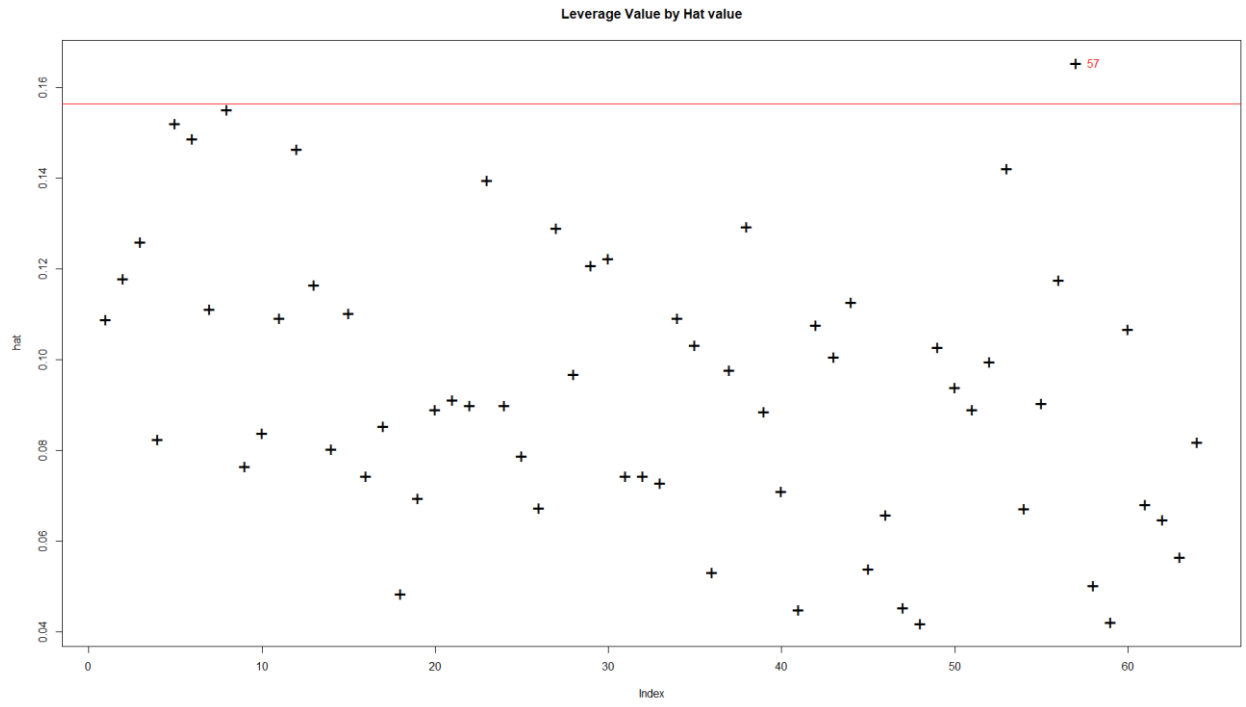


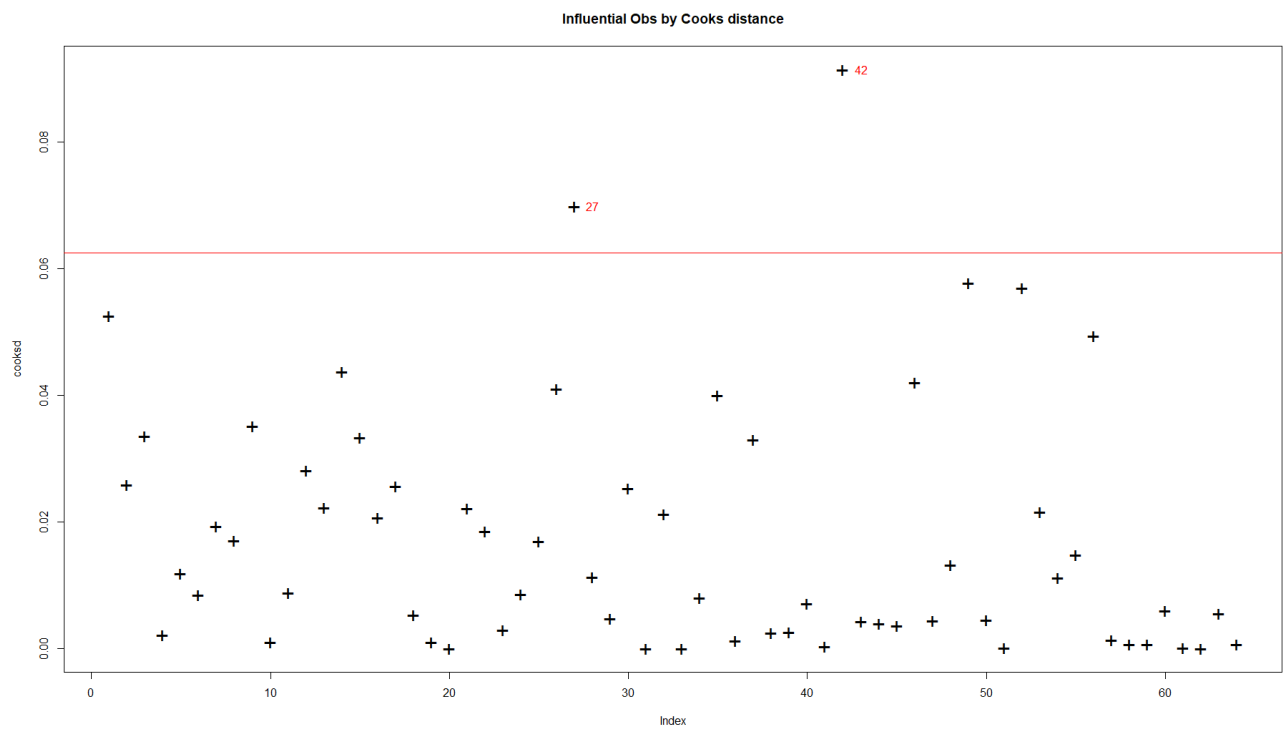
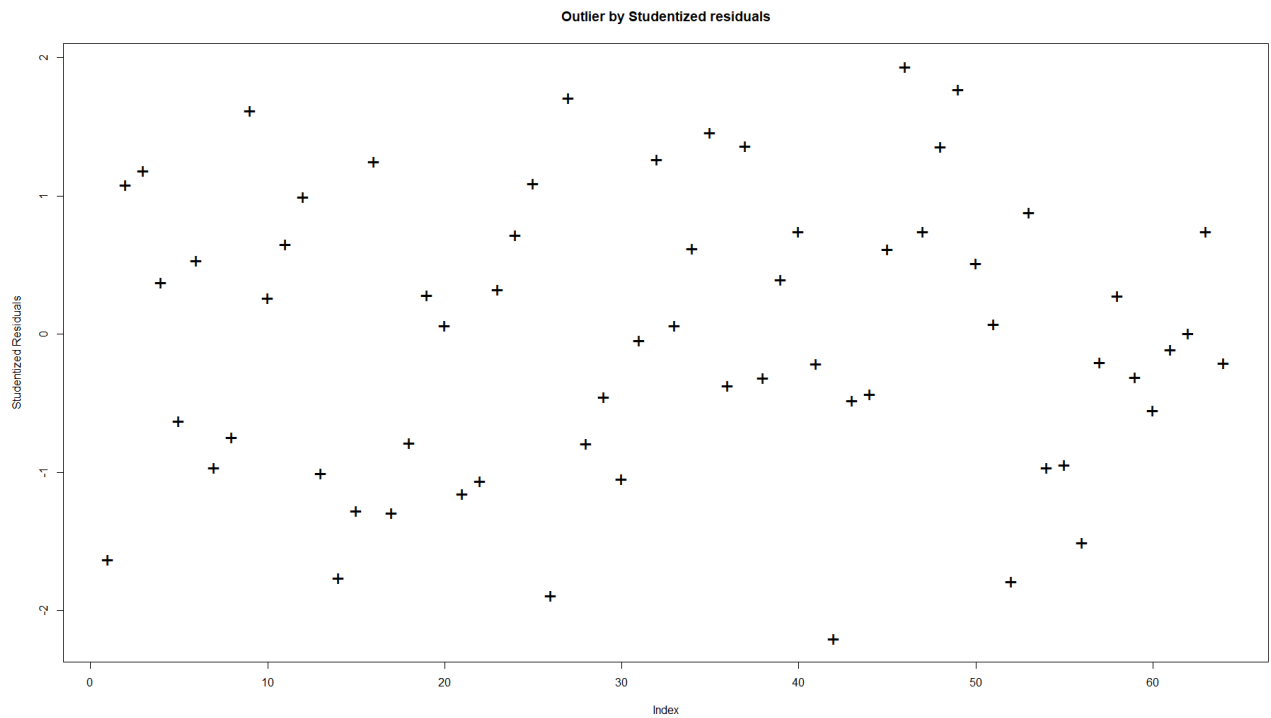
```
df4 = df3[-c(40,41,35),]  
attach(df4)  
lillie.test(df4$y)
```

```
> lillie.test(df4$y)  
  
Lilliefors (Kolmogorov-Smirnov) normality test  
  
data: df4$y  
D = 0.14139, p-value = 0.00281
```

Değerler veri setinden çıkarılmıştır. Yeni veri setinin p-value değeri  $0.00281 < \alpha = 0.05$  olduğu için veri setinin normal dağılım göstermediği 0.05 anlamlılık ile söylenmektedir. Artık incelemesine devam edilmektedir.

## Beşinci Artık İncelemesi





```
df5 = df4[-c(43,27,57,42),]
attach(df5)
lillie.test(df5$y)
```



```
> lillie.test(df5$y)
```

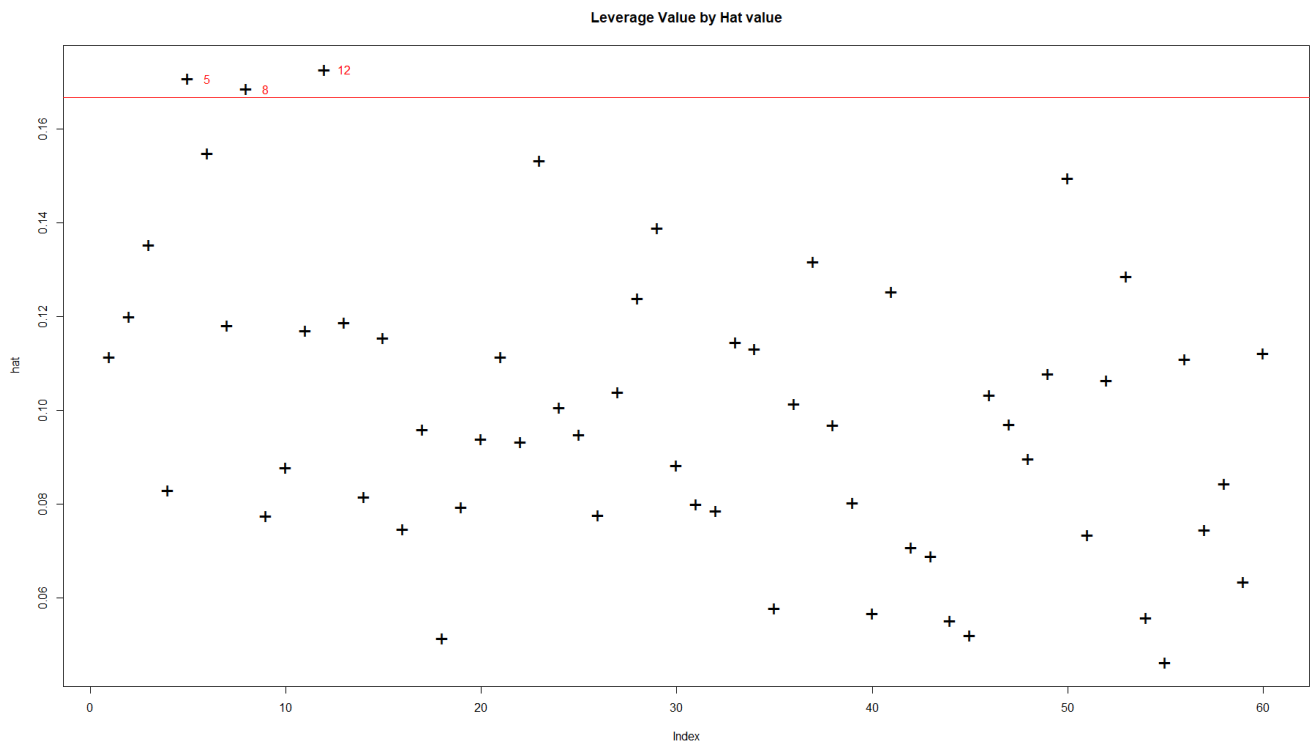
Lilliefors (Kolmogorov-Smirnov) normality test

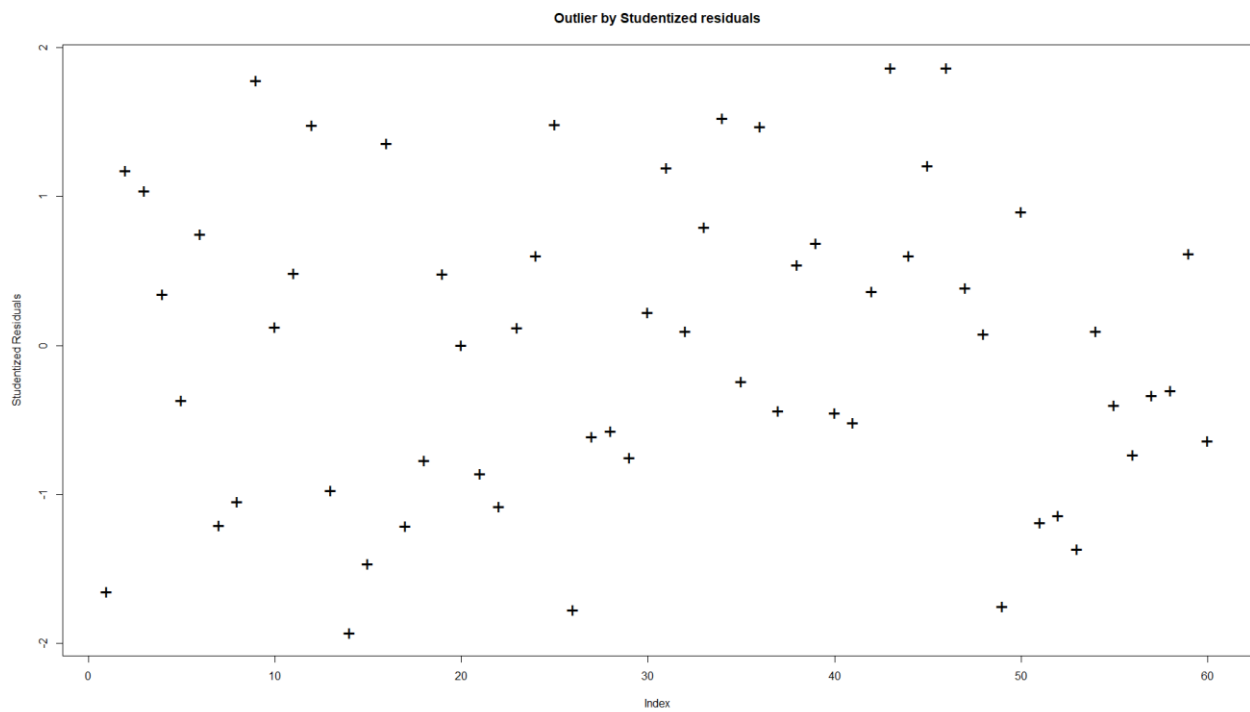
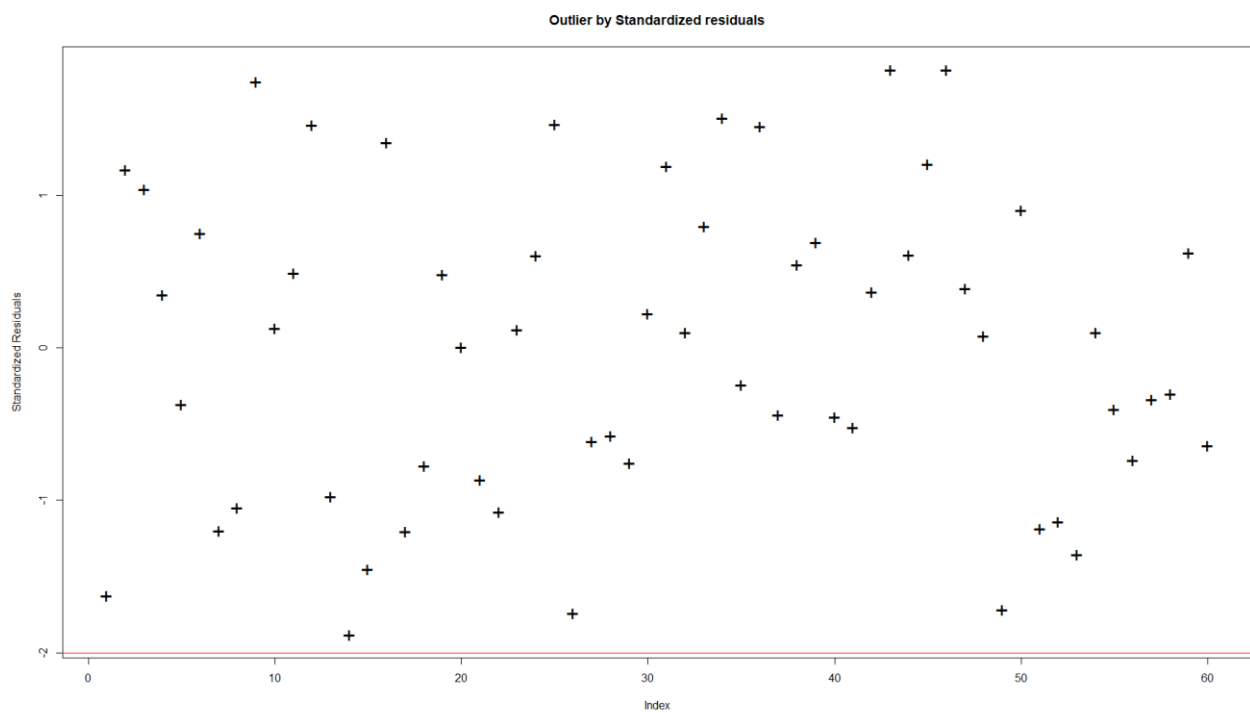
data: df5\$y

