



HACETTEPE
ÜNİVERSİTESİ

REGRESYON ÇÖZÜMLEMESİ

HAZIRLAYAN : Şükriye Nur Şencan

Öncelikle veri import edilmiştir.

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

	y	x1	x2	x3	x4
1	912	0.40	62.4	4.71	VS2
2	449	0.32	60.6	4.40	SI1
3	710	0.30	61.7	4.28	VS1
4	480	0.32	61.6	4.40	SI1
5	773	0.32	63.3	4.38	VS2
6	827	0.41	61.7	4.70	VS2
7	984	0.35	60.7	4.58	VS1
8	672	0.32	63.4	4.34	SI1
9	561	0.32	63.3	4.36	VS2
10	561	0.32	63.1	4.34	VS1
11	743	0.33	60.2	4.47	VS1
12	702	0.32	62.0	4.37	VS2
13	556	0.36	61.9	4.57	SI1
14	666	0.33	62.3	4.41	VS2
15	743	0.33	62.6	4.40	VS2

Showing 1 to 15 of 100 entries, 5 total columns

Veri setinin ismi veri olarak aşağıdaki gibi tanımlanmıştır. Veri seti 100 gözlemden oluşmaktadır. Y bağımlı değişken; X1,X2,X3 bağımsız değişkenlerdir. X4 değişkeni ise faktör olarak atanmıştır.

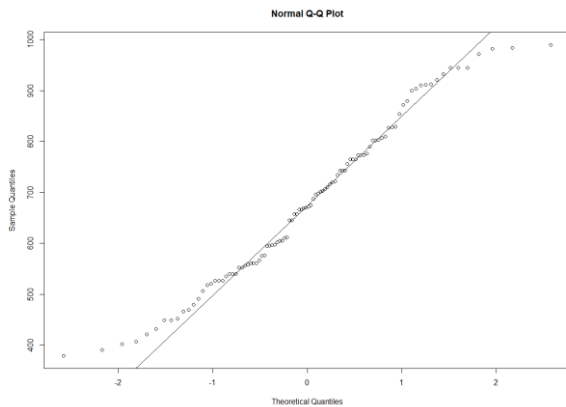
```
veri <- odev  
attach(veri)  
veri[,5] <- as.factor(x4)  
attach(veri)  
x4
```

```
> x4  
[1] VS2 SI1 VS1 SI1 VS2 VS2 VS1 SI1 VS2 VS1 VS1 VS2 SI1 VS2 VS2 SI1 VS2 SI1 SI1 VS1 SI1 SI1 VS1  
[24] SI1 VS2 VS2 VS2 VS2 VS1 VS2 VS1 VS2 VS2 SI1 SI1 SI1 SI1 SI1 VS2 VS2 VS1 SI1 VS2 SI1 VS2 SI1  
[47] SI1 VS2 VS2 VS2 VS1 VS1 VS1 VS2 VS2 SI1 VS2 VS2 VS2 VS2 VS2 VS1 SI1 VS1 VS2 SI1 SI1 SI1 SI1  
[70] VS2 VS1 VS2 VS1 VS2 VS2 SI1 VS1 VS2 VS2 VS2 SI1 VS2 SI1 VS2 VS2 SI1 SI1 VS2 VS2 VS1 SI1 VS2  
[93] SI1 VS2 SI1 VS2 SI1 VS2 VS2 SI1  
Levels: SI1 VS1 VS2
```

Veri setindeki X4 nitel değişkeninin; SI1 ,VS1, VS2'den oluştuğu görülmektedir.

Normallik ve Doğrusallık

Q-Q Plot Grafiği



Veri setinde noktalar 45 derecelik doğru etrafında dağıldığı için grafiğe göre bağımlı değişkenin normal dağılım gösterdiği söylenebilir.

Normallik testi

```
> #install.packages("nortest")
> library(nortest)
> lillie.test(y)

Lilliefors (Kolmogorov-Smirnov) normality test

data: y
D = 0.081775, p-value = 0.09594
```

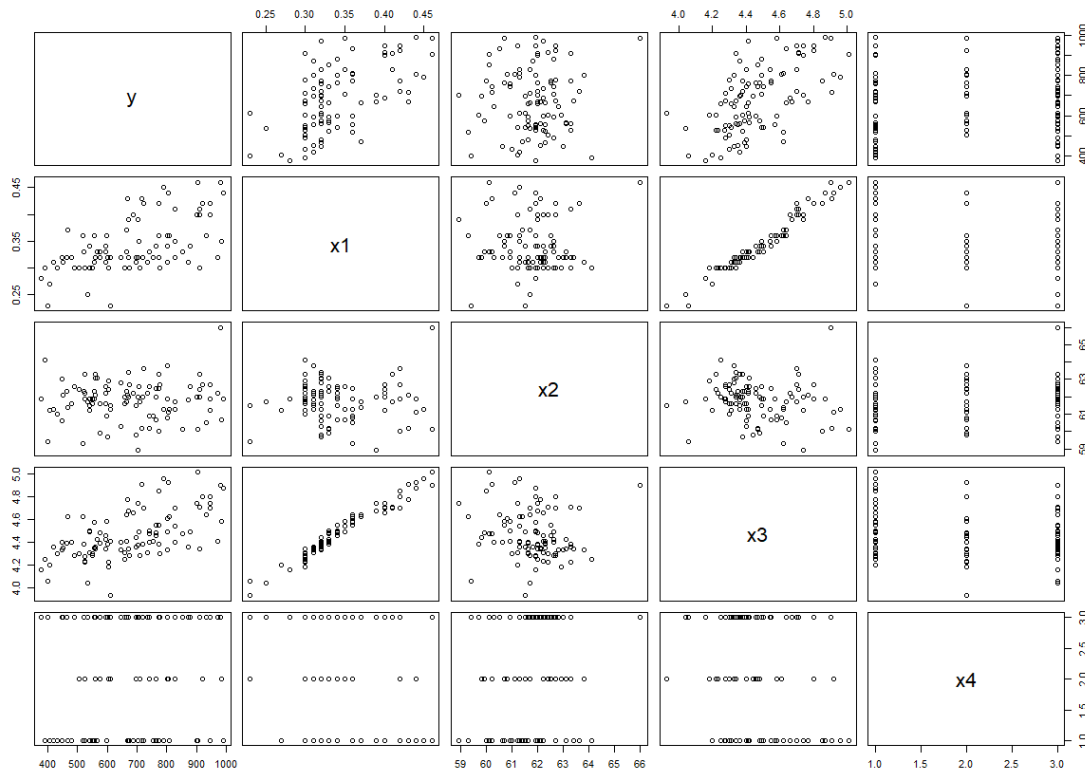
H0: Verilerin dağılışı ile normal dağılım arasında fark yoktur.

HS: Verilerin dağılışı ile normal dağılım arasında fark vardır.

Kolmogrov-Smirnov Normallik Testine $p \text{ value} = 0.09 > \alpha = 0,05$ olduğundan H0 hipotezi kabul edilir. Veriler normal dağılıma uygun dağılış gösterir.

Doğrusallık

```
#Doğrusallık
pairs(veri)
```



Y bağımlı değişkeni ile X1, X3 bağımsız değişkenleri arasında doğrusala yakın bir ilişki görülürken , X2 değişkeni ile Y arasında doğrusal bir ilişki görülmemektedir. Ayrıca X1 ve X3 bağımsız değişkenleri arasında ilişki olduğu görülmektedir dolayısıyla çoklu bağlantı sorunu olduğu söylenebilir.