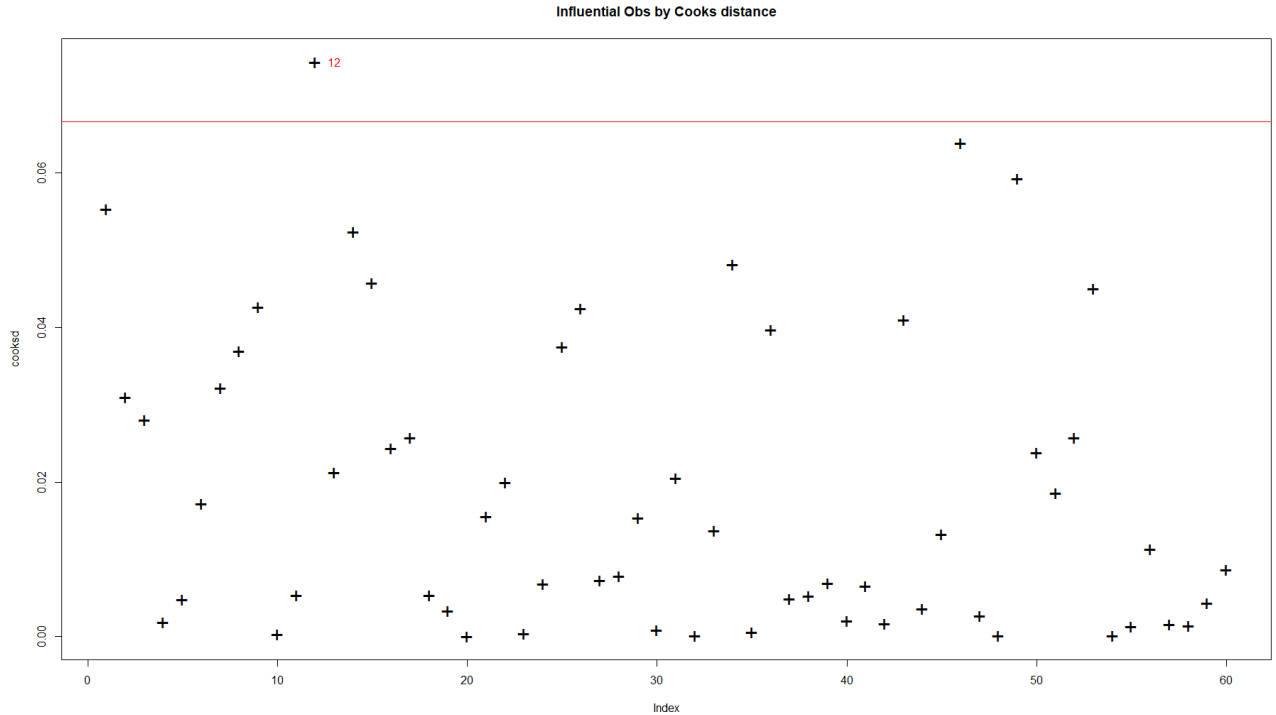
D = 0.13462, p-value = 0.008641

Değerler veri setinden çıkarılmıştır. Yeni veri setinin p-value değeri  $0.008641 < \alpha = 0.05$  olduğu için veri setinin normal dağılım göstermediği  $0.05$  anlamlılık ile söylenmektedir. Artık incelemesine devam edilmektedir.

## Altıncı Artık İncelemesi





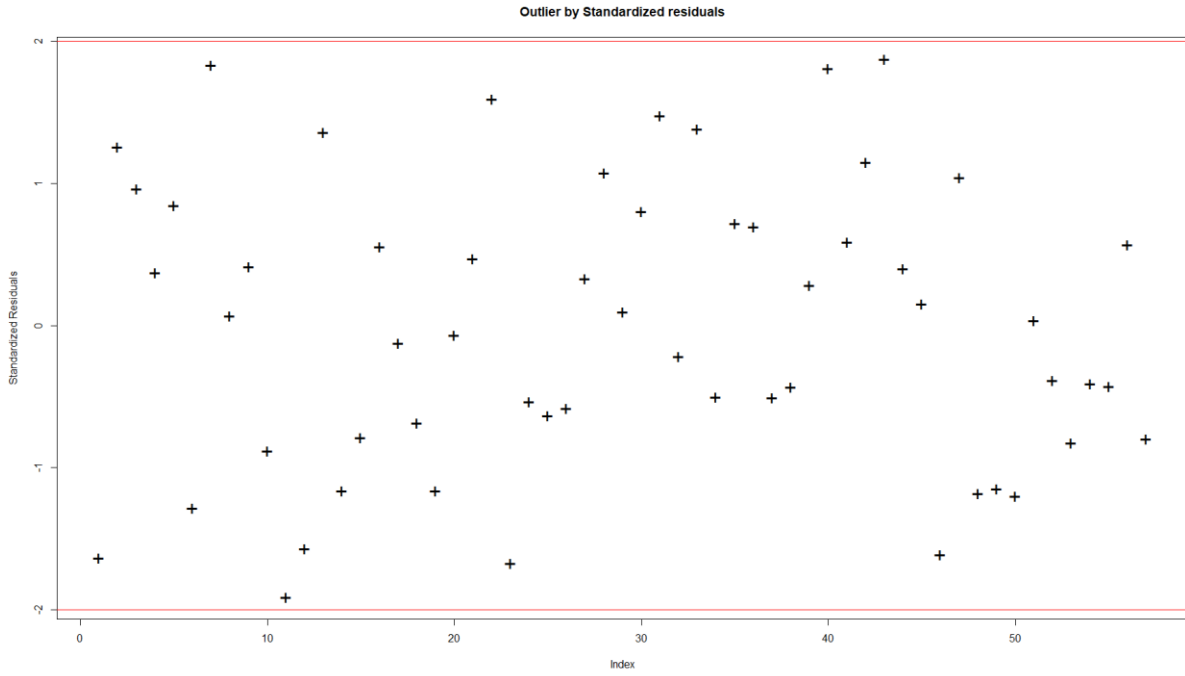
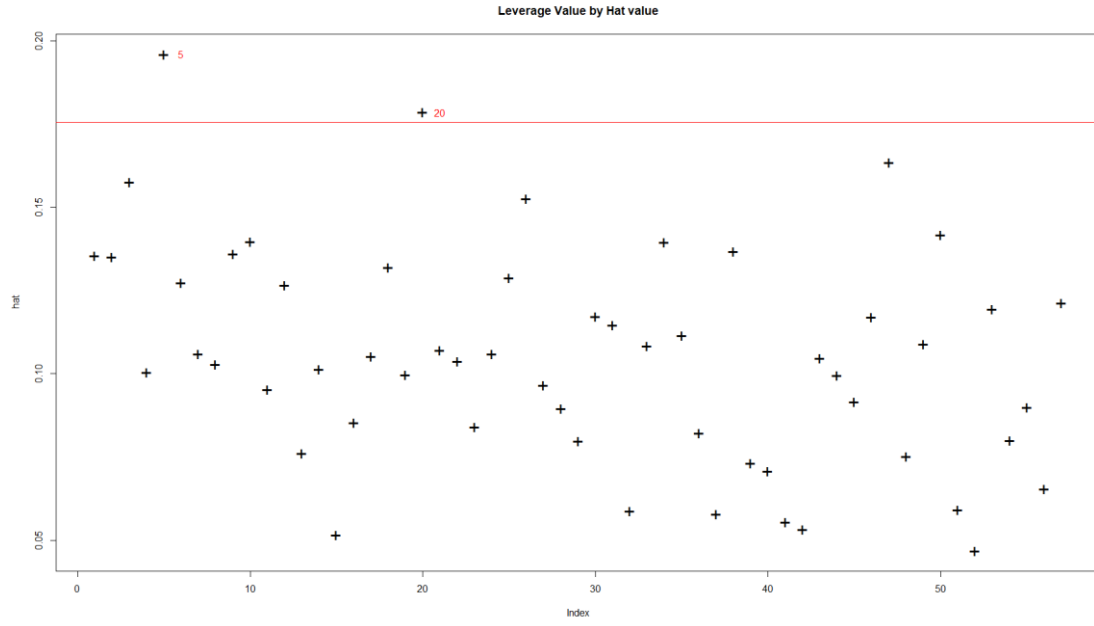


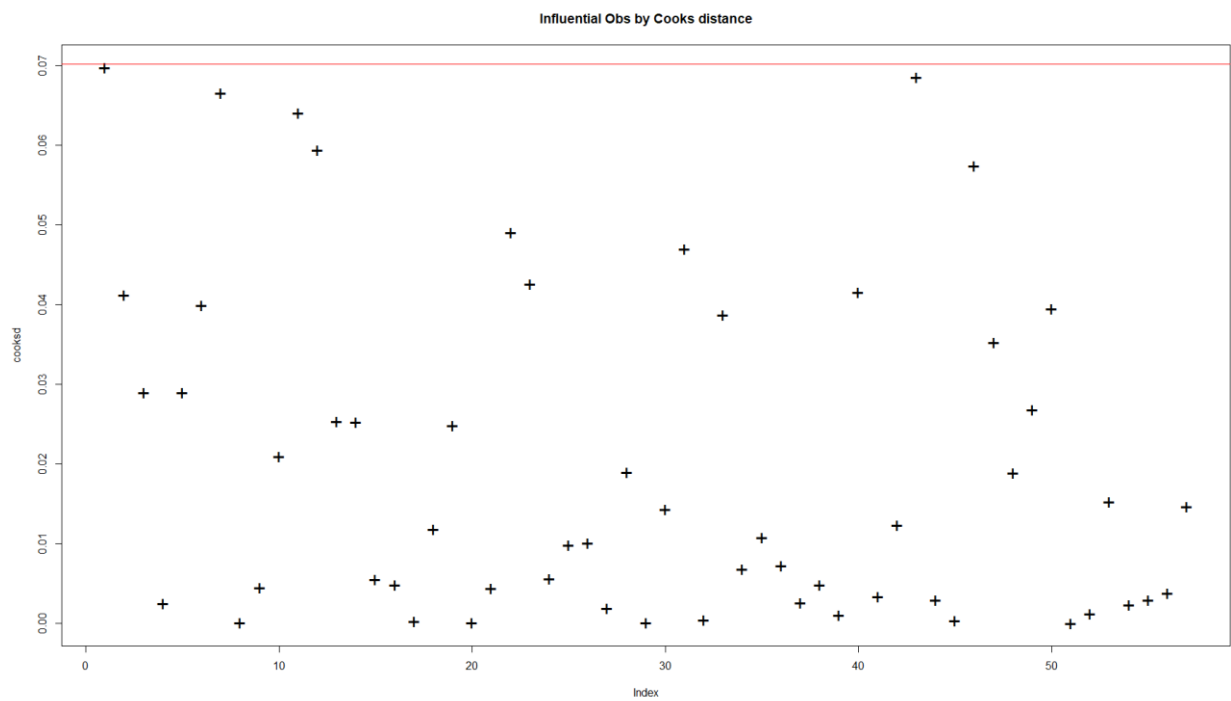
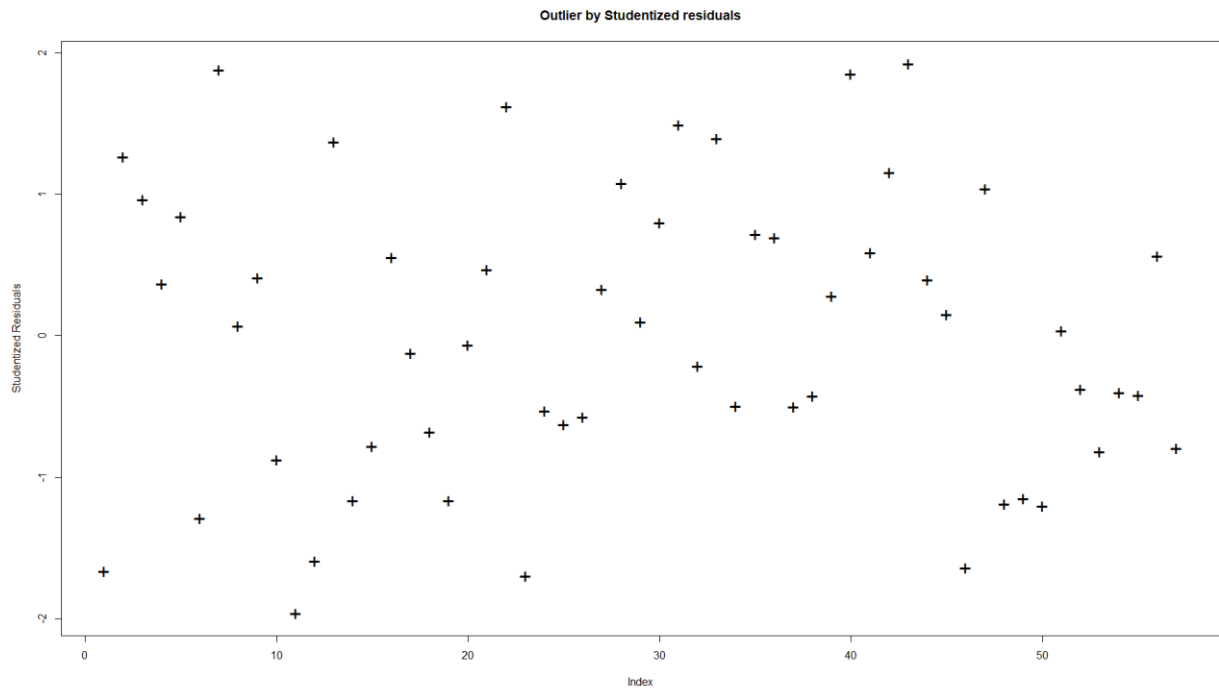
```
df6 = df5[-c(12,5,8),]  
attach(df6)  
lillie.test(df6$y)
```

```
> lillie.test(df6$y)  
  
Lilliefors (Kolmogorov-Smirnov) normality test  
  
data: df6$y  
D = 0.14501, p-value = 0.004426
```

Değerler veri setinden çıkarılmıştır. Yeni veri setinin p-value değeri  $0.004426 < \alpha = 0.05$  olduğu için veri setinin normal dağılım göstermediği 0.05 anlamlılık ile söylenmektedir. Artık incelemesine devam edilmektedir.

## Yedinci Artık İncelemesi



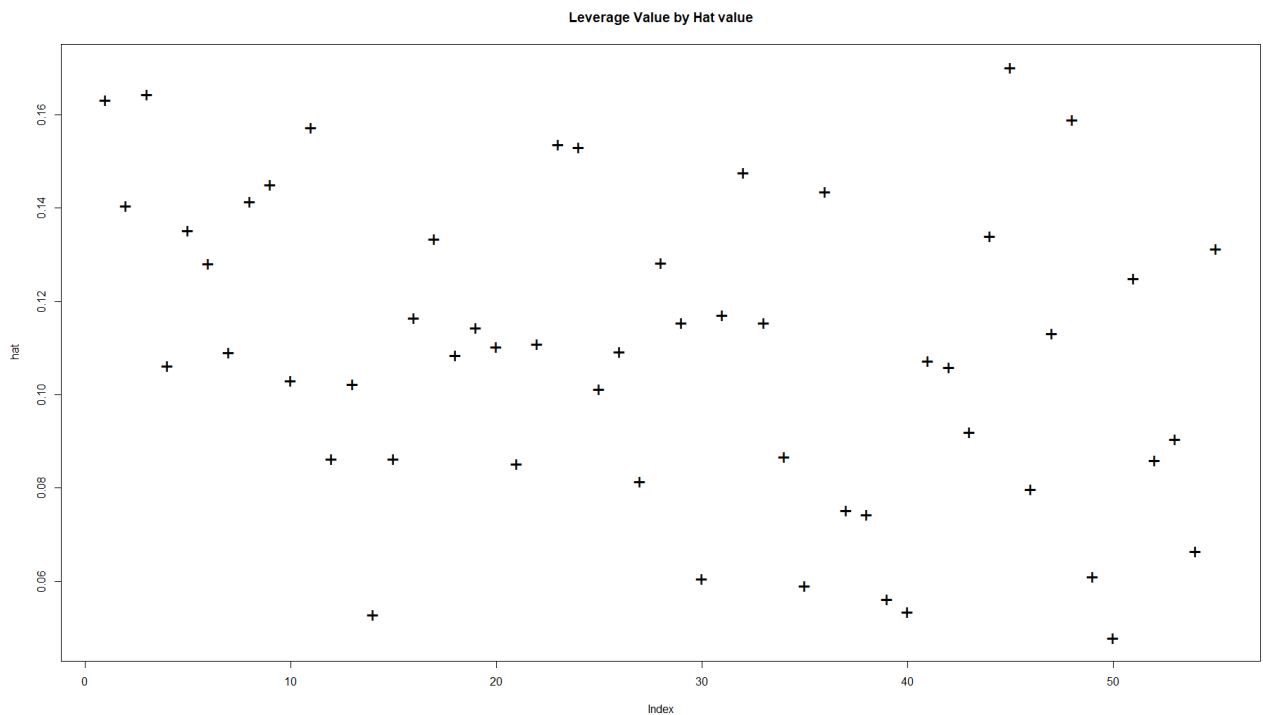


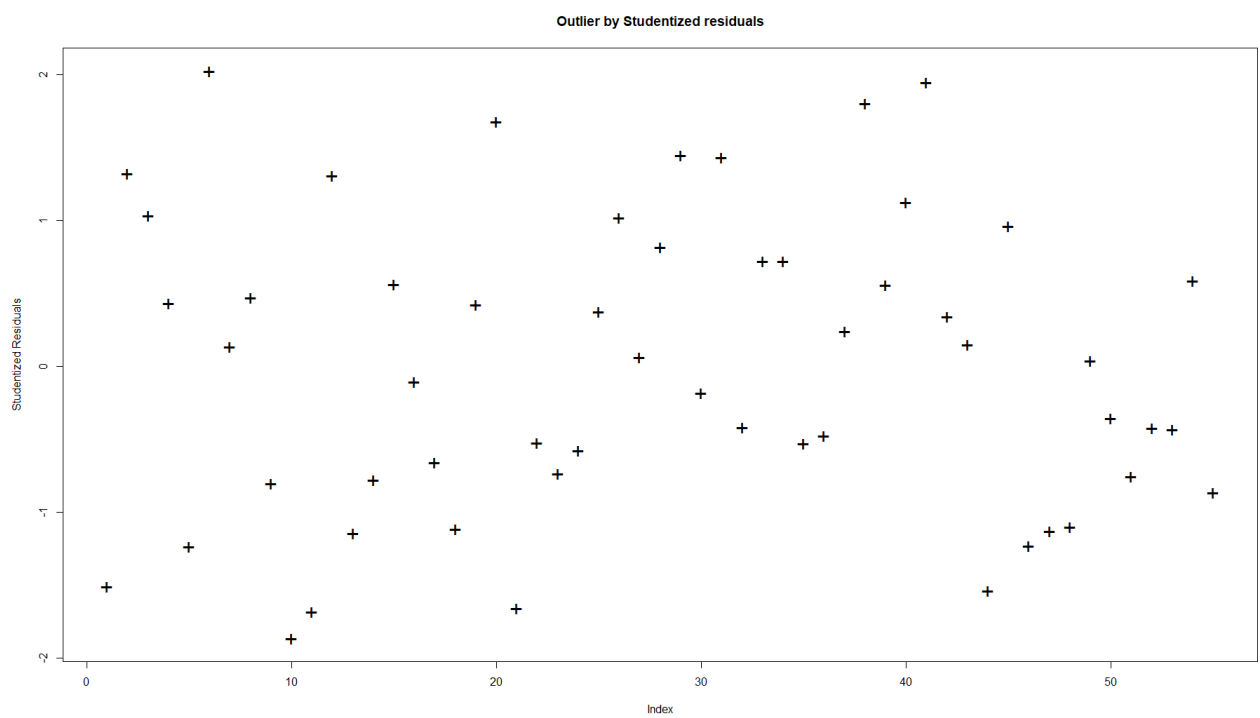
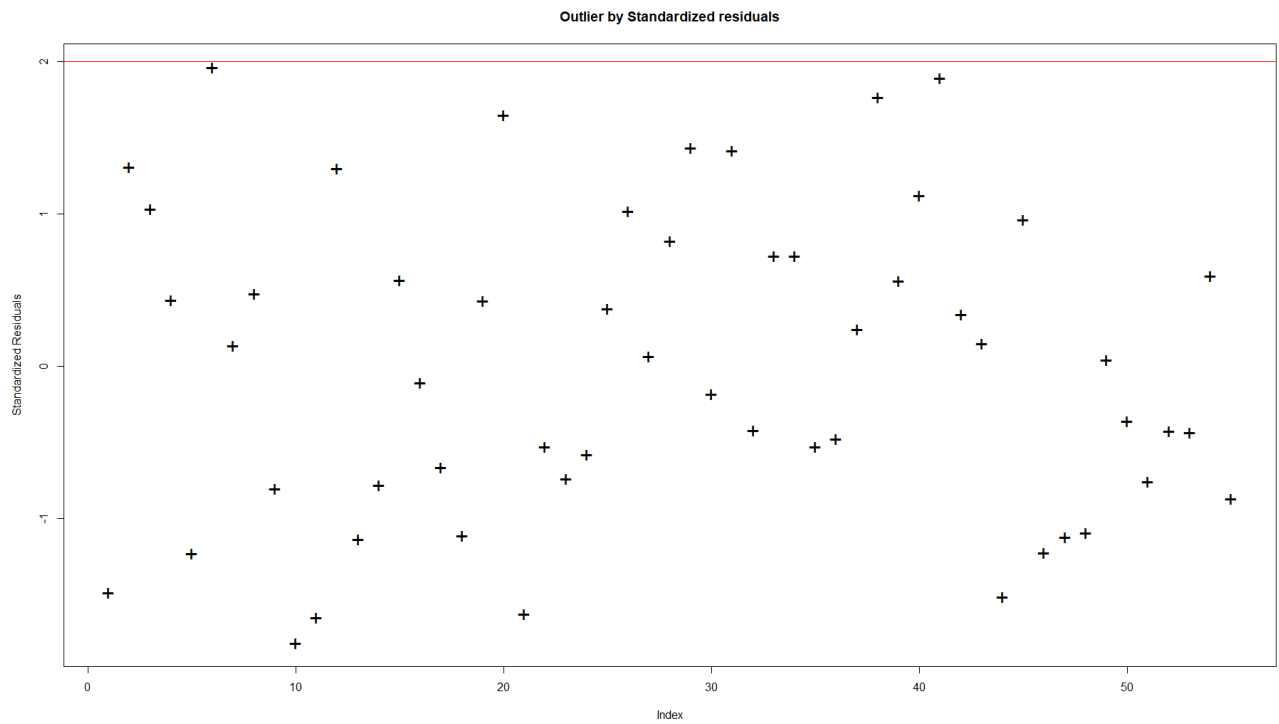
```
df7 = df6[-c(5,20),]  
attach(df7)  
lillie.test(df7$y)
```

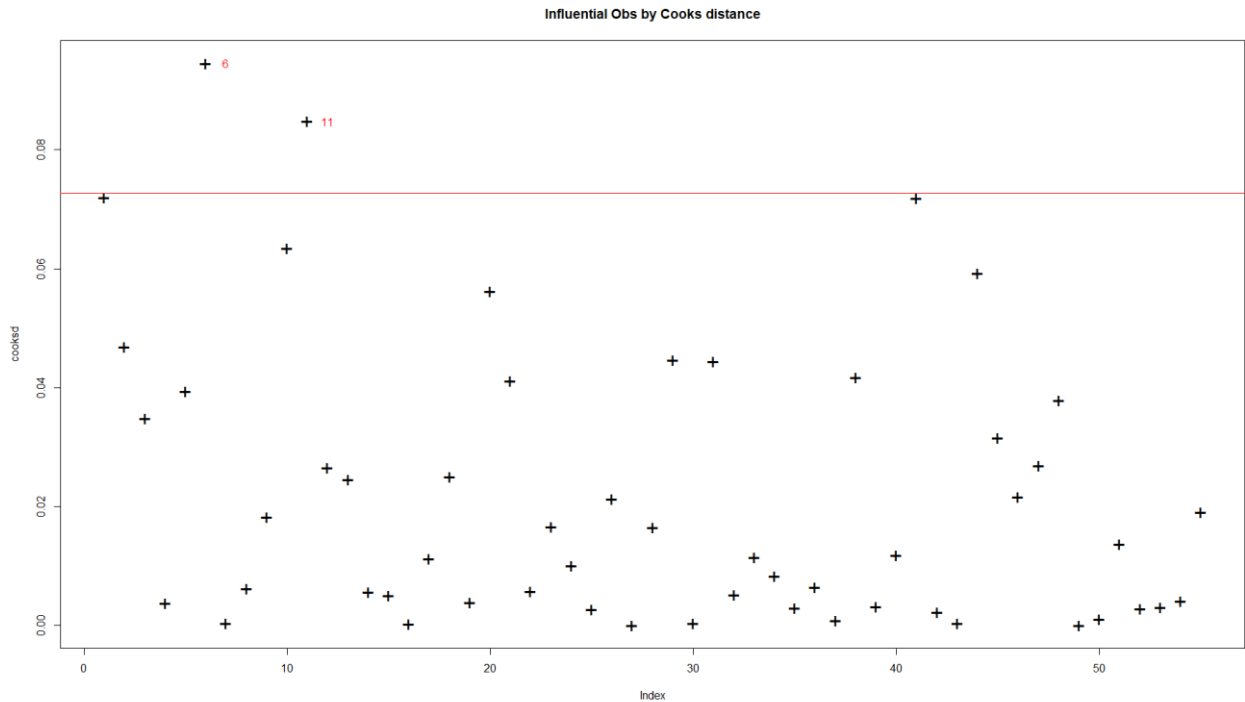
```
> lillie.test(df7$y)  
  
Lilliefors (Kolmogorov-Smirnov) normality test  
  
data: df7$y  
D = 0.1498, p-value = 0.003547
```

Değerler veri setinden çıkarılmıştır. Yeni veri setinin p-value değeri  $0.003547 < \alpha = 0.05$  olduğu için veri setinin normal dağılım göstermediği 0.05 anlamlılık ile söylenmektedir. Artık incelemesine devam edilmektedir.

## Sekizinci Artık İncelemesi







```
df8 = df7[-c(6,11),]  
attach(df8)  
lillie.test(df8$y)
```

```
> lillie.test(df8$y)
```

Lilliefors (Kolmogorov-Smirnov) normality test

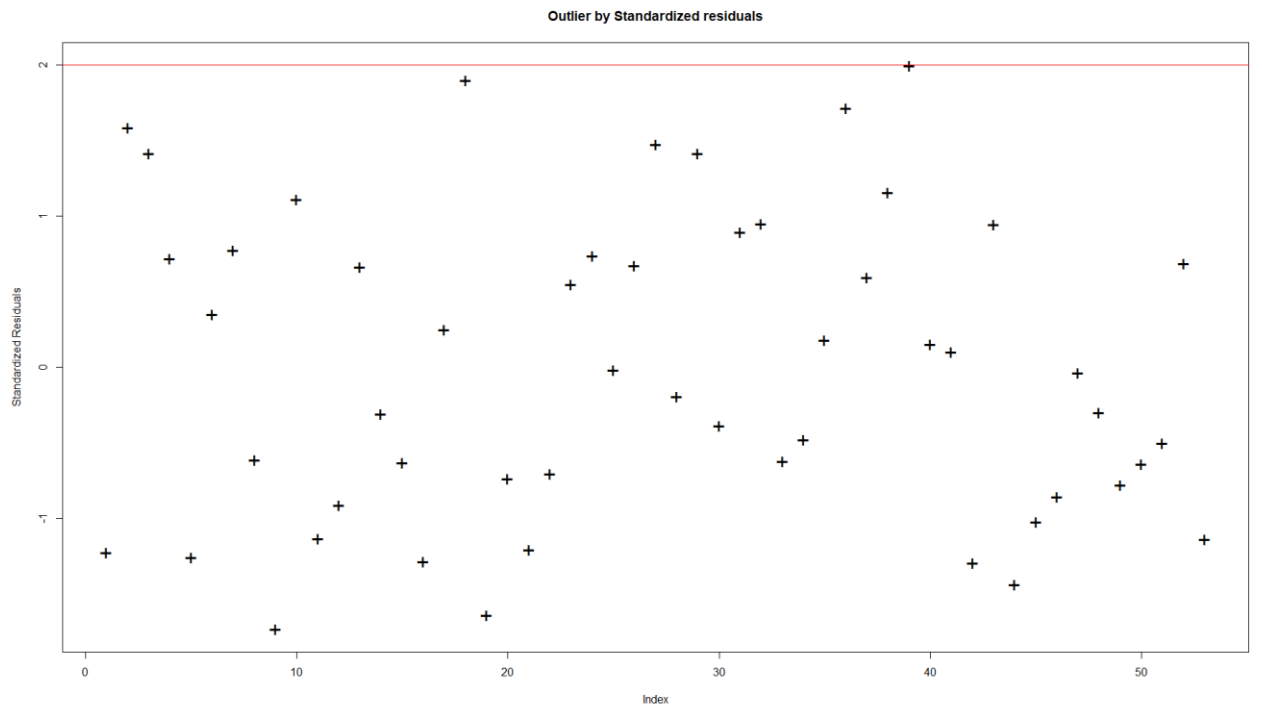
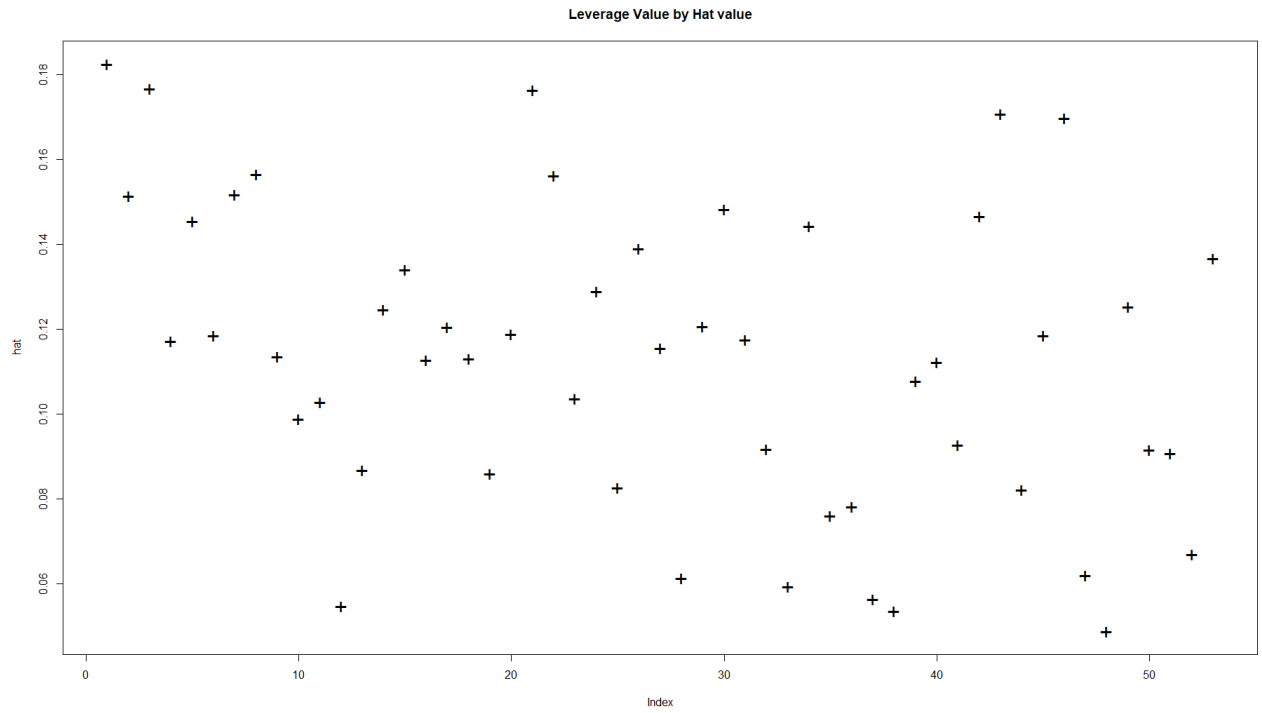
data: df8\$y

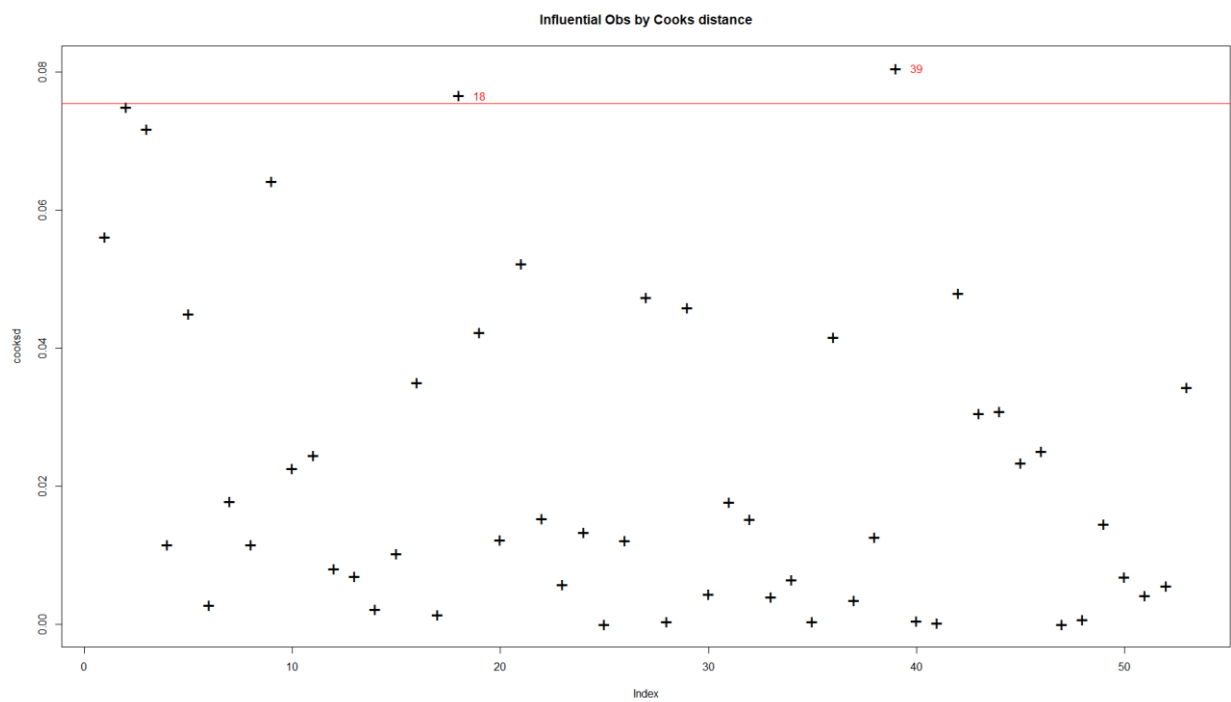
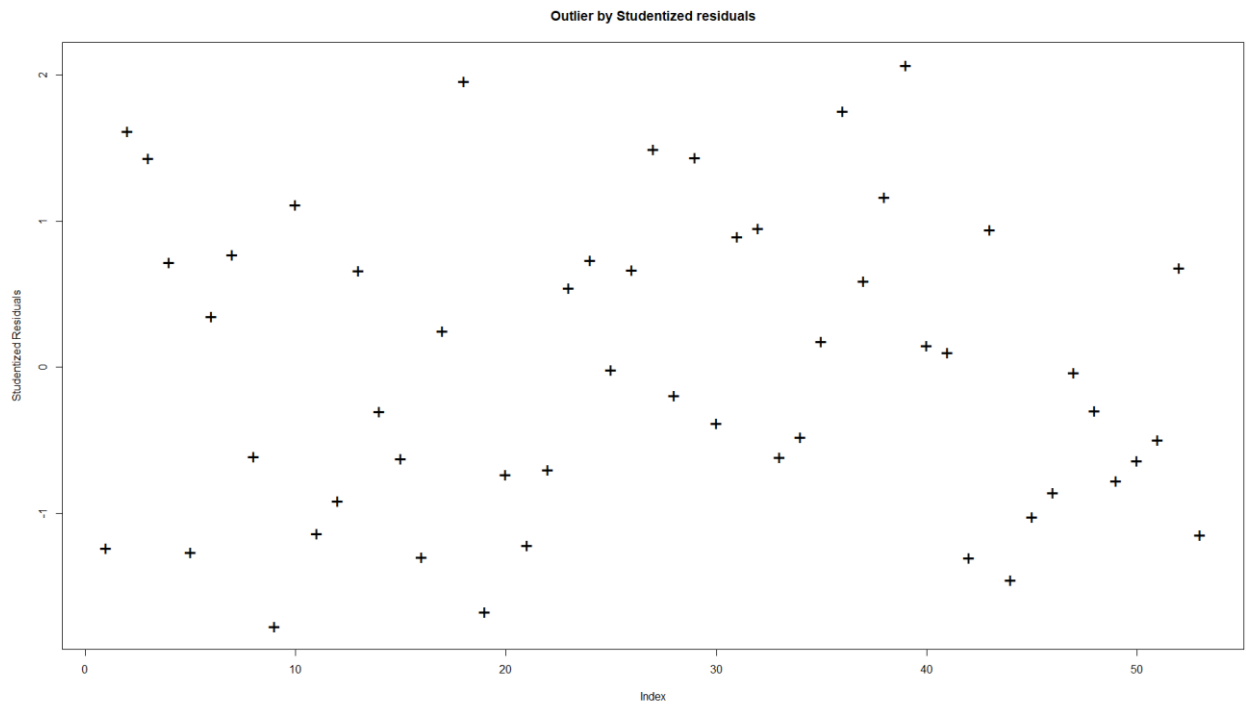
D = 0.15379, p-value = 0.00313

Değerler veri setinden çıkarılmıştır. Yeni veri setinin p-value değeri  $0.00313 < \alpha = 0.05$  olduğu için veri setinin normal dağılım göstermediği 0.05 anlamlılık ile söylenmektedir. Artık incelemesine devam edilmektedir.