Artıkların İncelenmesi

```
> sonuc<-lm(y~x1+x2+x3+x4)
> sonuc

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

Coefficients:
(Intercept)          x1          x2          x3          x4VS1          x4VS2
-6003.87        -3268.02         32.26        1282.90         130.72          99.04
```

Bağımlı değişken y olmak üzere x1, x2 ve x3 bağımsız değişkenleri girilmiştir. S11 değişkeninin kılavuz değişken olduğu görülmektedir.

```
sonuc<-lm(y~x1+x2+x3+x4)
sonuc
summary(sonuc)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -6003.87    1929.92   -3.111 0.002469 **
x1          -3268.02    1692.26   -1.931 0.056476 .
x2             32.26     14.98    2.153 0.033854 *
x3           1282.90     388.49    3.302 0.001357 **
x4VS1         130.72      34.83    3.753 0.000303 ***
x4VS2          99.04      27.31    3.626 0.000467 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 115.9 on 94 degrees of freedom
Multiple R-squared:  0.4981, Adjusted R-squared:  0.4715
F-statistic: 18.66 on 5 and 94 DF, p-value: 7.732e-13
```

H0: Model istatistiksel olarak anlamlı değildir.

HS: Model istatistiksel olarak anlamlıdır.

Tabloda görüldüğü üzere P value değeri 0.05'den oldukça küçüktür. H0 hipotezi reddedilir. Model istatistiksel olarak anlamlı olarak gözükmemektedir. Ancak bağımsız değişkenlerden X1 değişkeninin anlamsız, X2 ve X3 değişkenlerinin anlamlı olduğu görülmektedir. Ayrıca bağımlı değişkenin %49'unun bağımsız değişkenlerce açıklandığı yorumlaması yapılmasına rağmen bağımsız değişkenlerden X1 anlamsız olduğu için varsayım bozulumlarından şüphelenebiliriz.

Aykırı Değer İncelemesi

1.Artık İncelemesi

Uç değer: $h_{ii} > 2*(4+1)/100 = 0.1$ olan veriler uç değerdir ve veri setinden 6,23,28,31,41,63,64,68,73,78,87,91,95,96 numaralı değişkenler atılmıştır.

```
$hat
[1] 0.04487761 0.04391011 0.05987492 0.03727871 0.05063789 0.11595094 0.09530708 0.06681800
[9] 0.03910580 0.07057085 0.07887883 0.02284580 0.03382617 0.02219692 0.02466927 0.07320113
[17] 0.02938675 0.03231416 0.04478692 0.08435735 0.03564918 0.04913484 0.13702400 0.04021887
[25] 0.02186820 0.05277847 0.02573038 0.12262874 0.08706261 0.07190762 0.10770451 0.04305875
[33] 0.06164039 0.06052550 0.04520965 0.04306708 0.03211930 0.06890767 0.02748833 0.03523601
[41] 0.14531803 0.06932263 0.02761678 0.03131688 0.03783492 0.05167124 0.07134838 0.02870702
[49] 0.02523359 0.02715233 0.07201233 0.06488913 0.08838192 0.03705127 0.02855952 0.04772508
[57] 0.02635055 0.02959021 0.06150332 0.03932431 0.04530393 0.05982241 0.12650507 0.12545529
[65] 0.02322901 0.03873965 0.06470738 0.10745165 0.04442064 0.04690027 0.08645000 0.03604474
[73] 0.12472843 0.02949093 0.02191439 0.07122911 0.09145605 0.23212653 0.02747379 0.05050528
[81] 0.03827320 0.06835738 0.05849472 0.03609010 0.07687041 0.06697303 0.10510253 0.02660130
[89] 0.03680187 0.09123367 0.12255640 0.02657952 0.07470981 0.05368576 0.10007379 0.10933851
[97] 0.07141665 0.05775456 0.02654242 0.04792509
```

Standartlaştırılmış Artıklar: -2 ile +2 aralığında olmayan değerler aykırı değerlerdir ve veri setinden atılır. Buna göre 80 ve 89 gözlemleri veri setinden atılmıştır.

```
$std.res
[1] 0.606094042 -0.891262029 0.733291157 -0.899254347 0.553760042 0.474793473 1.516744641
[8] 0.970622482 -1.089942415 -1.104351607 0.162484712 0.404104956 -1.082442824 -0.157263853
[15] 0.542869662 -1.038741515 -1.173350673 -0.781810697 0.183239610 -0.483262549 1.164962420
[22] -0.601446387 0.622131037 -1.170921595 0.449020322 1.453836852 -0.218301901 -0.610403139
[29] -0.688583287 0.976489833 -0.126156364 1.026336527 -1.200751195 0.764836885 -1.014317724
[36] 1.147651623 -0.551451092 -0.618548917 0.234587692 -1.094074887 1.961075239 1.675094214
[43] -0.633937999 -1.978272447 1.074611666 0.314473762 -0.681932404 1.016843805 -0.367065550
[50] 0.533063075 -1.521415368 0.101609516 -1.383501999 -0.522751059 0.008416957 0.430289334
[57] -0.984073105 -1.002180376 0.602784904 -1.248558251 -1.439653529 1.109856484 0.158476996
[64] -1.520525881 -0.807306160 1.768258690 -0.505162851 -1.322773519 0.288477596 0.803280939
[71] -0.206284797 0.130090033 0.693493524 0.664132136 -1.014406224 -0.072890080 -0.092503742
[78] -0.247729847 1.775351633 2.022806520 1.614667049 0.628413890 1.297101237 -1.649300121
[85] 0.750526643 0.756734979 0.411304363 0.337345367 2.273068167 0.287744780 -0.410592477
[92] -1.968270902 0.119148302 0.050068644 0.003841445 0.070015656 -1.267768782 -0.100636872
[99] -1.871261441 1.802666331
```