## Dokuzuncu Artık İncelemesi





```
df9 = df8[-c(39,18),]
attach(df9)
lillie.test(df9$y)
```

```
> lillie.test(df9$y)
```

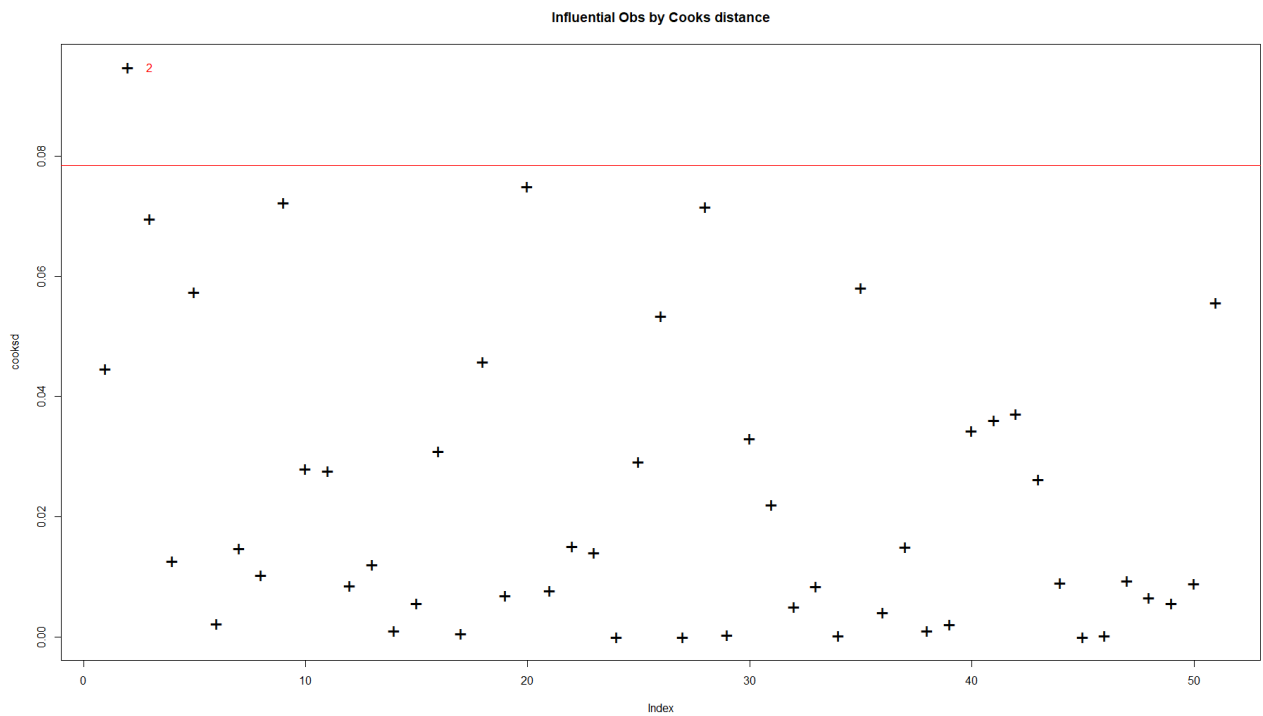
Lilliefors (Kolmogorov-Smirnov) normality test

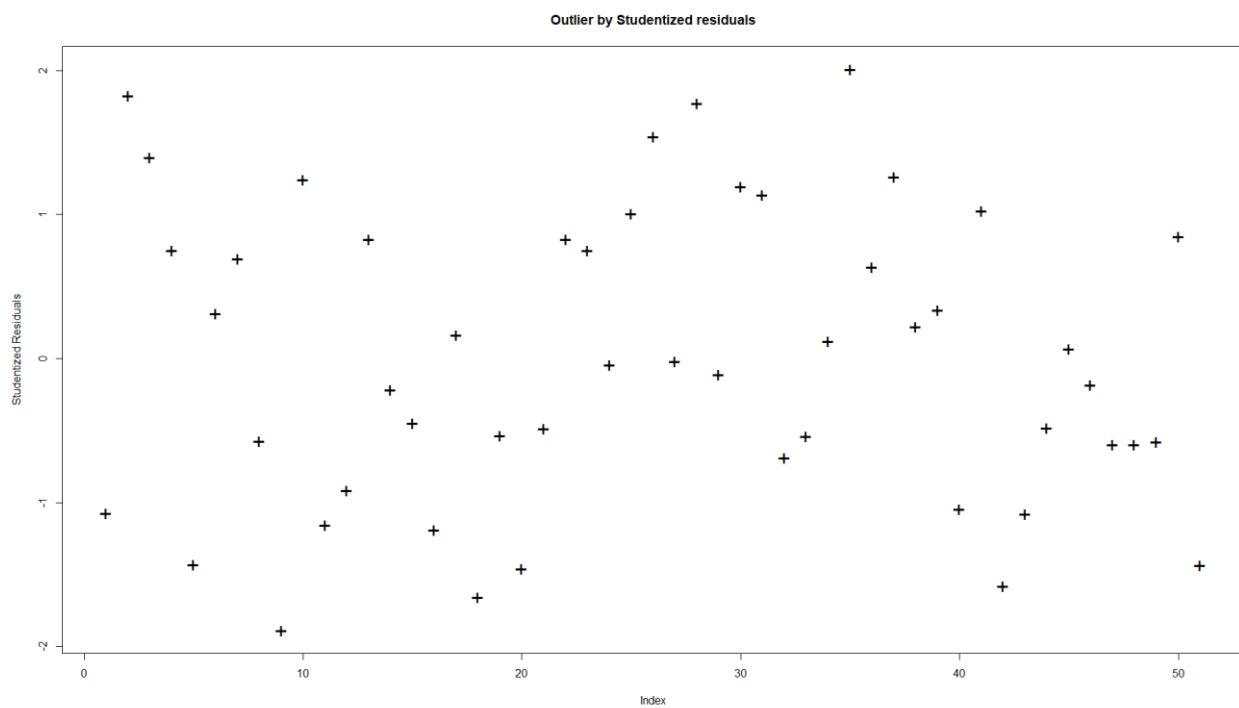
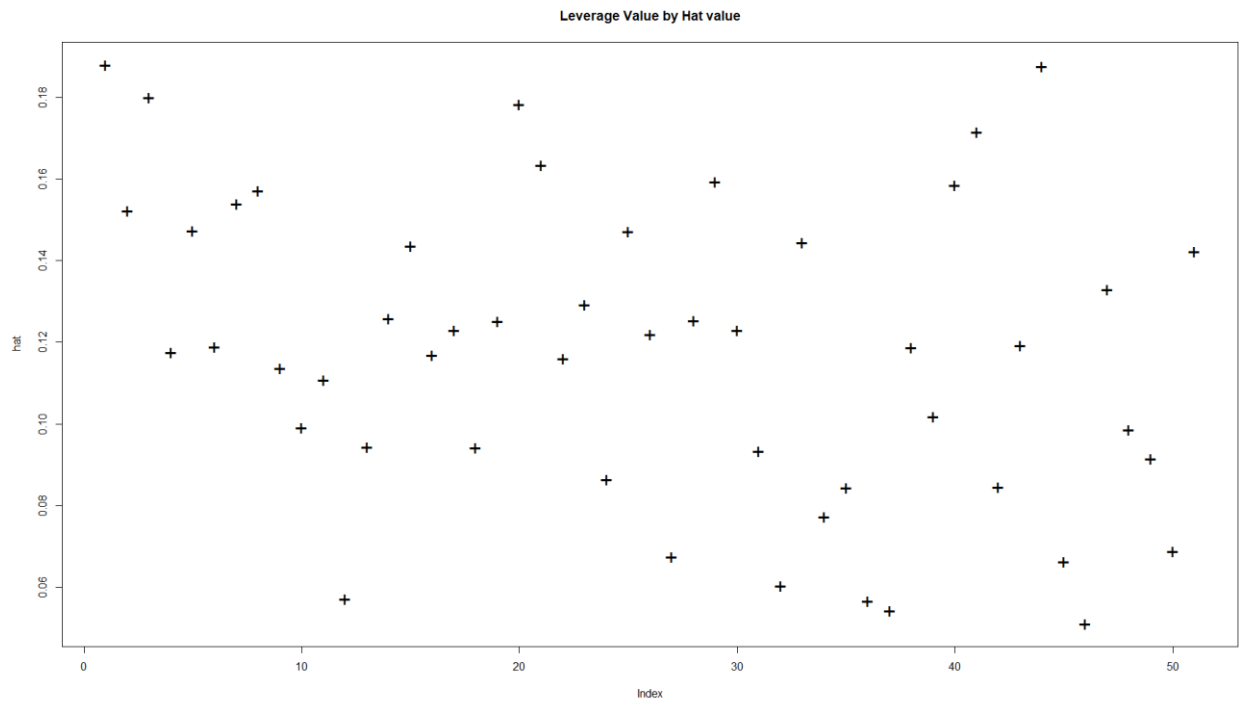
data: df9\$y

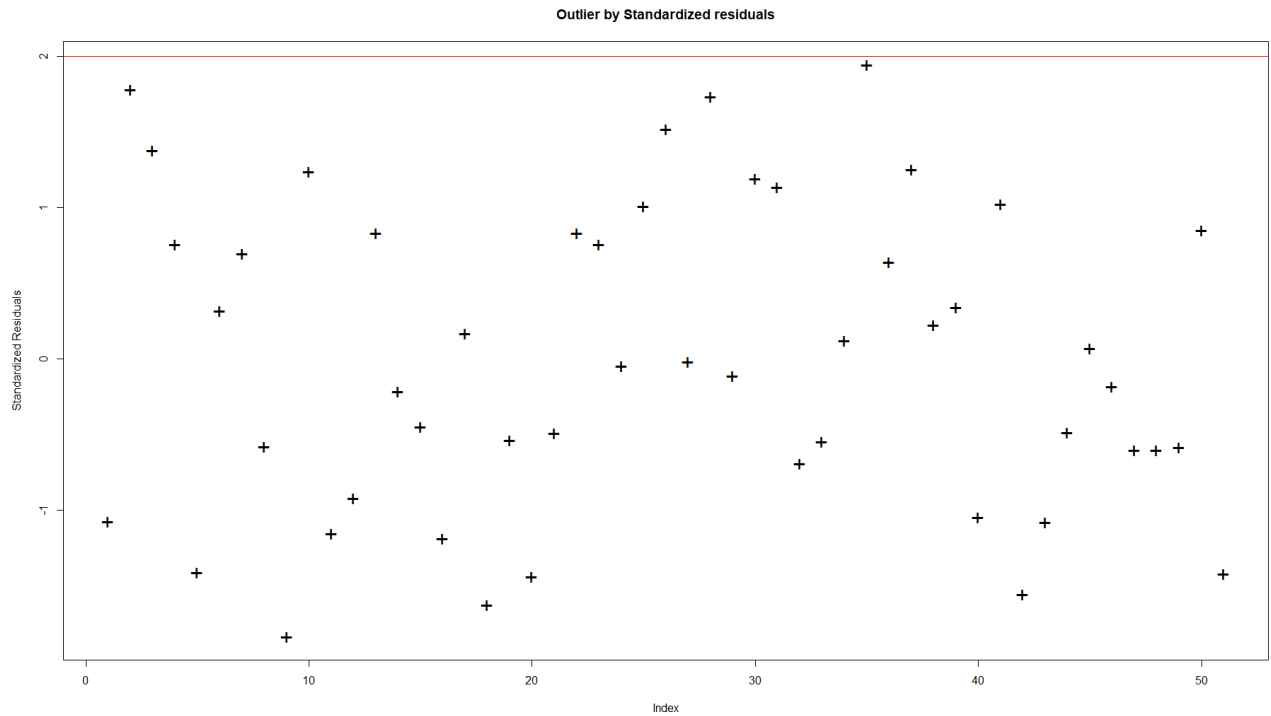
D = 0.15229, p-value = 0.004733

Değerler veri setinden çıkarılmıştır. Yeni veri setinin p-value değeri  $0.004733 < \alpha = 0.05$  olduğu için veri setinin normal dağılım göstermediği 0.05 anlamlılık ile söylenmektedir. Artık incelemesine devam edilmektedir.

## Onuncu Artık İncelemesi





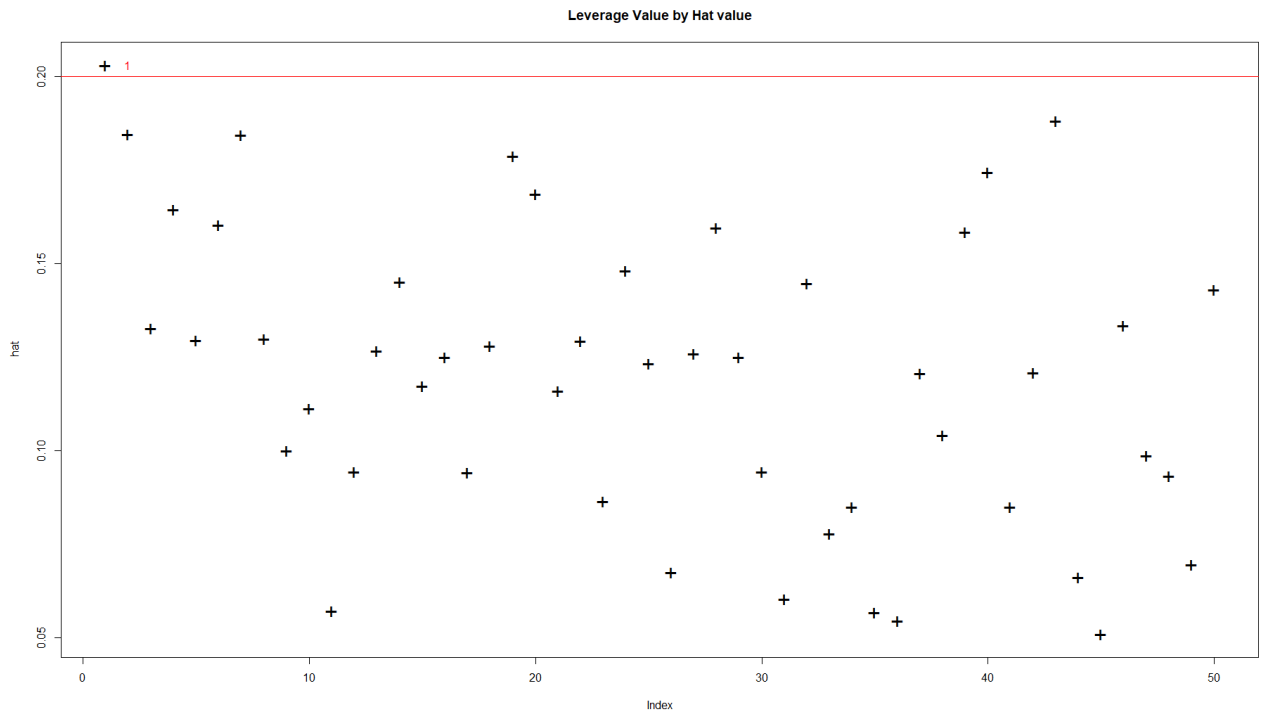
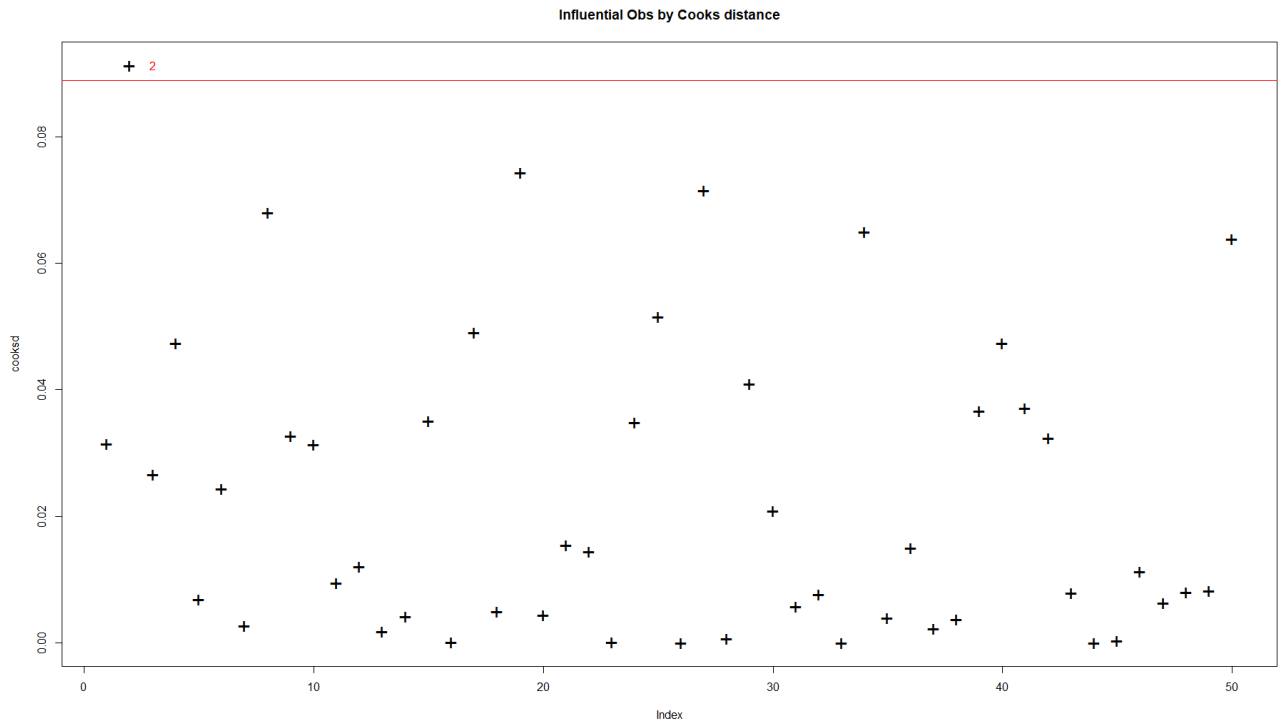


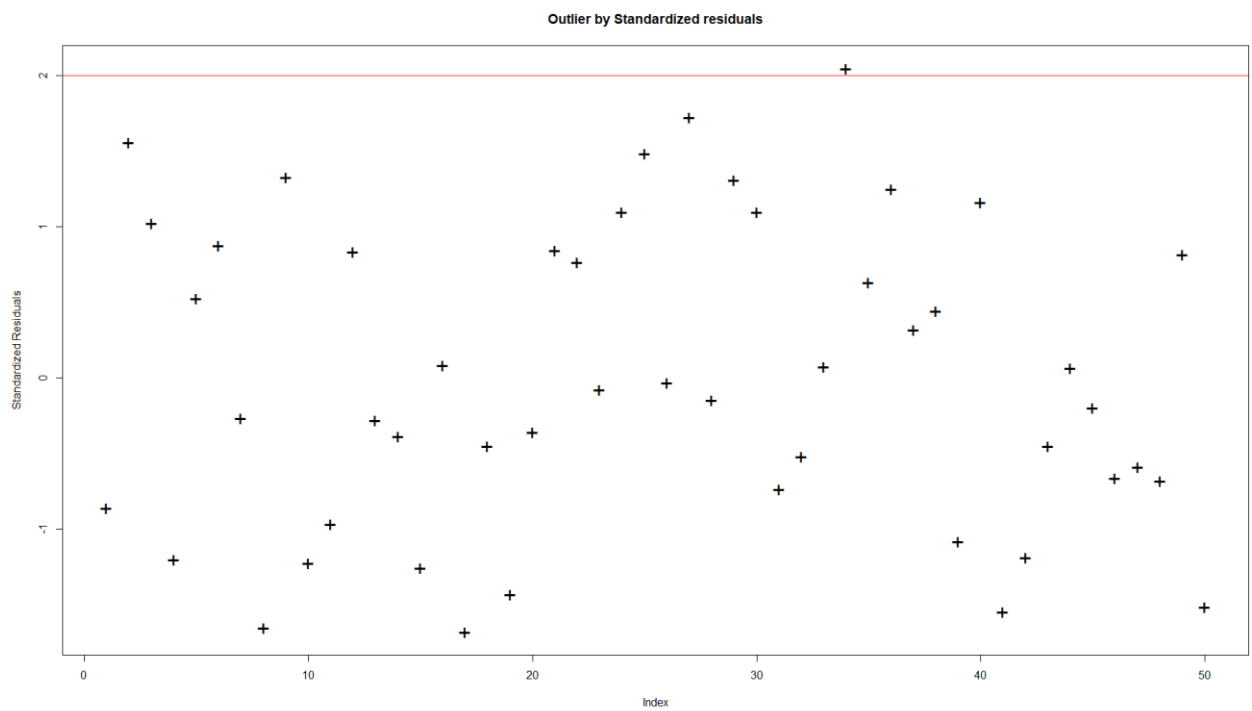
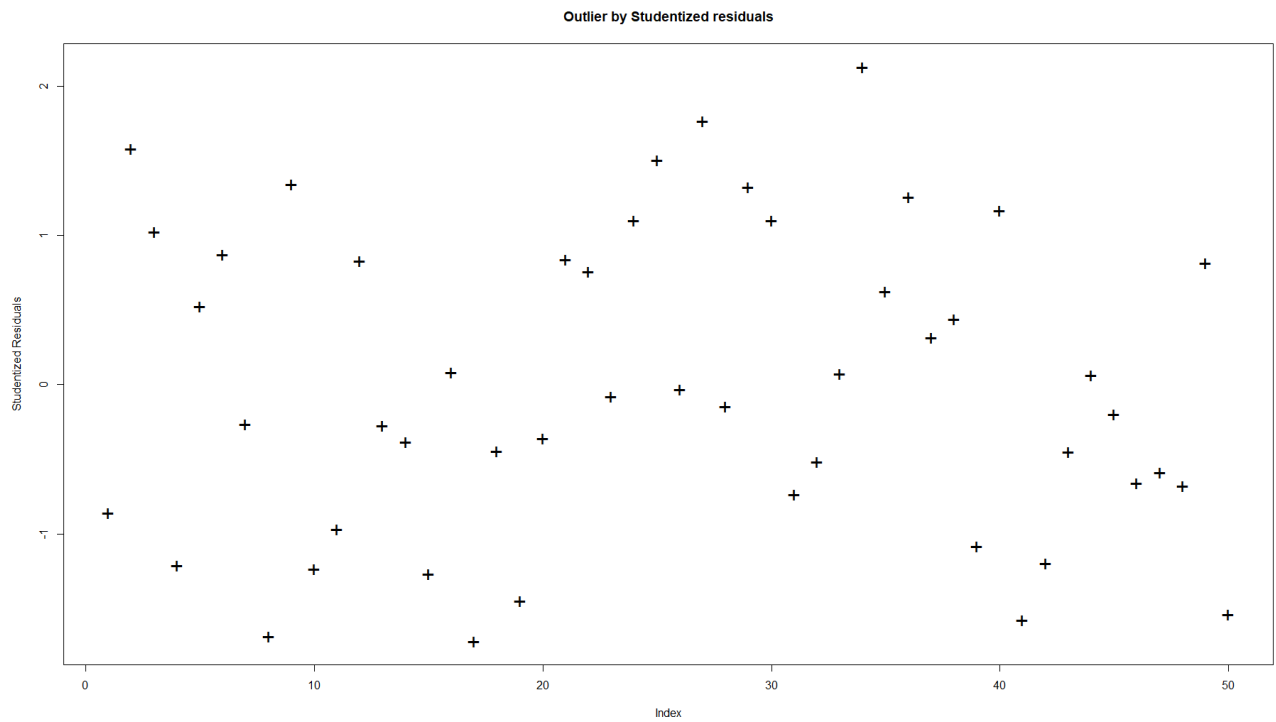
```
df10 = df9[-c(2),]  
attach(df10)  
lillie.test(df10$y)
```

```
> lillie.test(df10$y)  
  
Lilliefors (Kolmogorov-Smirnov) normality test  
  
data: df10$y  
D = 0.15179, p-value = 0.005655
```

Değerler veri setinden çıkarılmıştır. Yeni veri setinin p-value değeri  $0.005655 < \alpha = 0.05$  olduğu için veri setinin normal dağılım göstermediği 0.05 anlamlılık ile söylenmektedir. Artık incelemesine devam edilmektedir.

## On Birinci Artık İncelemesi



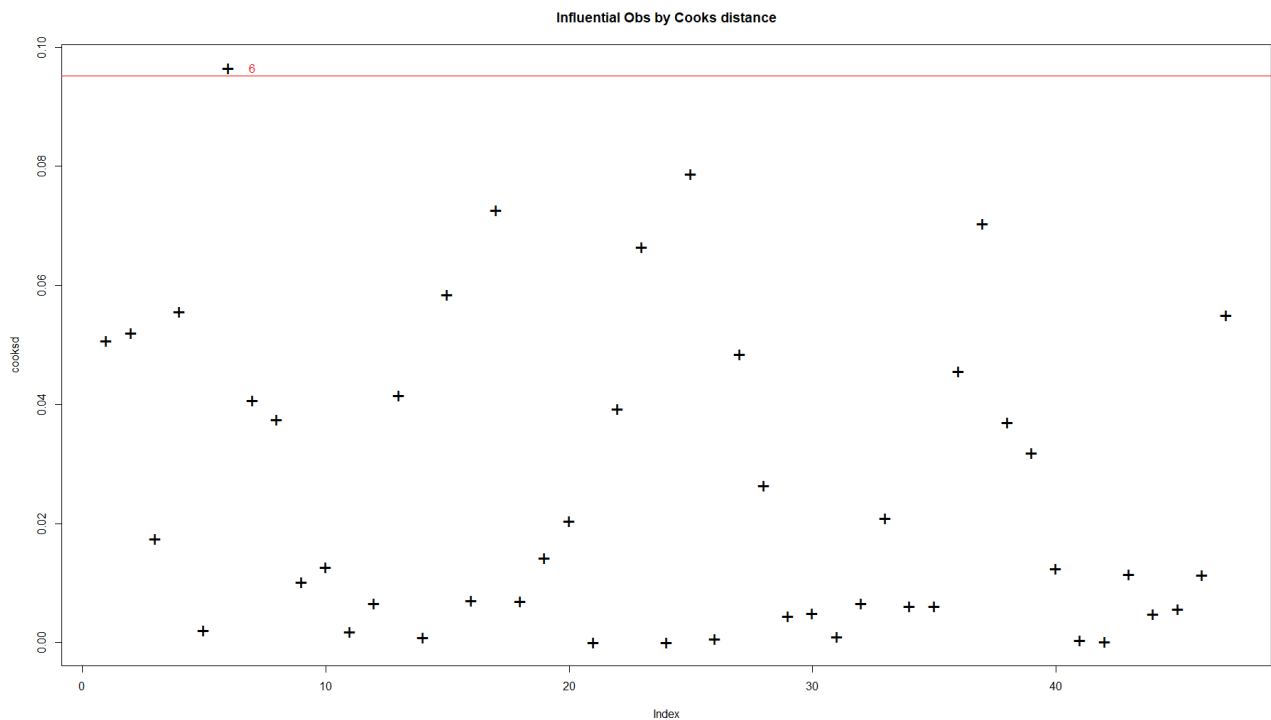


```
df11 <- df10[-c(1,2,34),]  
attach(df11)  
lillie.test(df11$y)
```

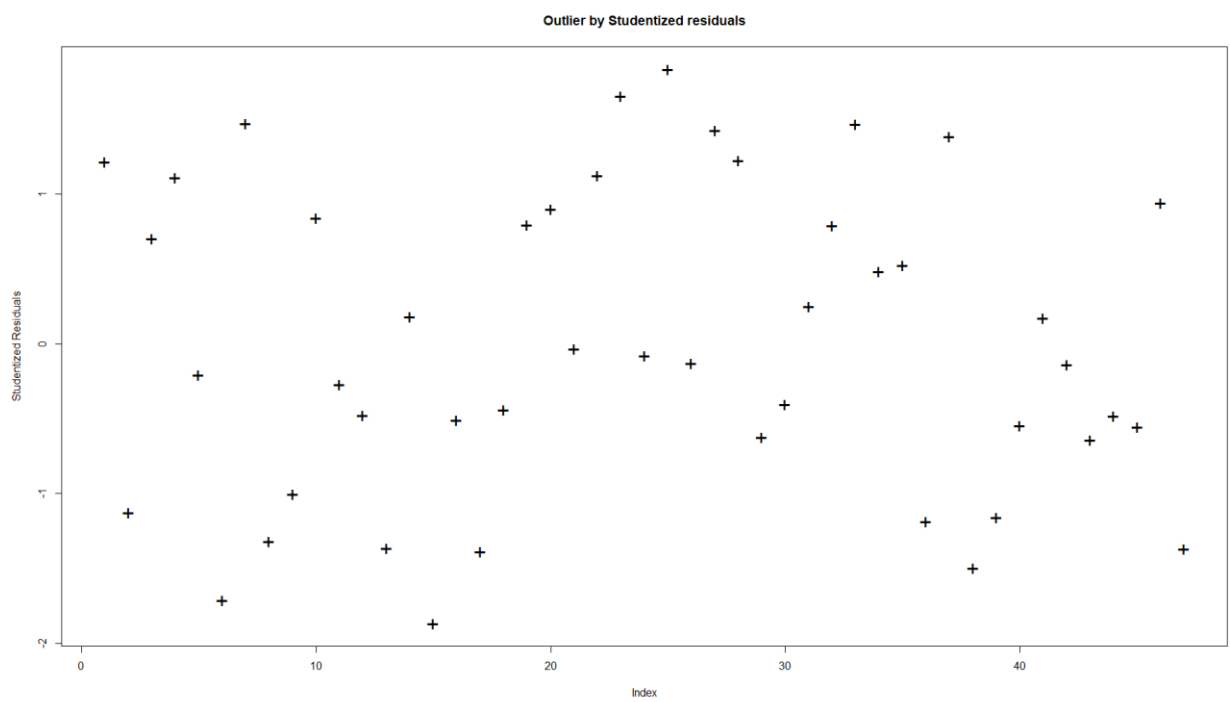
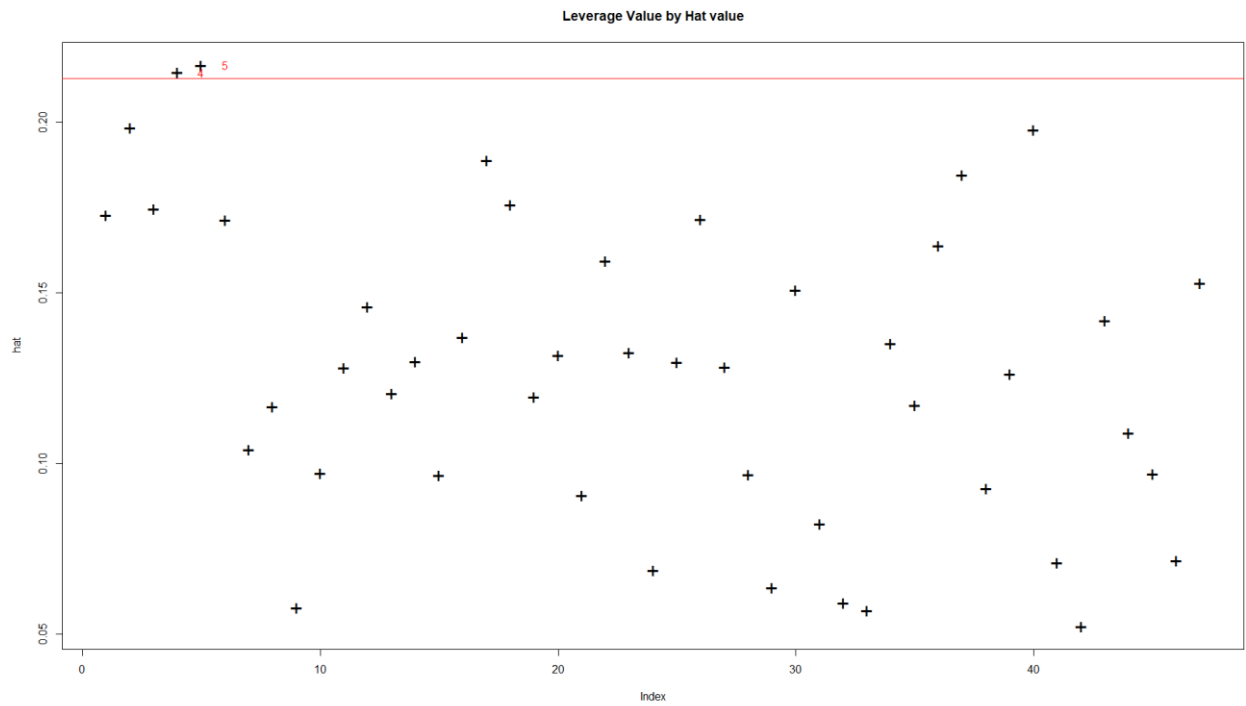
```
> lillie.test(df11$y)  
  
Lilliefors (Kolmogorov-Smirnov) normality test  
  
data: df11$y  
D = 0.16386, p-value = 0.002869
```

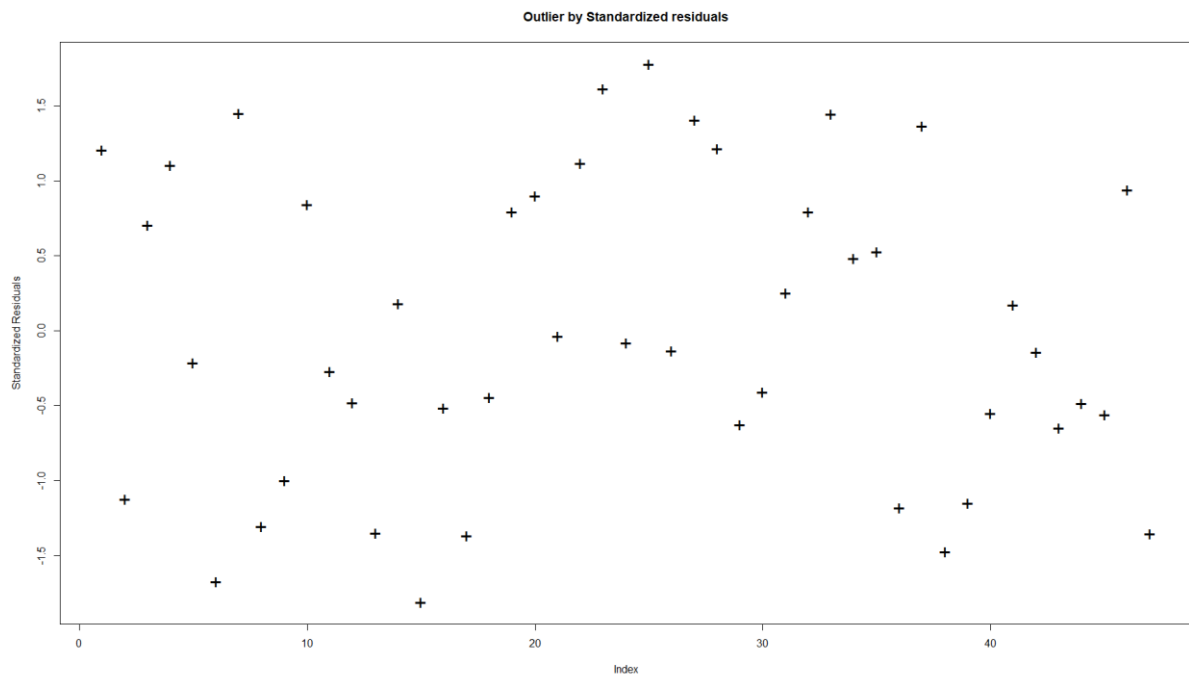
Değerler veri setinden çıkarılmıştır. Yeni veri setinin p-value değeri  $0.002869 < \alpha = 0.05$  olduğu için veri setinin normal dağılım göstermediği 0.05 anlamlılık ile söylenmektedir. Artık incelemesine devam edilmektedir.