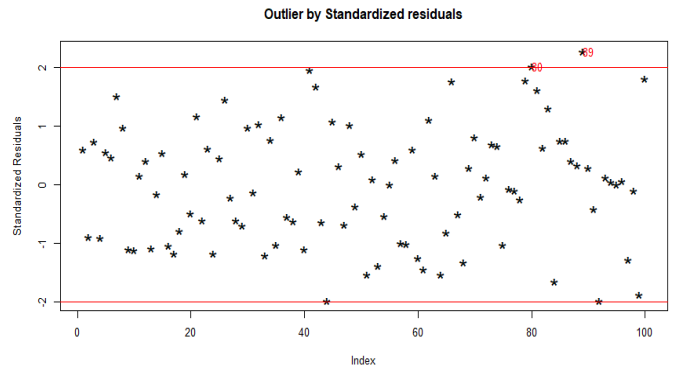
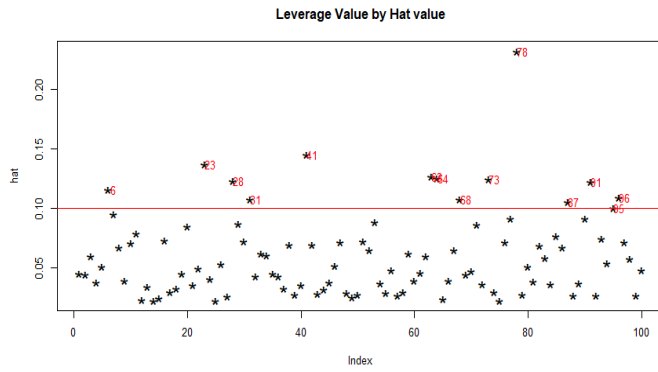
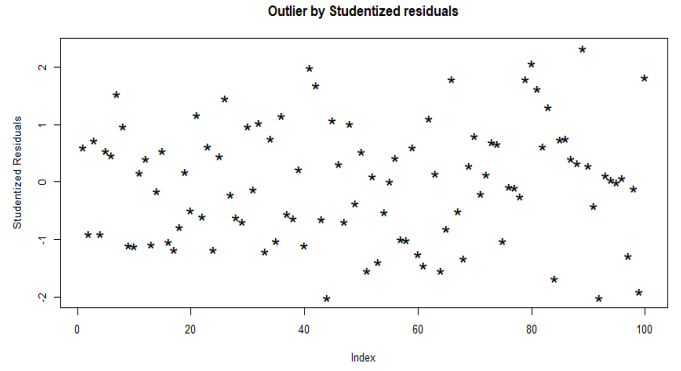
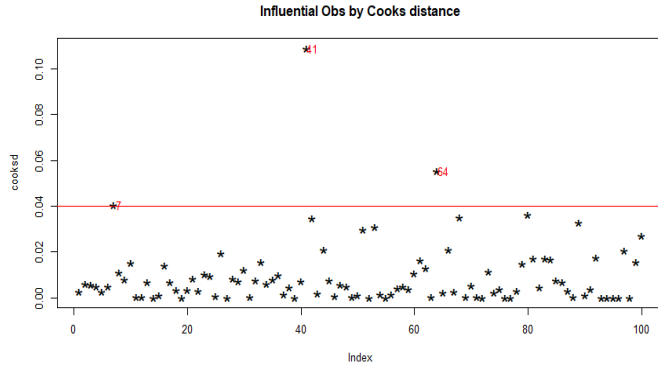
Student True Artıklar: -3 ile +3 aralığında olmayan değerler aykırı değerlerdir. Burada aykırı değer bulunmamıştır.

```
$stud.res
[1] 0.604042966 -0.890278231 0.731475403 -0.898330692 0.551707269 0.472828527 1.527462277
[8] 0.970320519 -1.091045600 -1.105657723 0.161640821 0.402299312 -1.083443330 -0.156445688
[15] 0.540822792 -1.039182892 -1.175734596 -0.780181683 0.182294886 -0.481283376 1.167205720
[22] -0.599393076 0.620090917 -1.173264518 0.447105282 1.462620403 -0.217192678 -0.608354510
[29] -0.686644764 0.976245972 -0.125494150 1.026631128 -1.203613526 0.763135984 -1.014475037
[36] 1.149613228 -0.549399396 -0.616505910 0.233404881 -1.095235509 1.991786513 1.691599074
[43] -0.631909217 -2.010008382 1.075507082 0.312961227 -0.679979470 1.017029574 -0.365369803
[50] 0.531023285 -1.532284573 0.101073146 -1.390351610 -0.520720481 0.008372070 0.428416569
[57] -0.983905960 -1.002203899 0.600732194 -1.252326947 -1.448028058 1.111242107 0.157652842
[64] -1.531366135 -0.805798834 1.788830613 -0.503152068 -1.328137817 0.287066137 0.801753288
[71] -0.205231063 0.129407864 0.691566268 0.662145375 -1.014564530 -0.072503379 -0.092014574
[78] -0.246489088 1.796254504 2.057292476 1.628802195 0.626379454 1.301886920 -1.664768322
[85] 0.748770661 0.755002280 0.409479359 0.335749480 2.325767920 0.286336263 -0.408769358
[92] -1.999408416 0.118521790 0.049802273 0.003820958 0.069644052 -1.271928070 -0.100105531
[99] -1.896948493 1.824872648
```

Cook Uzaklığı: $n(100) > 50$ olduğu için $4/100=0.04$ değerinden büyük olan değerler etkin gözlemlerdir. Buna göre ; 7,41,64 gözlemleri etkin gözlemlerdir, veri setinden atılmıştır.

```
$cooks
[1] 2.876732e-03 6.080305e-03 5.707698e-03 5.218841e-03 2.726062e-03 4.927835e-03 4.039222e-02
[8] 1.124285e-02 8.057892e-03 1.543379e-02 3.768057e-04 6.363269e-04 6.836852e-03 9.357243e-05
[15] 1.242351e-03 1.420352e-02 6.947199e-03 3.401813e-03 2.623846e-04 3.586014e-03 8.361554e-03
[22] 3.115396e-03 1.024260e-02 9.575513e-03 7.512705e-04 1.962842e-02 2.097640e-04 8.679428e-03
[29] 7.536199e-03 1.231311e-02 3.201786e-04 7.899588e-03 1.578523e-02 6.281159e-03 8.119325e-03
[36] 9.879452e-03 1.681926e-03 4.719236e-03 2.592464e-04 7.286323e-03 1.089814e-01 3.483398e-02
[43] 1.902295e-03 2.108714e-02 7.568244e-03 8.980645e-04 5.954739e-03 5.093252e-03 5.813189e-04
[50] 1.321808e-03 2.993705e-02 1.194061e-04 3.092850e-02 1.752421e-03 3.471312e-07 1.546515e-03
[57] 4.368080e-03 5.104267e-03 3.968618e-03 1.063532e-02 1.639213e-02 1.306280e-02 6.062179e-04
[64] 5.527686e-02 2.583231e-03 2.100172e-02 2.942510e-03 3.510759e-02 6.447495e-04 5.292010e-03
[71] 6.711443e-04 1.054683e-04 1.142237e-02 2.233812e-02 3.842600e-03 6.791012e-05 1.435601e-04
[78] 3.092007e-03 1.484003e-02 3.627452e-02 1.729250e-02 4.829213e-03 1.742170e-02 1.697461e-02
[85] 7.817673e-03 6.850813e-03 3.311428e-03 5.183348e-04 3.290242e-02 1.385372e-03 3.924529e-03
[92] 1.763052e-02 1.910399e-04 2.370304e-05 2.734964e-07 1.002997e-04 2.060191e-02 1.034630e-04
[99] 1.591260e-02 2.726285e-02
```

Grafiklere bakıldığında;



Aykırı değerler, uç değerler ve etkin gözlemler aşağıdaki gibi kodlanarak veri setinden çıkarılmıştır.