## On İkinci Artık İncelemesi









```
df12 <- df11[-c(5,4,6),]  
attach(df12)  
lillie.test(df12$y)
```

```
> lillie.test(df12$y)
```

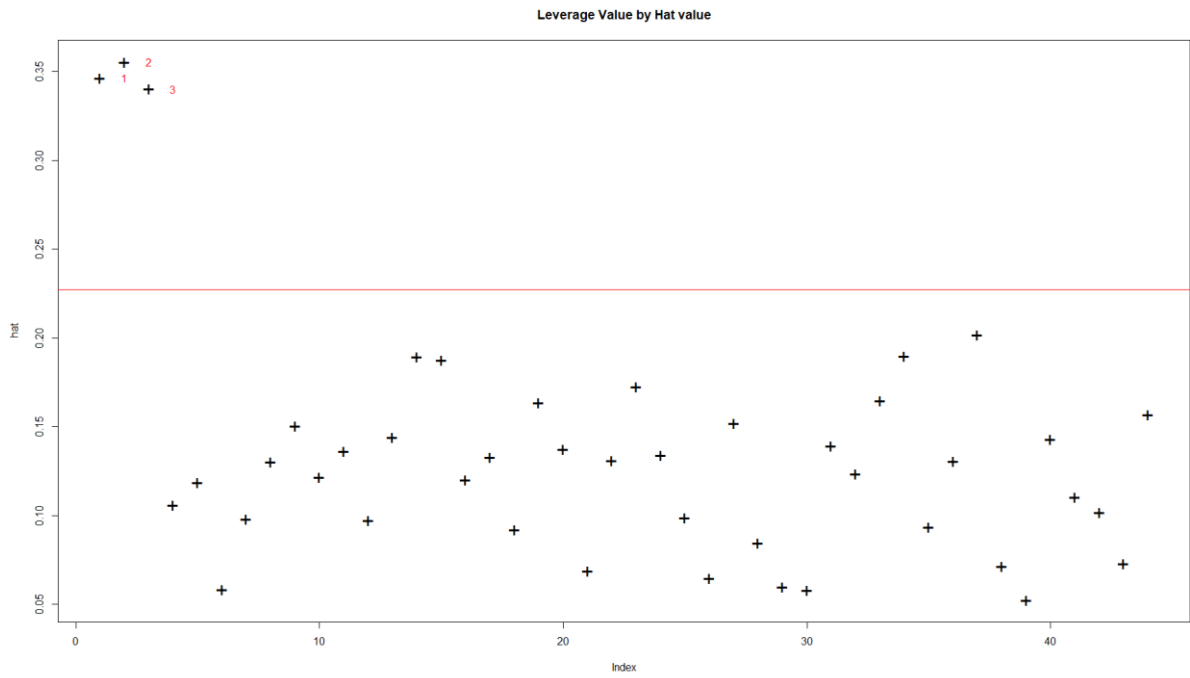
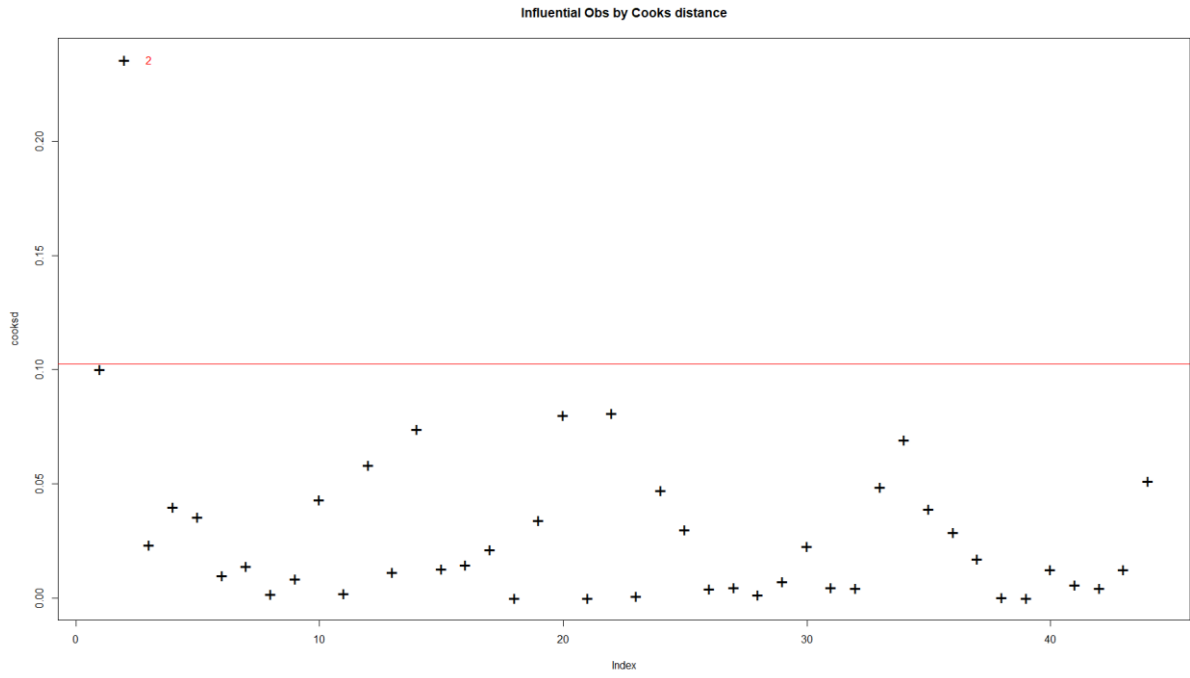
Lilliefors (Kolmogorov-Smirnov) normality test

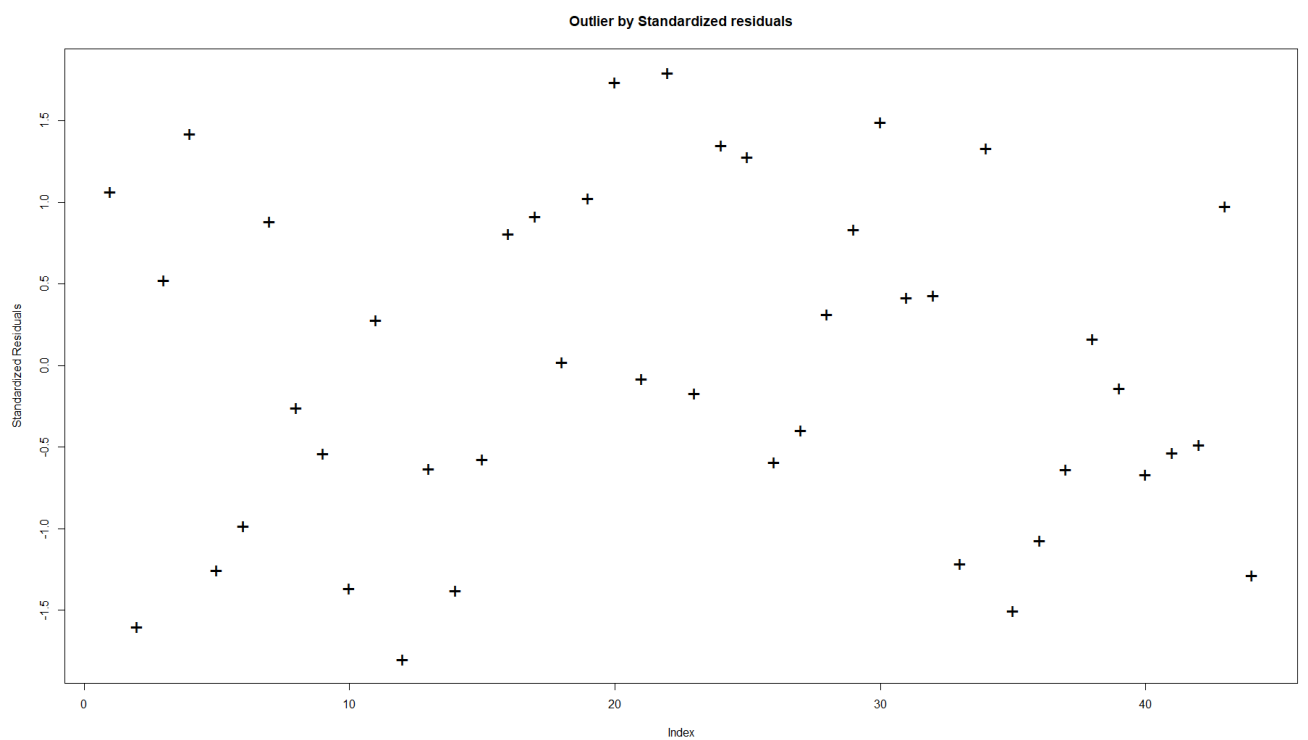
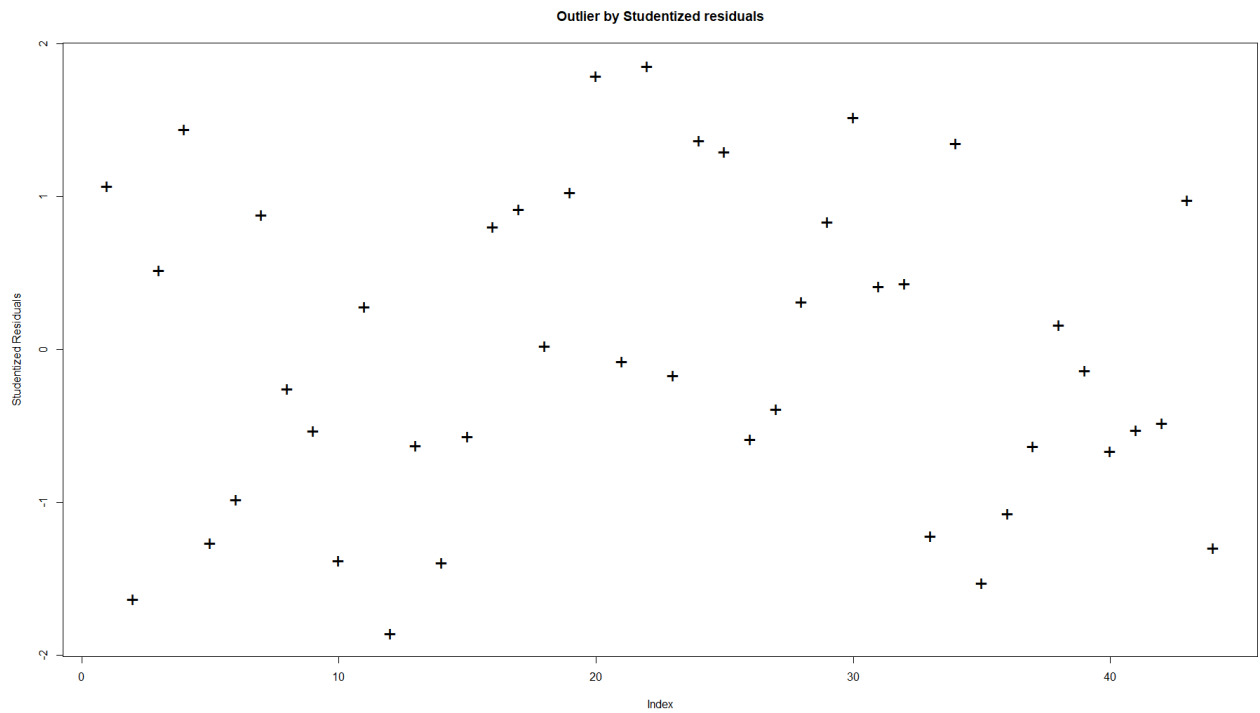
data: df12\$y

D = 0.16809, p-value = 0.003145

Değerler veri setinden çıkarılmıştır. Yeni veri setinin p-value değeri  $0.003145 < \alpha = 0.05$  olduğu için veri setinin normal dağılım göstermediği 0.05 anlamlılık ile söylenmektedir. Artık incelemesine devam edilmektedir.

## On Üçüncü Artık İncelemesi





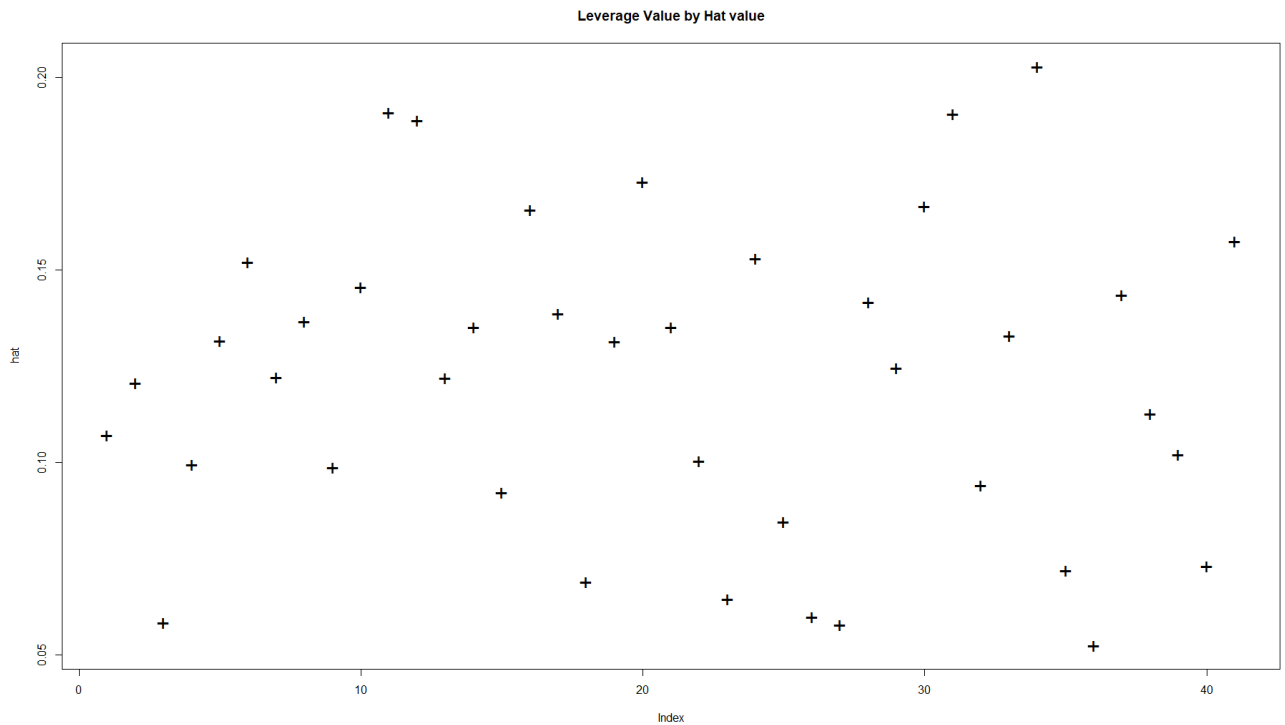
```
df13 <- df12[-c(1,2,3),]  
attach(df13)  
lillie.test(df13$y)
```