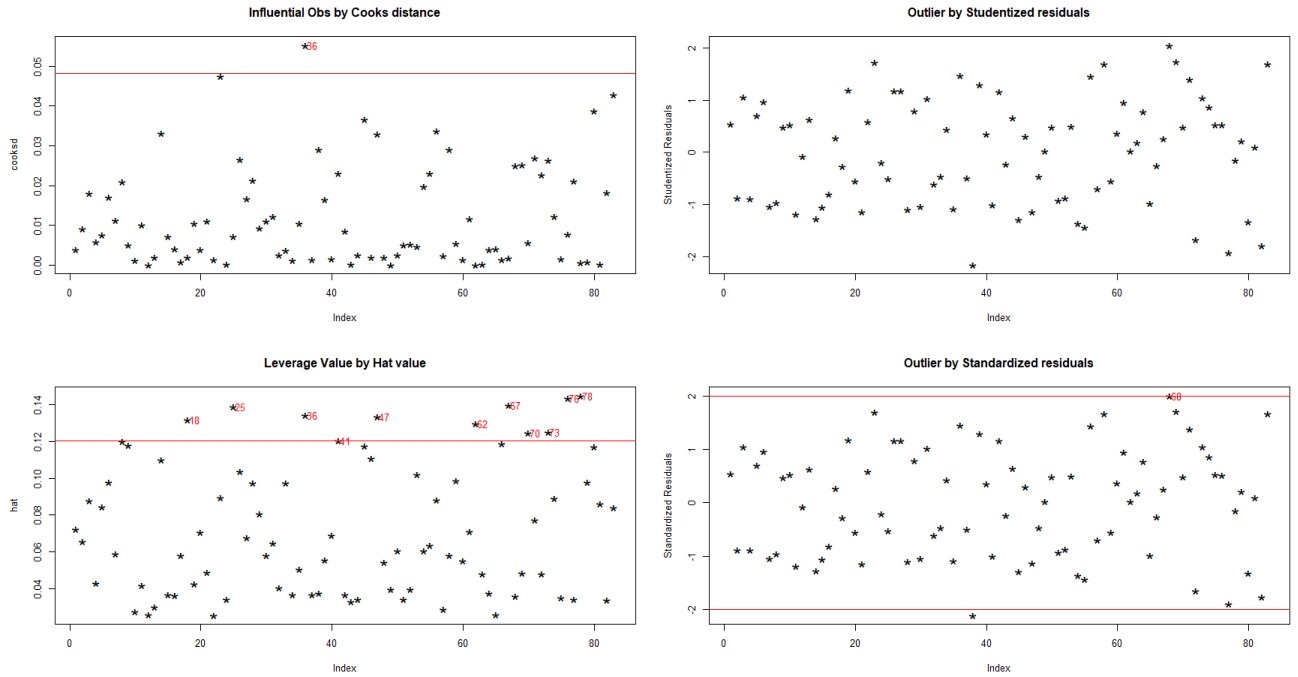
```
veri1 <- veri[-c(7,41,64,6,23,80,89,28,31,41,63,64,68,73,78,87,91,95,96),]
attach(veri1)
lillie.test(veri1$y)
```

```
Lilliefors (Kolmogorov-Smirnov) normality test
data: veri1$y
D = 0.0936, p-value = 0.0694
```

H0: Verilerin dağılışı ile normal dağılım arasında fark yoktur.

Verinin normalliği test edilmiştir ve $p=0.06 > \alpha = 0,05$ olduğundan H0 hipotezi kabul edilir. Veriler normal dağılıma uygun dağılış gösterir.

2.Artık İncelemesi



Cook uzaklığına bakıldığında 36. Gözlem etkin gözlemdir ve veri setinden çıkarılmıştır.

Hat değerlerine bakıldığında; 18,25,36,41,47,62,67,70,73,76,78 gözlemleri uç değerlerdir ve veri setinden atılmıştır.

Student True artıklara bakıldığında aykırı değer görülmemektedir.

Standartlaştırılmış artıklara bakıldığında 68. Gözlem aykırı değerdir ve veri setinden atılmıştır.

Aykırı değerler, uç değerler ve etkin gözlemler aşağıdaki gibi kodlanarak veri setinden çıkarılmıştır.

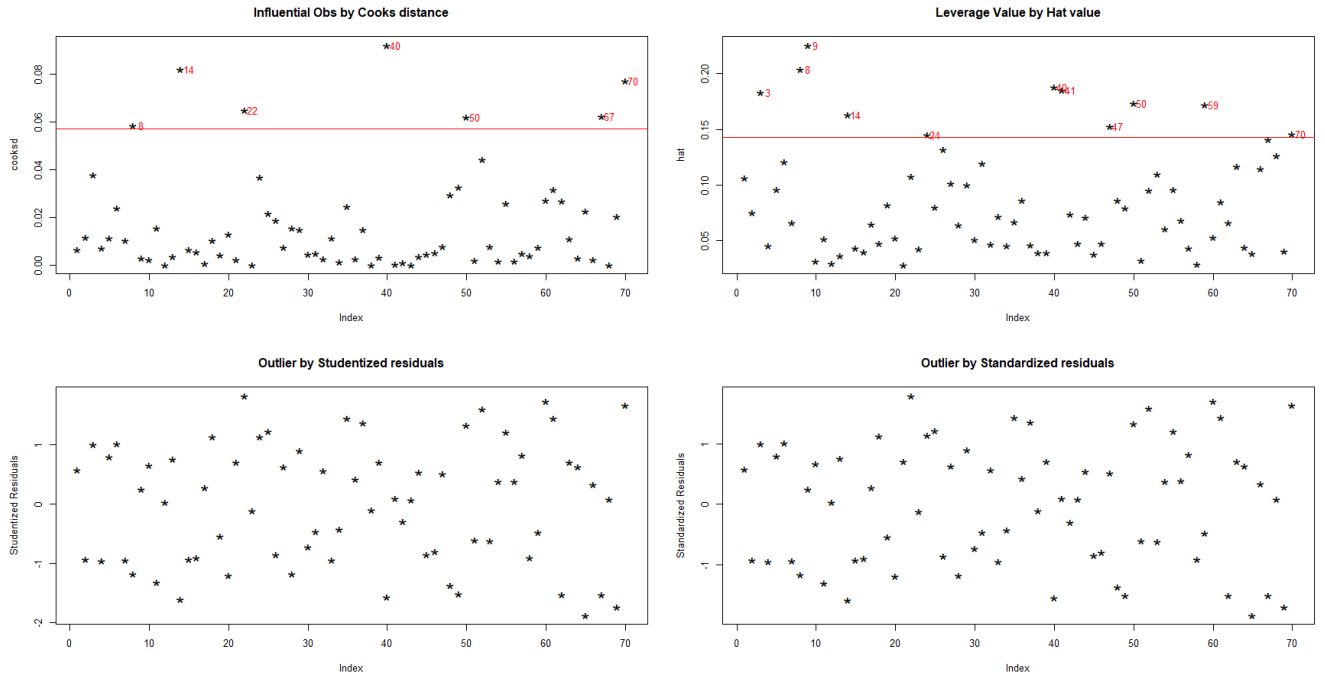
```
veri2 <- veri1[-c(36,18,25,41,47,62,67,70,73,76,78,38,68),]  
attach(veri2)  
lillie.test(veri2$y)
```

```
> lillie.test(veri2$y)  
  
Lilliefors (Kolmogorov-Smirnov) normality test  
data: veri2$y  
D = 0.090035, p-value = 0.1723
```

H0: Verilerin dağılışı ile normal dağılım arasında fark yoktur.

Verinin normalliği test edilmiştir ve $p=0.1723 > \alpha = 0,05$ olduğundan H0 hipotezi kabul edilir. Veriler normal dağılıma uygundur.

3.Artık İncelemesi



Cook uzaklığı için grafiğe bakıldığında 8,14,22,40,50,67,70 gözlemleri etkin gözlemdir ,veri setinden çıkarılır.

Hat değerlerine bakıldığında ; 3,8,9,14,24,40,41,47,50,59,70 gözlemleri uç değerlerdir, veri setinden atılır.

Student True artıklar ve Standartlaştırılmış artıklar bulunmamıştır.

Aykırı değerler, uç değerler ve etkin gözlemler veri setinden çıkarılmıştır.

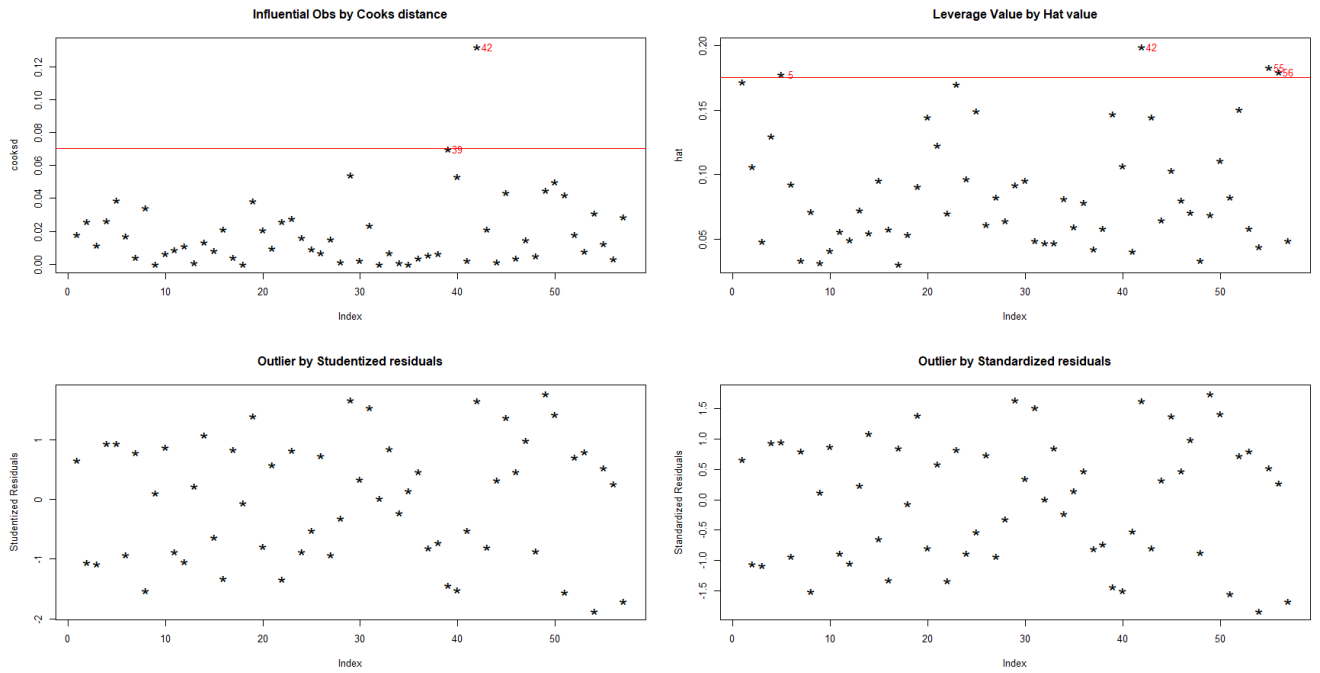
```
veri3 <- veri2[-c(22,40,67,70,3,8,9,14,24,41,47,50,59),]  
attach(veri3)  
lillie.test(veri3$y)
```

```
> lillie.test(veri3$y)  
  
Lilliefors (Kolmogorov-Smirnov) normality test  
  
data: veri3$y  
D = 0.085982, p-value = 0.368
```

H0: Verilerin dağılışı ile normal dağılım arasında fark yoktur.

Verinin normalliği test edilmiştir ve $p=0.368 > \alpha = 0,05$ olduğundan H0 hipotezi kabul edilir. Veriler normal dağılıma uygundur.

4.Artık İncelemesi



Cook değerlerine bakıldığında 39,42 gözlemleri etkin gözlemlerdir, veri setinden çıkarılmıştır.

Hat değerlerine bakıldığında 5,42,55,56 değişkenleri uç değerlerdir, veri setinden atılır.

Standartlaştırılmış artıklar ve Student True artıklar bulunmamıştır.

Aykırı değerler, uç değerler ve etkin gözlemler veri setinden çıkarılmıştır.

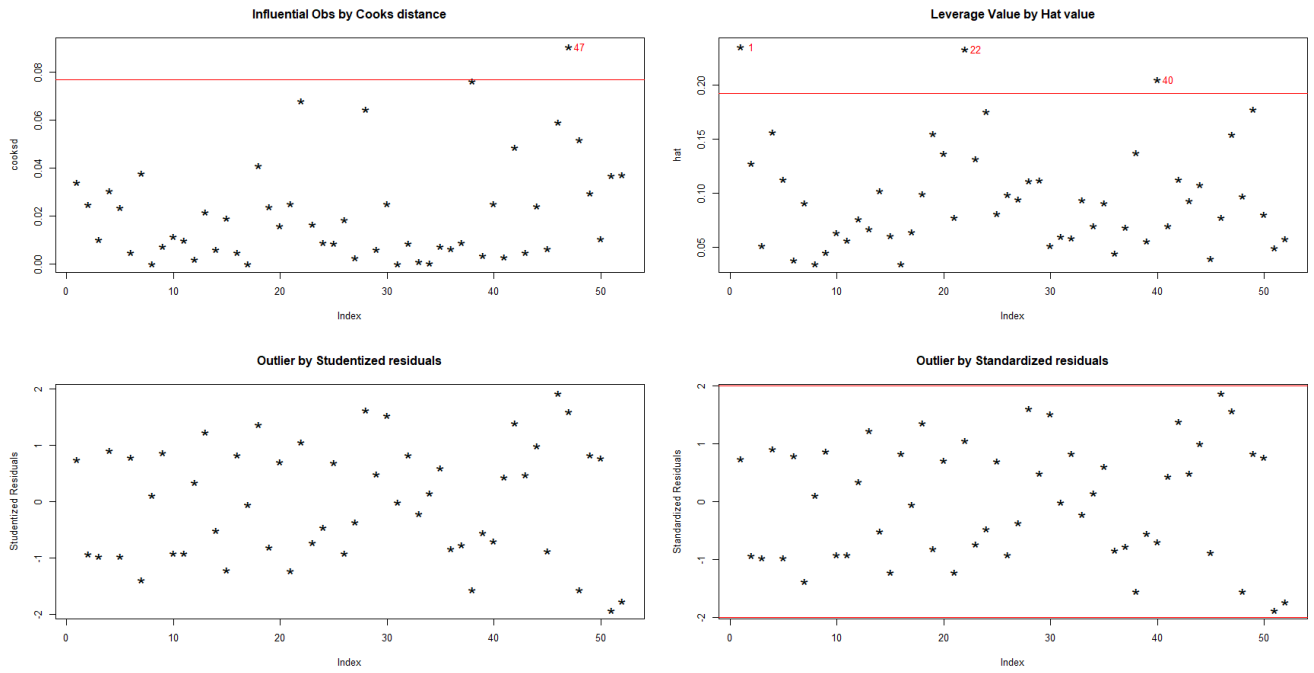
```
veri4 <- veri3[-c(5,39,42,55,56),]  
attach(veri4)  
lillie.test(veri4$y)  
sonuc4 <- lm(y~x1+x2+x3+x4,data=veri4)  
sonuc4
```

```
> lillie.test(veri4$y)  
  
Lilliefors (Kolmogorov-Smirnov) normality test  
data: veri4$y  
D = 0.1056, p-value = 0.1571
```

H0: Verilerin dağılışı ile normal dağılım arasında fark yoktur.