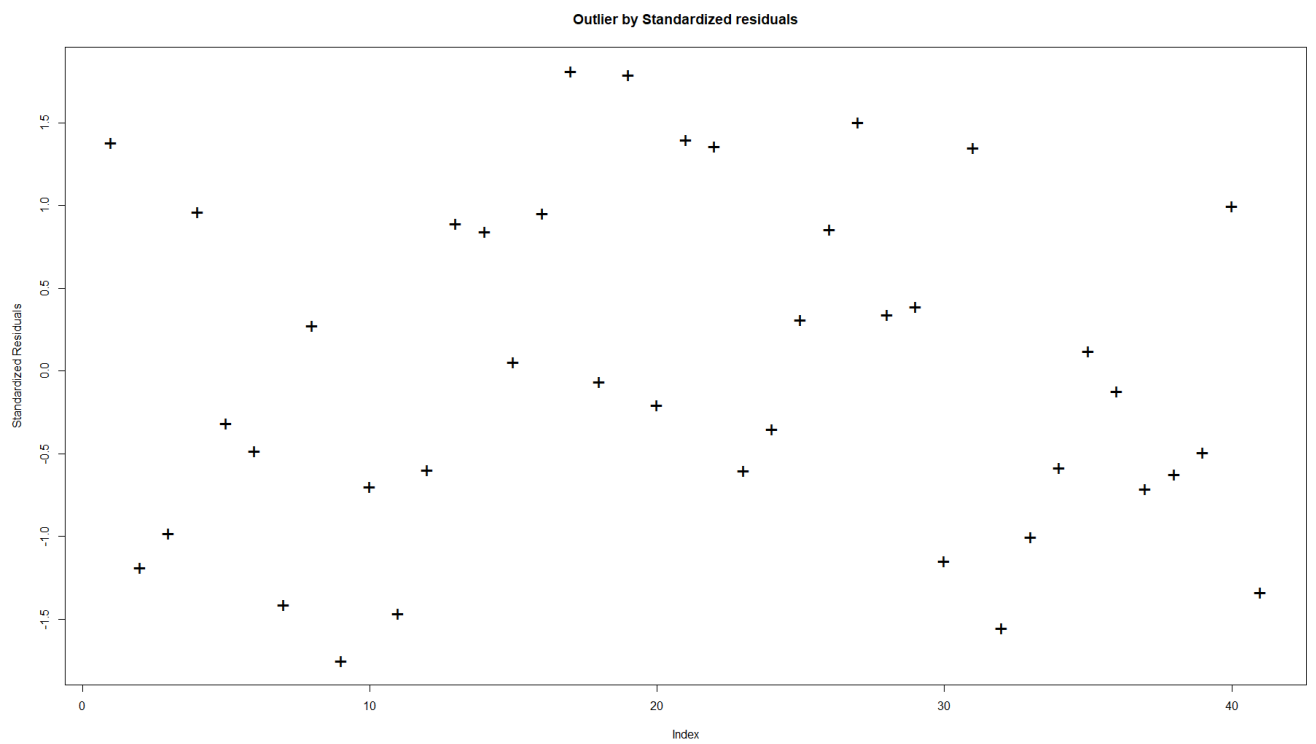
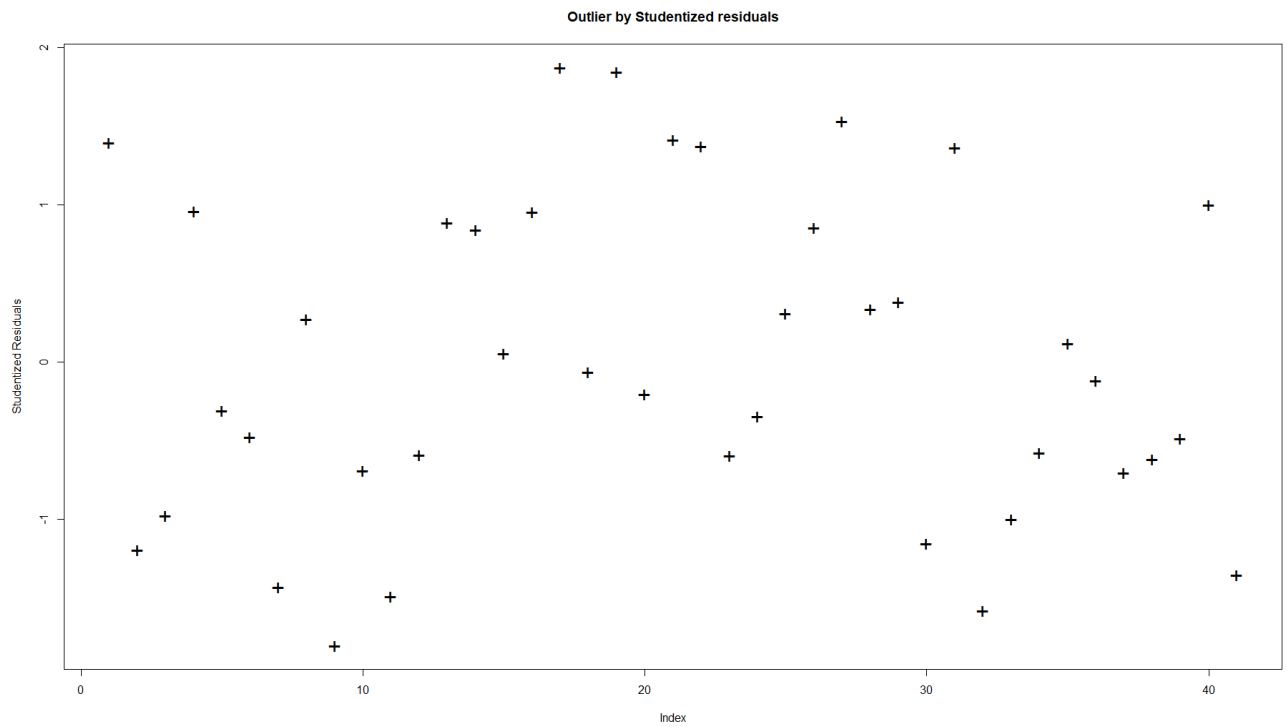
Lilliefors (Kolmogorov-Smirnov) normality test

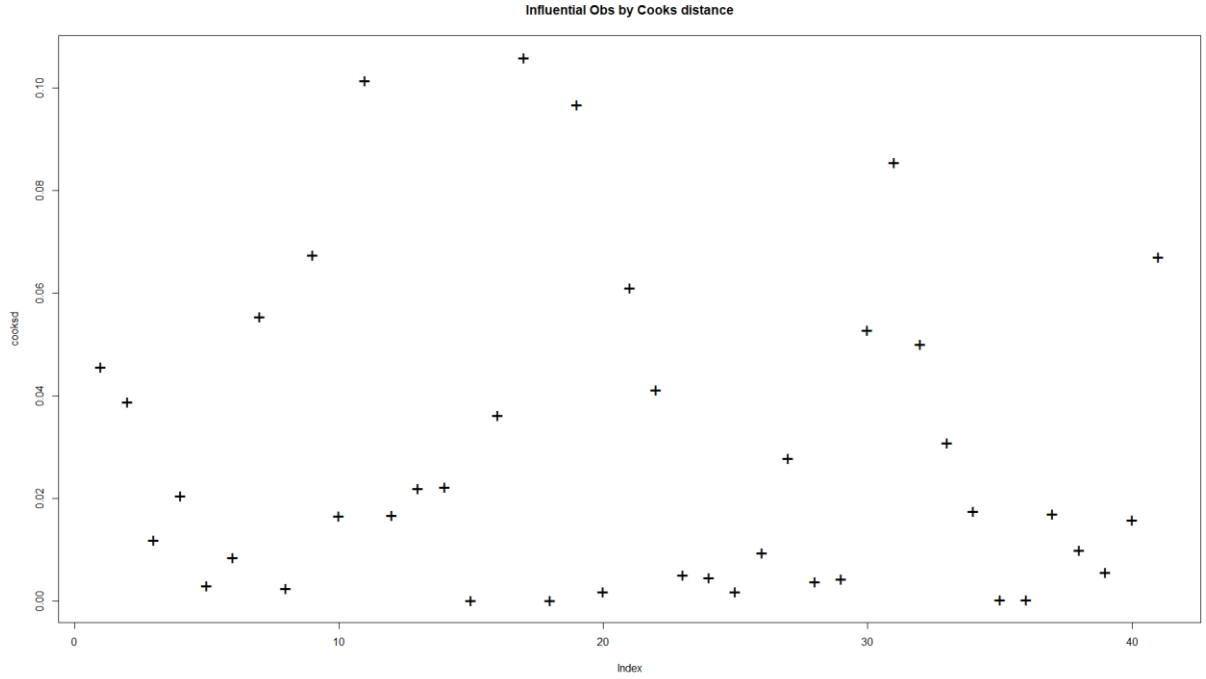
```
data: df13$y  
D = 0.16286, p-value = 0.007878
```

Değerler veri setinden çıkarılmıştır. Yeni veri setinin p-value değeri  $0.007878 < \alpha = 0.05$  olduğu için veri setinin normal dağılım göstermediği 0.05 anlamlılık ile söylenmektedir. Artık incelemesine devam edilmektedir.

## On Dördüncü Artık İncelemesi







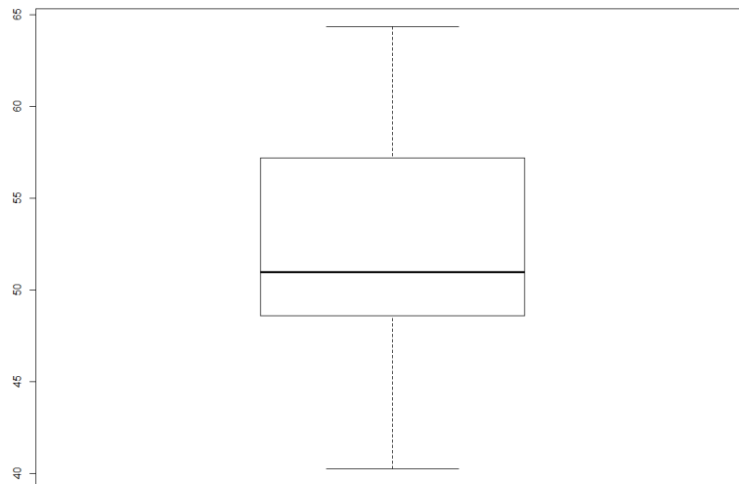
Grafiklerde görüldüğü gibi artık incelemesi sona ermiştir. Veri seti normal dağılıyormuş gibi devam edilmektedir. Oluşan yeni veri setinin bağımlı değişkeni için boxplot grafiği çizdirilmiştir. Aykırı değer olmadığı görülmektedir.

```
> lillie.test(df13$y)
```

Lilliefors (Kolmogorov-Smirnov) normality test

data: df13\$y

D = 0.16286, p-value = 0.007878





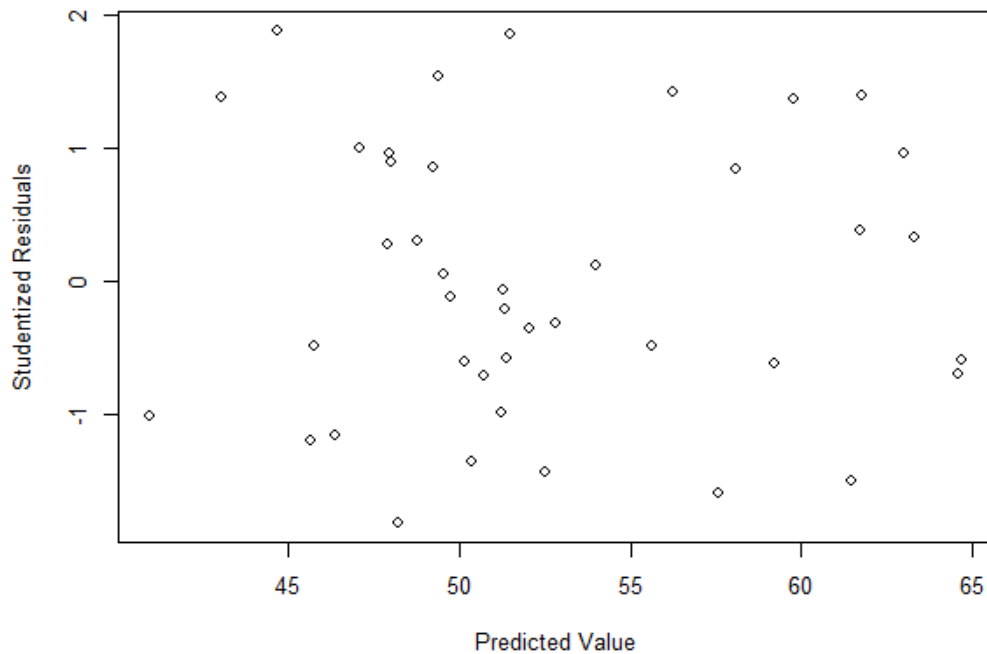


$H_0$ : Model istatistiksel olarak anlamlı değildir.

$H_1$ : Model istatistiksel olarak anlamlıdır.

P-value değeri, alfa 0.05 değerinden çok küçük bir değere sahip olduğu için  $H_0$  hipotezi reddedilir. Modelin istatistiksel olarak anlamlı olduğu 0.05 anlamlılık düzeyinde söylenebilir. Bağımsız değişkenlerin hepsi anlamlıdır. Ayrıca bağımlı değişkenin %98.78'i bağımsız değişkenlerce açıklandığı görülmektedir. Varsayım bozulumlarından söz etmek mümkün değildir.

## Değişen Varyanslılık



Student tipi artıklarla kestirim değerleri arasında çizilmiş olan grafik incelendiğinde yapının rasgele olduğu görülmektedir. Grafik sonucunda değişen varyanslılık sorunu bulunmamaktadır.

$H_0$ : Varyanslar homojendir.

$H_1$ : Varyanslar homojen değildir.

```

> library(lmtest)
> bptest(model13)

studentized Breusch-Pagan test

data:  model13
BP = 3.5291, df = 4, p-value = 0.4735

```

“lmtest” kütüphanesi yardımıyla Breusch and Pagan testi uygulanabilmektedir.

Test sonucu p-value değeri  $0.4735 > \alpha = 0.05$  olduğu için  $H_0$  hipotezi reddedilemez. Varyansların homojen olduğu 0.95 güven düzeyinde söylenebilir.

## Öz ilişki Sorunu

$H_0$ : Öz ilişki yoktur.

$H_1$ : Pozitif yönlü öz ilişki vardır.

```

> #Özilişki sorunu
> dwtest(model13)

Durbin-Watson test

data:  model13
DW = 1.7682, p-value = 0.1819
alternative hypothesis: true autocorrelation is greater than 0

```

Durbin-Watson test sonucunda p-value değeri  $0.1819 > \alpha = 0.05$  olduğu için  $H_0$  hipotezi reddedilemez. Bu durumda özilişki olmadığı 0.05 anlamlılık ile söylenebilir.

## Çoklu Bağlantı Sorunu

Çoklu bağlantı sorunu incelemesinin daha doğru sonuç vermesi için nitel değişken olan "x4" oluşturulan yeni veri setinden çıkarılmıştır. Bu çıkarma işlemi için "dplyr" kütüphanesi yardımıyla "select" komutu etkin hale getirilerek uygulanmıştır. Yeni model ve bu modele ait artık incelemesinde korelasyon katsayıları elde edilmiştir.

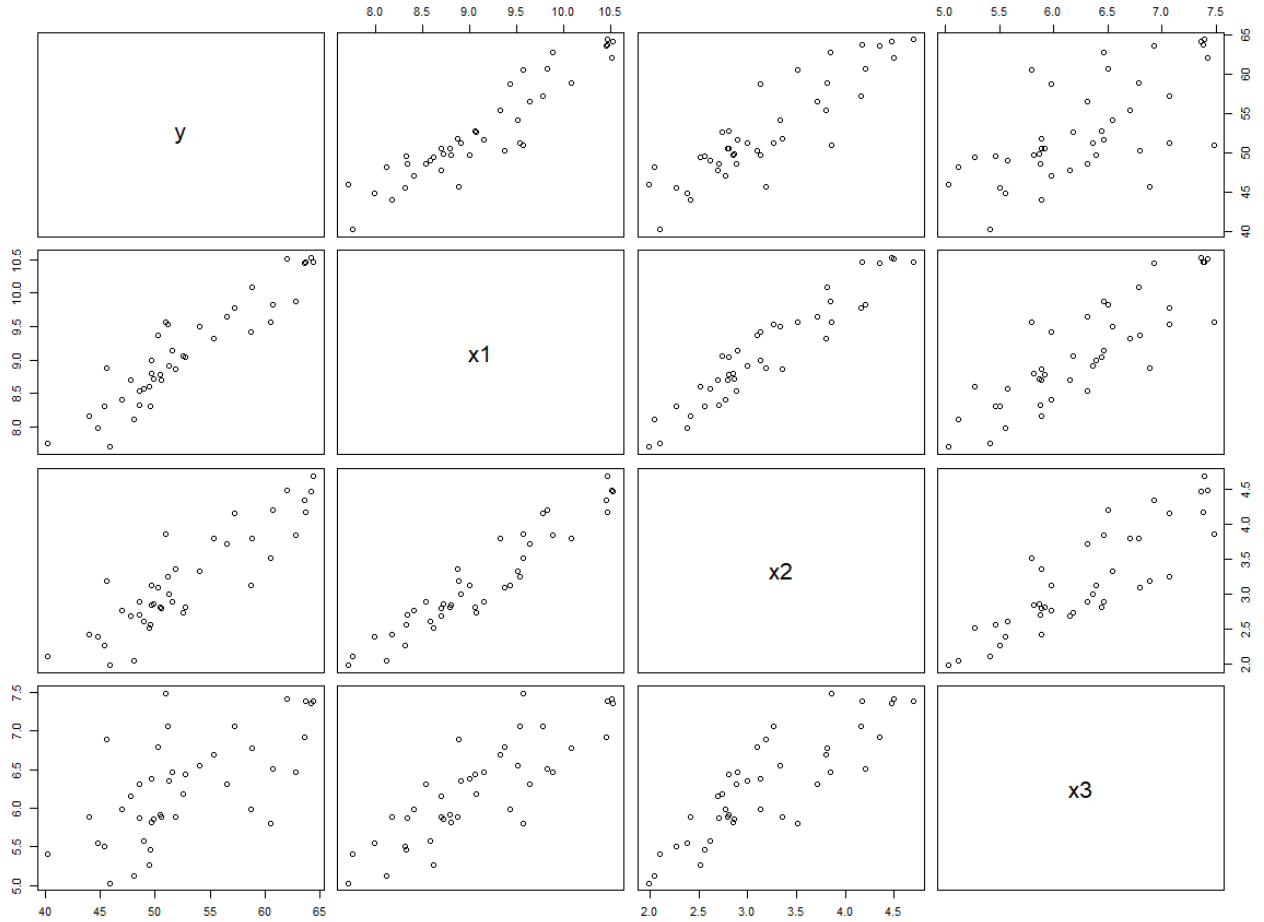
```
#Çoklu bağlantı sorunu
df14 <- select(df13, 1:4)
model14 <- lm( y ~. , data=df14)
model14

artik14 <- ls.diag(model14)
artik14

attach(df14)
```

```
$correlation
              (Intercept)              x1              x2              x3
(Intercept)  1.0000000 -0.8808253  0.8809701 -0.2122561
x1           -0.8808253  1.0000000 -0.8119856 -0.2412955
x2           0.8809701 -0.8119856  1.0000000 -0.2891450
x3          -0.2122561 -0.2412955 -0.2891450  1.0000000
```

X1 ve x2 değişkenleri arasında ilişki olduğu görülmektedir. Bu durumda çoklu bağlantı sorunundan şüphelenilmelidir.