Verinin normalliği test edilmiştir ve $p=0.1571 > \alpha = 0,05$ olduğundan H0 hipotezi kabul edilir. Veriler normal dağılıma uygundur.

5.Artık İncelemesi



Cook uzaklığına bakıldığında 47. Gözlem etkin gözlemdir, veri setinden çıkarılır.

Hat değerlerine bakıldığında ; 1,22,40 değerleri uç değerlerdir, veri setinden çıkarılır.

Uç değerler ve etkin gözlemler veri setinden çıkarılmıştır.

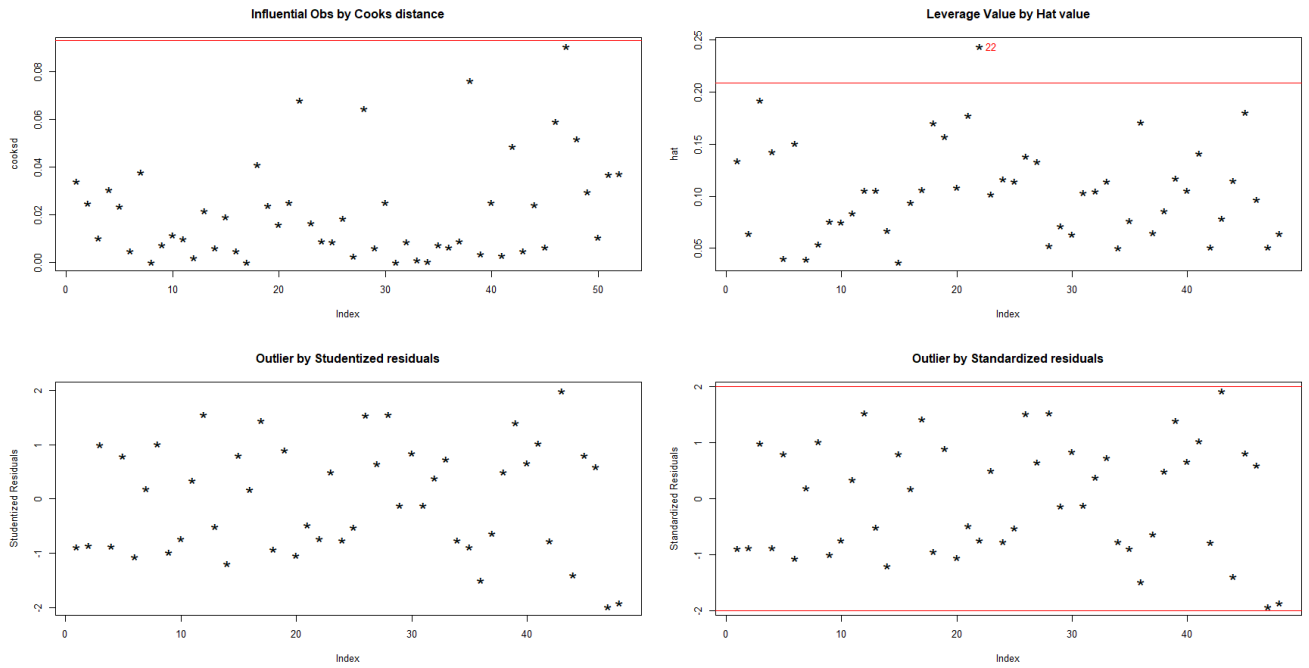
```
veri5 <- veri4[-c(1,22,40,47),]  
attach(veri5)  
lillie.test(veri5$y)  
sonuc5 <- lm(y~x1+x2+x3+x4,data=veri5)  
sonuc5
```

```
> lillie.test(veri5$y)  
  
Lilliefors (Kolmogorov-Smirnov) normality test  
  
data: veri5$y  
D = 0.10156, p-value = 0.2472
```

H0: Verilerin dağılışı ile normal dağılım arasında fark yoktur.

Verinin normalliği test edilmiştir ve $p=0.2472 > \alpha = 0,05$ olduğundan H0 hipotezi kabul edilir. Veriler normal dağılıma uygundur.

6.Artık İncelemesi



Cook değerine bakıldığında etkin gözlem bulunmamıştır.

Hat değerine bakıldığında 22. Gözlem uç değerdir, veri setinden çıkarılır.

Student True artıklar ve Standartlaştırılmış artıklar bulunmamıştır.

Uç değer veri setinden çıkarılmıştır.

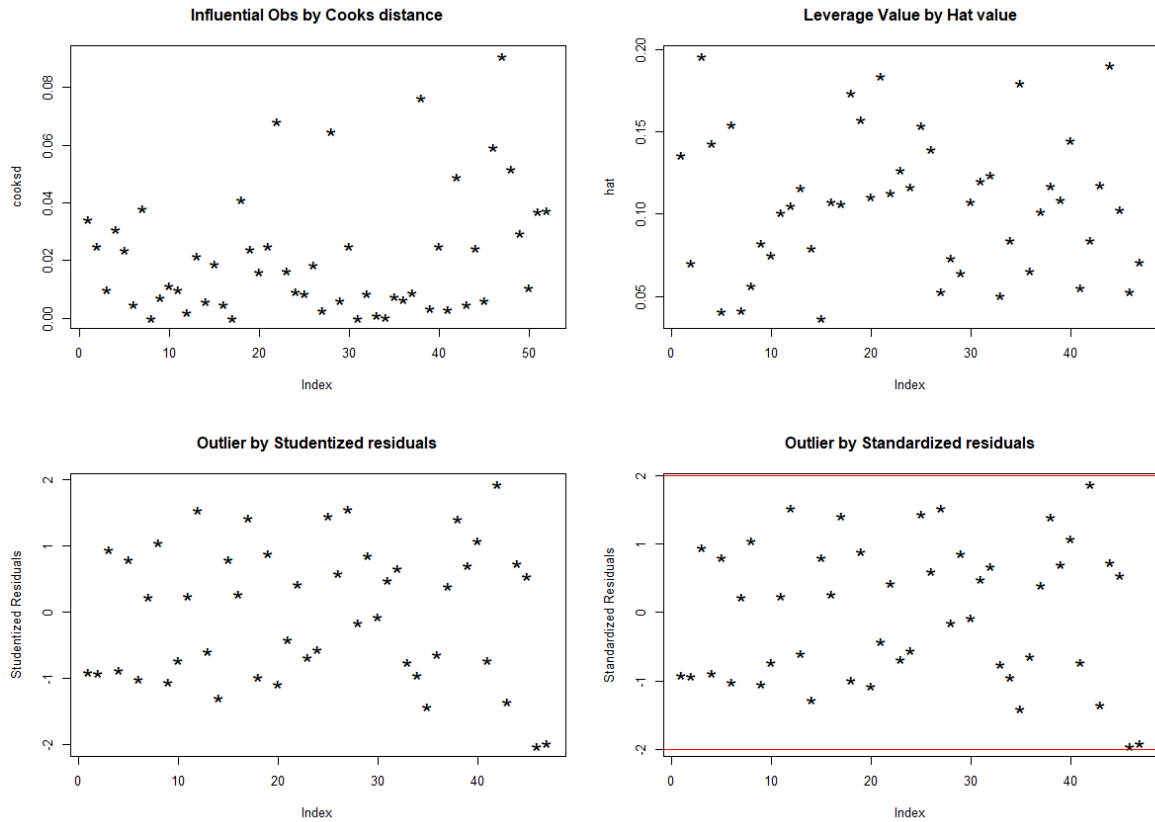
```
veri6 <- veri5[-c(22),]  
attach(veri6)  
shapiro.test(veri6$y)  
sonuc6 <- lm(y~x1+x2+x3+x4,data=veri6)  
sonuc6
```

```
> shapiro.test(veri6$y)  
  
Shapiro-Wilk normality test  
  
data: veri6$y  
W = 0.97056, p-value = 0.2783
```

H0: Verilerin dağılışı ile normal dağılım arasında fark yoktur.

Verinin normalliği test edilmiştir ve $p=0.2783 > \alpha = 0,05$ olduğundan H0 hipotezi kabul edilir. Veriler normal dağılıma uygundur.

7.Artık İncelemesi



Tüm aykırı değerler, uç değerler ,etkin gözlemler veri setinden atılmıştır. Artık incelemesi sona ermiştir.

```
> glimpse(veri6)
Observations: 47
Variables: 5
$ y <int> 449, 480, 773, 561, 702, 556, 666, 743, 491, 540, 540, 765, 432, 421, 720, 706, 880,...
$ x1 <dbl> 0.32, 0.32, 0.32, 0.32, 0.32, 0.36, 0.33, 0.33, 0.30, 0.34, 0.30, 0.34, 0.30, 0.31, ...
$ x2 <dbl> 60.6, 61.6, 63.3, 63.3, 62.0, 61.9, 62.3, 62.6, 62.6, 61.5, 61.2, 62.0, 61.0, 61.3, ...
$ x3 <dbl> 4.40, 4.40, 4.38, 4.36, 4.37, 4.57, 4.41, 4.40, 4.28, 4.49, 4.31, 4.48, 4.30, 4.36, ...
$ x4 <fct> SI1, SI1, VS2, VS2, VS2, SI1, VS2, VS2, VS2, SI1, SI1, SI1, SI1, SI1, VS2, VS2, VS2,...
```

Artık incelemesi sonunda veri setinde 47 gözlem kalmıştır.

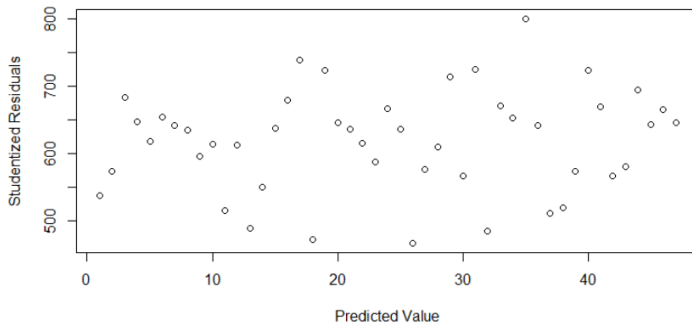
```
> sonuc6=lm(y~x1+x2+x3+x4,data=veri6)
> sonuc6

Call:
lm(formula = y ~ x1 + x2 + x3 + x4, data = veri6)

Coefficients:
(Intercept)      x1      x2      x3      x4vs2
   -7724.66   -6025.12    35.53   1826.74    85.92
```

Artık incelemesinin sonunda nitel değişken(X4) 2 düzeye inmiştir ve klavuz değişken SI1'dir.

Değişen Varyanslılık



Grafik incelendiğinde yapının rasgele olduğu görülmektedir. Grafik sonucunda değişen varyanslılık sorunu bulunmamaktadır.

```
> library(lmtest)
> bptest(sonuc6)

studentized Breusch-Pagan test

data: sonuc6
BP = 3.2902, df = 4, p-value = 0.5105
```

H0: Varyanslar homojendir.

Hs: Varyanslar homojen değildir.

Breusch and Pagan testi uygulanmıştır. Test sonucu p-value değeri $0.5105 > \alpha = 0.05$ olduğu için H0 hipotezi reddedilemez. Varyansların homojen olduğu 0.95 güven düzeyinde söylenebilir.

Öz İlişki Sorunu

```
> #Özilişki sorunu
> dwtest(sonuc6)

Durbin-Watson test

data: sonuc6
DW = 2.0286, p-value = 0.5365
alternative hypothesis: true autocorrelation is greater than 0
```

Model tablosunda test değeri $d=2.0286$ olarak bulunmuştur. Alternatif hipotez kurulmasında bu değer önemlidir. Eğer 2 ve 4 arasındaki değerlerde bir öz ilişki varsa bu öz ilişkinin negatif olacağı bilinmektedir. Alternatif hipotez negatif yönlü öz ilişki vardır şeklinde kurulmuştur.

H0: Öz ilişki yoktur.

Hs: Negatif yönlü öz ilişki vardır.

p value değeri $=0.5365 > \alpha = 0.05$ olduğu için H0 hipotezi reddedilemez. Öz ilişki sorununun olmadığı %95 güvenle söylenir.

Durbin-Watson değeri $=2.0286$; 2'ye yakın olduğu için p değerine bakılmadan da öz ilişki yoktur denebilir.

Çoklu Bağlantı Sorunu

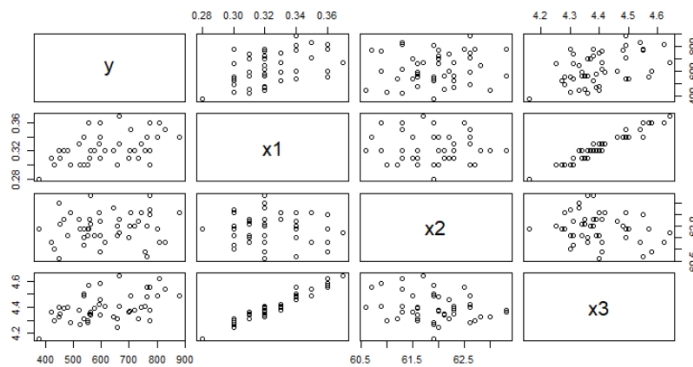
```
inf6 <- ls.diag(sonuc6)
inf6
```

```
attach(veri6)
pairs(veri6)
```

Modeldeki nitel değişken (X4) modelden atılmıştır. Modele ait korelasyon katsayıları aşağıdaki gibidir.

```
$correlation
      (Intercept)      x1      x2      x3      x4vs2
(Intercept)  1.0000000  0.8463748 -0.8732830 -0.87479089  0.20255884
x1           0.8463748  1.0000000 -0.5245304 -0.98111761 -0.08525960
x2          -0.8732830 -0.5245304  1.0000000  0.53406656 -0.46906842
x3          -0.8747909 -0.9811176  0.5340666  1.00000000  0.09832787
x4vs2        0.2025588 -0.0852596 -0.4690684  0.09832787  1.00000000
```

Bağımsız değişkenler arasındaki ilişki incelendiğinde X1 ve X3 arasında büyük bir ilişki olduğu görülmektedir. Bu durumda çoklu bağlantıdan şüphelenilmelidir.