"pairs(df14)" kodu ile veri setindeki değişkenler arasında doğrusallık grafiği oluşturulmuştur. Bağımlı değişken ile bağımsız değişkenler arasında ilişki vardır ve doğrusal model oluşturulabilir. Ancak bağımsız değişkenlerde "x1-x2" birbirleriyle ilişkili durumdadır. Çoklu bağlantı olduğu söylenebilir.

```
> vif(model14)
      x1      x2      x3
10.7120 11.0090  3.9823
```

"DAAG" kütüphanesi ya da "car" kütüphanesi yardımıyla modelin vif değerleri görülmektedir. X1 ve x2 değişkenlerinin çoklu bağlantıdan etkilendiği VİF değerlerine bakılarak söylenebilmektedir.

```
> library(perturb)
> colldiag(model14)
Condition
Index   Variance Decomposition Proportions
      intercept x1      x2      x3
1    1.000 0.000    0.000 0.000 0.000
2   12.586 0.023    0.000 0.106 0.000
3   43.929 0.094    0.022 0.136 0.990
4   94.019 0.883    0.978 0.758 0.009
```

Koşul sayısı 30'dan büyük olduğu durumda çoklu bağlantıdan etkilenilmektedir. Bu durumda 2 tane güçlü çoklu bağlantı vardır. Çoklu bağlantının yaratmış olduğu sorundan "x1 ve x2" bağımsız değişkenleri ve "x3 ve x2" bağımsız değişkenlerinin etkilendiği görülmektedir. Bu durumda çoklu bağlantı yapısı "x1-x2" ve "x3-x2" arasında oluşacaktır.

```
ortalama1<-mean(x1)
kt1<-sum((x1-ortalama1)^2)
skx1<-(x1-ortalama1)/(kt1^0.5)

ortalama2<-mean(x2)
kt2<-sum((x2-ortalama2)^2)
skx2<-(x2-ortalama2)/(kt2^0.5)

ortalama3<-mean(x3)
kt3<-sum((x3-ortalama3)^2)
skx3<-(x3-ortalama3)/(kt3^0.5)

x <- cbind(skx1,skx2,skx3)
sm <- eigen (t(x)%*%x)
signif(sm$values,3)
```

```
> signif(sm$values,3)
[1] 2.7700 0.1770 0.0508
```

```
> signif(sm$vectors,3)
      [,1] [,2] [,3]
[1,] -0.584 -0.412 0.6990
[2,] -0.585 -0.384 -0.7150
[3,] -0.563 0.826 0.0169
```

Öz değerlerden sıfıra yakın olan 0,0508'dir ve 1 tane güçlü çoklu bağlantı olduğu söylenebilir. Bu öz değere ait olan öz vektör sütununa bakıldığında "**x1 ve x2**" değişkenleri katsayılarından "**0,6990x1-0,7150x2=0**" şeklinde çoklu bağlantının yapısı yazılabilir.

```
> V<-sm$vectors
> t(V)%*%V
      [,1] [,2] [,3]
[1,] 1.000000e+00 -1.665335e-16 -1.092876e-16
[2,] -1.665335e-16 1.000000e+00 -3.139849e-16
[3,] -1.092876e-16 -3.139849e-16 1.000000e+00
> #ilişki matrisi
> V %*% diag(sm$values) %*% t(V)
      [,1] [,2] [,3]
[1,] 1.0000000 0.9491451 0.8520458
[2,] 0.9491451 1.0000000 0.8563685
[3,] 0.8520458 0.8563685 1.0000000
```

Öz değer ve öz vektörlerin ilişki matrisi yukarıdaki kod ile oluşturulmuştur.

## Model Anlamlılığı

Artık incelemesi sonucunda oluşan model13'tür. Artık incelemesinin ardından x4 değişkeni, 2 düzeye düşmüştür. Kılavuz değişkeninin X42 olduğu görülmektedir.

Çoklu bağlantı sorunu için nitel değişken çıkarılmıştı bunun sonucunda oluşan yeni model14'tür.

Ho: Model istatistiksel olarak anlamlı değildir.

Hs: Model istatistiksel olarak anlamlıdır.

```
> model14
Call:
lm(formula = y ~ ., data = df14)

Coefficients:
(Intercept)      x1      x2      x3
-2.722      8.837      3.062     -5.550

> summary(model14)

Call:
lm(formula = y ~ ., data = df14)

Residuals:
    Min       1Q   Median       3Q      Max
-2.3545 -1.1961 -0.1853  1.1893  2.5987

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.7216     5.9278  -0.459   0.64883
x1           8.8372     1.0002   8.835 1.20e-10 ***
x2           3.0620     1.0892   2.811  0.00785 **
x3          -5.5502     0.7105  -7.812 2.42e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.502 on 37 degrees of freedom
Multiple R-squared:  0.947,    Adjusted R-squared:  0.9427
F-statistic: 220.3 on 3 and 37 DF,  p-value: < 2.2e-16

> model13
Call:
lm(formula = y ~ ., data = df13)

Coefficients:
(Intercept)      x1      x2      x3      x43
-1.761      8.624      3.170     -5.232     -2.527

> summary(model13)

Call:
lm(formula = y ~ ., data = df13)

Residuals:
    Min       1Q   Median       3Q      Max
-1.21707 -0.47213 -0.08701  0.60989  1.23051

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.7605     2.8882  -0.610   0.546
x1           8.6237     0.4875  17.689 < 2e-16 ***
x2           3.1699     0.5305   5.975 7.51e-07 ***
x3          -5.2321     0.3472 -15.069 < 2e-16 ***
x43          -2.5268     0.2307 -10.955 5.14e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7314 on 36 degrees of freedom
Multiple R-squared:  0.9878,    Adjusted R-squared:  0.9864
F-statistic: 726.7 on 4 and 36 DF,  p-value: < 2.2e-16
```

P-value değerleri , alfa 0.05 değerinden çok küçük bir değere sahip olduğu için Ho hipotezleri reddedilir. Modellerin istatistiksel olarak anlamlı olduğu 0.05 anlamlılık düzeyinde söylenebilir. Bağımsız değişkenlerin hepsi anlamlıdır.

Ayrıca model14'e göre bağımlı değişkenin %94.7'si bağımsız değişkenlerce açıklandığı görülmektedir.

Model13'e göre ise bağımlı değişkenin %98.78'i bağımsız değişkenlerce açıklanabilmektedir.

$$\text{Model 13 Denklemi} = -1.7605 + 8.6237x_1 + 3.1699x_2 - 5.2321x_3 - 2.5268x_{43} + (-) 0.7314$$

(2.8882)      (0.4875)      (0.5305)      (0.3472)      (0.2307)

## Model 13 Katsayı Yorumları

```
> #regresyon katsayıları güven aralığı
> confint(model13,level=.99)
```

	0.5 %	99.5 %
(Intercept)	-9.614807	6.093758
x1	7.297916	9.949438
x2	1.727119	4.612710
x3	-6.176294	-4.287811
x43	-3.154013	-1.899498

### Sabit terim yorumu :

Ho: Sabit terimin modele katkısı anlamsızdır.

Ho: Sabit terimin modele katkısı anlamlıdır.

Çıktıda görüldüğü gibi sabit terimin p-value değeri  $0.546 > \alpha = 0.05$  olduğu için Ho hipotezi reddedilemez.

Sabit terim denklem gereği bir katsayıdır.

X1, x2, x3, x43 değişkenleri sıfır olduğunda bağımlı değişkenin alacağı ortalama değer -9.614807 ve 6.093758 aralığında olduğunu 0.99 güven düzeyinde söyleyebiliriz.

### X1 değişken yorumu:

Ho: x1 değişkeninin modele katkısı anlamsızdır.

Ho: x1 değişkeninin modele katkısı anlamlıdır.

Çıktıda görüldüğü gibi x1 değişkeninin p-value değeri  $2e-16 < \alpha = 0.05$  olduğu için Ho hipotezi reddedilir.

X1 değişkeninin modele katkısı anlamlıdır.

Diğer değişkenler sabit tutulduğunda x2 bağımsız değişkenindeki 1 birimlik artış bağımlı değişkeni ortalama 8.6237 birim arttırmaktadır.



X2, x3, x43 değişkenleri sabit olduğunda x1 değişkenindeki 1 birimlik artış bağımlı değişkenin alacağı değeri ortalama 7.297916 ve 9.949438 aralığında arttırdığı 0.99 güven düzeyinde söylenebilir.

#### X2 değişken yorumu:

Ho: x2 değişkeninin modele katkısı anlamsızdır.

Ho: x2 değişkeninin modele katkısı anlamlıdır.

Çıktıda görüldüğü gibi x2 değişkeninin p-value değeri  $7.51e-07 < \alpha = 0.05$  olduğu için Ho hipotezi reddedilir.

X2 değişkeninin modele katkısı anlamlıdır.

Diğer değişkenler sabit tutulduğunda x2 bağımsız değişkenindeki 1 birimlik artış bağımlı değişkeni ortalama 3.1699 birim arttırmaktadır.

X1, x3, x43 değişkenleri sabit olduğunda x2 değişkenindeki 1 birimlik artış bağımlı değişkenin alacağı ortalama değeri 1.727119 ve 4.612710 aralığında arttırdığı 0.99 güven düzeyinde söylenebilir.

#### X3 değişken yorumu:

Ho: x3 değişkeninin modele katkısı anlamsızdır.

Ho: x3 değişkeninin modele katkısı anlamlıdır.

Çıktıda görüldüğü gibi x3 değişkeninin p-value değeri  $2e-16 < \alpha = 0.05$  olduğu için Ho hipotezi reddedilir.

X3 değişkeninin modele katkısı anlamlıdır.

Diğer değişkenler sabit tutulduğunda x3 bağımsız değişkenindeki 1 birimlik artış bağımlı değişkeni ortalama 5.2321 birim azaltmaktadır.

X1, x2, x43 değişkenleri sabit olduğunda x3 değişkenindeki 1 birimlik artış bağımlı değişkenin alacağı ortalama değeri -6.176294 ve -4.287811 aralığında arttırdığı 0.99 güven düzeyinde söylenebilir.

### X4 değişken yorumu

Ho: X42 ve X43 arasında fark yoktur.

Hs: X42 ve X43 arasında fark vardır.

Çıktıda görüldüğü gibi p-value değeri  $5.14e-13 < \alpha = 0.05$  olduğu için Ho hipotezi reddedilir.

X42 ve x43 arasında istatistiksel olarak farklılık bulunmaktadır. X42'nin x43'e göre daha etkili olduğu söylenebilir.

## Uyum Kestirimi

Veri seti içerisinde bulunan x1, x2, x3, x4 değerleri koda girilmiştir. Y bağımlı değişkeninin ortalama değeri 52.80171 çıkmıştır.

```
> #uyum kestirimi
> predict(model13,data.frame(x1=9.068291,x2=2.741881,x3=6.179469,x4="2"), interval='confidence')
      fit      lwr      upr
1 52.80171 52.26374 53.33968
```

Girilen x1, x2, x3, x4 değerleri için y miktarının 52.26374 ile 53.33968 aralığında değişim gösterdiği söylenebilir.

## Ön Kestirim

Veri setinde bulunmayan x1, x2, x3, x4 değerleri koda girilmiştir. Y bağımlı değişkeninin ortalama değeri 138.3604 çıkmıştır.

```
> #önkestirim
> predict(model13,data.frame(x1=10,x2=31,x3=8,x4="3"), interval='confidence')
      fit      lwr      upr
1 138.3604 109.4836 167.2372
```

Girilen x1, x2, x3, x4 değerleri için y miktarının 109.4836 ile 167.2372 aralığında değişim gösterdiği söylenebilir.

# Değişken Seçimi Yöntemi

## İleriye Doğru Seçim Yöntemi

```
> library(stats)
> lm.null <- lm(df13$y ~ 1)
> forward <- step(lm.null , df13$y~ x1+x2+x3+x4, direction = "forward")
Start: AIC=151.56
df13$y ~ 1
```

	Df	Sum of Sq	RSS	AIC
+ x1	1	1352.33	221.84	73.224
+ x2	1	1237.40	336.77	90.339
+ x3	1	654.61	919.57	131.524
<none>			1574.17	151.564
+ x4	1	40.07	1534.10	152.507

```
Step: AIC=73.22
df13$y ~ x1
```

	Df	Sum of Sq	RSS	AIC
+ x3	1	120.570	101.27	43.075
+ x4	1	79.527	142.32	57.024
<none>			221.84	73.224
+ x2	1	0.751	221.09	75.085

```
Step: AIC=43.07
df13$y ~ x1 + x3
```

	Df	Sum of Sq	RSS	AIC
+ x4	1	62.921	38.353	5.263
+ x2	1	17.824	83.450	37.138
<none>			101.274	43.075

```
Step: AIC=5.26
df13$y ~ x1 + x3 + x4
```

	Df	Sum of Sq	RSS	AIC
+ x2	1	19.096	19.257	-20.9842
<none>			38.353	5.2632

```
Step: AIC=-20.98
df13$y ~ x1 + x3 + x4 + x2
```

```
> forward

Call:
lm(formula = df13$y ~ x1 + x3 + x4 + x2)

Coefficients:
(Intercept)      x1      x3      x43      x2
    -1.761      8.624     -5.232     -2.527      3.170
```

Bağımlı değişken y olduğu durumda birinci adımda modele  $x_1$  değişkeni girmiştir. Daha sonra sırayla  $x_3$ ,  $x_4$  ve  $x_2$  değişkenleri modele girmiştir. Tüm değişkenler modele alındığında model anlamlı hale gelmiştir.

```
> summary(forward)

Call:
lm(formula = df13$y ~ x1 + x3 + x4 + x2)

Residuals:
    Min       1Q   Median       3Q      Max
-1.21707 -0.47213 -0.08701  0.60989  1.23051

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.7605     2.8882  -0.610   0.546
x1             8.6237     0.4875  17.689 < 2e-16 ***
x3            -5.2321     0.3472 -15.069 < 2e-16 ***
x43           -2.5268     0.2307 -10.955 5.14e-13 ***
x2             3.1699     0.5305   5.975 7.51e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7314 on 36 degrees of freedom
Multiple R-squared:  0.9878,    Adjusted R-squared:  0.9864
F-statistic: 726.7 on 4 and 36 DF,  p-value: < 2.2e-16
```

P - value değerlerinden görüldüğü üzere modeldeki değişkenler %95 güven düzeyinde anlamlıdır. Modelin de anlamlı olduğu görülmektedir. (p=2.2e-16)

En iyi denklem =  $-1.7605 + 8.6237x_1 + 3.1699x_2 - 5.2321x_3 - 2.5268x_4 + (-) 0.7314$

(2.8882)      (0.4875)      (0.5305)      (0.3472)      (0.2307)

## Geriye Doğru Çıkarma Yöntemi

```
> #geriye doğru
> backward<-step(model13,direction="backward")
Start: AIC=-20.98
y ~ x1 + x2 + x3 + x4
```

	Df	Sum of Sq	RSS	AIC
<none>			19.257	-20.984
- x2	1	19.096	38.353	5.263
- x4	1	64.193	83.450	37.138
- x3	1	121.459	140.716	58.560
- x1	1	167.382	186.638	70.140

```
> backward

Call:
lm(formula = y ~ x1 + x2 + x3 + x4, data = df13)

Coefficients:
(Intercept)          x1          x2          x3          x43
    -1.761         8.624         3.170        -5.232        -2.527
```

Geriye doğru seçim yönteminin özelliğinden tüm değişkenler modelde olarak başlıyor. İlk model tüm bağımsız değişkenlerin modelde bulunduğu durumdur. Geriye doğru seçim yöntemi bu modeli en iyi model olarak kabul etmiştir.

```
> summary(backward)

Call:
lm(formula = y ~ x1 + x2 + x3 + x4, data = df13)

Residuals:
    Min       1Q   Median       3Q      Max
-1.21707 -0.47213 -0.08701  0.60989  1.23051

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.7605     2.8882  -0.610   0.546
x1             8.6237     0.4875  17.689 < 2e-16 ***
x2             3.1699     0.5305   5.975 7.51e-07 ***
x3            -5.2321     0.3472 -15.069 < 2e-16 ***
x43           -2.5268     0.2307 -10.955 5.14e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7314 on 36 degrees of freedom
Multiple R-squared:  0.9878,    Adjusted R-squared:  0.9864
F-statistic: 726.7 on 4 and 36 DF,  p-value: < 2.2e-16
```

P - value değerlerinden görüldüğü üzere modeldeki değişkenler %95 güven düzeyinde anlamlıdır. Modelin de anlamlı olduğu görülmektedir. (p=2.2e-16)

En iyi denklem =  $-1.7605 + 8.6237x_1 + 3.1699x_2 - 5.2321x_3 - 2.5268x_4 + (-) 0.7314$

(2.8882)      (0.4875)      (0.5305)      (0.3472)      (0.2307)

## Adımsal Regresyon Yöntemi

```
> #adımsal
> #library(MASS)
> step.model <- stepAIC(model13, direction = "both", trace = FALSE)
> step.model
```

Call:

```
lm(formula = y ~ x1 + x2 + x3 + x4, data = df13)
```

Coefficients:

(Intercept)	x1	x2	x3	x43
-1.761	8.624	3.170	-5.232	-2.527

```
> summary(step.model)
```

Call:

```
lm(formula = y ~ x1 + x2 + x3 + x4, data = df13)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.21707	-0.47213	-0.08701	0.60989	1.23051

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1.7605	2.8882	-0.610	0.546
x1	8.6237	0.4875	17.689	< 2e-16 ***
x2	3.1699	0.5305	5.975	7.51e-07 ***
x3	-5.2321	0.3472	-15.069	< 2e-16 ***
x43	-2.5268	0.2307	-10.955	5.14e-13 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7314 on 36 degrees of freedom

Multiple R-squared: 0.9878, Adjusted R-squared: 0.9864

F-statistic: 726.7 on 4 and 36 DF, p-value: < 2.2e-16

İlk model  $x_1$  değişkeni alınarak başlamıştır. Diğer bağımsız değişkenler anlamlı olduğu için yöntem tamamen ileri doğru seçim yöntemi gibi sonra ermiştir. En iyi model tüm değişkenlerin olduğu modeldir.

En iyi denklem =  $-1.7605 + 8.6237x_1 + 3.1699x_2 - 5.2321x_3 - 2.5268x_4 + (-) 0.7314$

(2.8882) (0.4875) (0.5305) (0.3472) (0.2307)

## Ridge Regresyon

```
> #ridge regresyon
> ridge <- lm.ridge(df14$y~x1+x2+x3 ,lambda = seq(0,1,0.05),data=df14)
> matplot(ridge$lambda,t(ridge$coef),type="l",xlab=expression(lambda),
+         ylab=expression(hat(beta)))
> abline(h=0,lwd=2)
> ridge$coef
```

	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
x1	6.782146	6.711402	6.643256	6.577545	6.514123	6.452851	6.393606	6.336271	6.280742	6.226921
x2	2.187674	2.221565	2.253302	2.283043	2.310927	2.337086	2.361637	2.384688	2.406340	2.426683
x3	-3.656406	-3.620737	-3.585522	-3.550757	-3.516435	-3.482548	-3.449091	-3.416056	-3.383437	-3.351227

```
> ridge$coef
```

	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
x1	6.174716	6.124044	6.074828	6.026994	5.980477	5.935212	5.891141	5.848211	5.806368	5.765567
x2	2.445802	2.463774	2.480670	2.496556	2.511495	2.525542	2.538750	2.551167	2.562840	2.573811
x3	-3.319420	-3.288009	-3.256988	-3.226351	-3.196090	-3.166201	-3.136676	-3.107511	-3.078699	-3.050234

```
> select(ridge)
modified HKB estimator is 0.03515669
modified L-W estimator is 0.06203113
smallest value of GCV at 0.15
> ridge$coef[,ridge$lam == 0.4]
      x1      x2      x3
6.280742 2.406340 -3.383437
```

Ridge regresyon çoklu bağlantıyı yakalayamamıştır. Dik iniş, çıkış yapan değişken görülmemektedir.