Grafikte de görüldüğü üzere X1 ve X3 bağımsız değişkenleri arasında doğrusal ve güçlü bir ilişki olduğunu söyleyebiliriz. Grafik yardımıyla da çoklu bağlantının olduğu görülmektedir.

```
> library(DAAG)
> vif(sonuc6)
      x1      x2      x3      x4vs2
26.8700  2.2764 28.0910  1.6485
```

X1 ve X3 değişkenlerinin VIF değerlerinin 10'dan büyük olduğu görülmektedir dolayısıyla çoklu bağlantıdan etkilenmiştir.

```
> #install.packages("perturb")
> library(perturb)
> colldiag(model.matrix(sonuc6))
Condition
Index  Variance Decomposition Proportions
      intercept (Intercept) x1      x2      x3      x4vs2
1      1.000 0.000      0.000      0.000 0.000 0.000 0.000 0.005
2      4.364 0.000      0.000      0.000 0.000 0.000 0.000 0.600
3     44.246 0.000      0.000      0.000 0.034 0.001 0.000 0.012
4    298.228 0.000      0.000      0.000 0.093 0.381 0.090 0.375
5    744.725 0.000      0.000      0.000 0.856 0.541 0.889 0.008
6 2732835791259962.000 1.000      1.000      0.018 0.076 0.021 0.001
```

Koşul sayısı 30'dan büyük olduğu durumda çoklu bağlantıdan etkilenilmektedir. 3 tane çoklu bağlantı olduğu görülmektedir. Çoklu bağlantı sorunundan X1 ve X3 bağımsız değişkenlerinin etkilendiği görülmektedir dolayısıyla çoklu bağlantı yapısı X1 ve X3 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)
[1] 2.0000 0.9830 0.0183

```

Özdeğerlerden 0'a en yakın olanı 0.0183'tür ve 1 tane güçlü çoklu bağlantı olduğu söylenebilir.

```

> signif(sm$vectors,3)
      [,1] [,2] [,3]
[1,]  0.689 0.2080 0.694
[2,] -0.175 0.9770 -0.120
[3,]  0.704 0.0386 -0.710

```

0'a en yakın olan özdeğere ait olan 3. özvektör sütununa bakılmıştır. Bu sütunda görüldüğü üzere X1 ve X3 değişkenlerinin katsayı değerleri X2 den daha büyüktür ve 0'a yakın değildir bu nedenle X1 ve X2 katsayılarından yararlanarak;

$0.694 \cdot X1 - 0.710 \cdot X3 = 0$ çoklu bağlantı yapısı yazılmıştır.

```

> V<-sm$vectors
> t(V)%*%V
      [,1] [,2] [,3]
[1,] 1.000000e+00 -3.469447e-18 1.665335e-16
[2,] -3.469447e-18 1.000000e+00 6.938894e-18
[3,] 1.665335e-16 6.938894e-18 1.000000e+00
> #ilişki matrisi
> V %*% diag(sm$values) %*% t(V)
      [,1] [,2] [,3]
[1,] 1.00000000 -0.04182007 0.9676206
[2,] -0.04182007 1.00000000 -0.2069466
[3,] 0.96762063 -0.20694659 1.0000000

```

Yukarıdaki tablodaki kod ile özvektörler ve özdeğerlerin ilişki matrisi elde edilmiştir.

Model Anlamlılığı

```
> sonuc6 <- lm( y ~. , data=veri6)
> sonuc6

Call:
lm(formula = y ~ ., data = veri6)

Coefficients:
(Intercept)      x1      x2      x3      x4vs2
-7724.66    -6025.12     35.53    1826.74     85.92

> summary(sonuc6)

Call:
lm(formula = y ~ ., data = veri6)

Residuals:
    Min       1Q   Median       3Q      Max
-200.07  -87.41   23.64   82.98  189.05

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -7724.66    3950.42  -1.955  0.0572 .
x1          -6025.12    3959.68  -1.522  0.1356
x2              35.53     36.58   0.971  0.3369
x3           1826.74     795.63   2.296  0.0267 *
x4vs2         85.92     41.71   2.060  0.0456 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 105.5 on 42 degrees of freedom
Multiple R-squared:  0.3543,    Adjusted R-squared:  0.2928
F-statistic: 5.761 on 4 and 42 DF,  p-value: 0.0008655
```

H0: Model istatistiksel olarak anlamlı değildir.

HS: Model istatistiksel olarak anlamlıdır.

Tablodaki p value değeri=0.0008655 < 0.05 olduğu için H0 hipotezi reddedilir. Modelin anlamlı olduğu 0.05 anlamlılık düzeyinde söylenebilir. Bağımsız değişkenlerden X3 anlamlıdır. X1 ve X2 ise anlamsızdır. Bağımlı değişkendeki toplam değişimin %35'ini bağımsız değişkenler açıklamaktadır.

Model denklemi = $-7724.66 - 6025.12 \cdot X1 + 35.53 \cdot X2 + 1826.74 \cdot X3 + 85.92$

(3950.42) (3959.68) (36.58) (795.63)

Katsayı yorumları

Sabit terim: X1,X2,X3 bağımsız değişkenleri 0 olduğunda bağımlı değişken ortalama -7724.666'dır.

H0: Sabit terimin modele katkısı önemsizdir.

HS: Sabit terimin modele katkısı önemlidir.

Sabit terim için p value değeri=0.0572 > $\alpha = 0.05$ olduğu için H0 hipotezi kabul edilir. Sabit terimin modele katkısı önemsizdir.

X1 Katsayısının Yorumu : X2,X3 bağımsız değişkenleri sabit tutulduğunda ; X1 bağımsız değişkenindeki 1 birimlik artış bağımlı değişkeni ortalama 6025.12 birim azaltır.

H0: X1 değişkeninin modele katkısı önemsizdir.

HS: X1 değişkeninin modele katkısı önemlidir.

X1 değişkeni için tablodan bakıldığında p value değeri=0.1356 > $\alpha = 0.05$ olduğu görülmektedir dolayısıyla H0 hipotezi kabul edilir. X1 bağımsız değişkeninin modele katkısı önemsizdir.

X2 Katsayısının Yorumu : X1,X3 bağımsız değişkenleri sabit tutulduğunda ; X2 bağımsız değişkenindeki 1 birimlik artış bağımlı değişkeni ortalama 35.53 birim artırır.

H0: X2 değişkeninin modele katkısı önemsizdir.

HS: X2 değişkeninin modele katkısı önemlidir.

X2 değişkeni için tablodan bakıldığında p value değeri=0.3369 > $\alpha = 0.05$ olduğu görülmektedir dolayısıyla H0 hipotezi kabul edilir. X2 bağımsız değişkeninin modele katkısı önemsizdir.

X3 Katsayısının Yorumu :

H0: X3 değişkeninin modele katkısı önemsizdir.

HS: X3 değişkeninin modele katkısı önemlidir.

X3 değişkeni için tablodan bakıldığında p value değeri=0.0267 < $\alpha = 0.05$ 'dır ve H0 hipotezi reddedilir. X1 bağımsız değişkeninin modele katkısı önemlidir. X1,X2 bağımsız değişkenleri sabit tutulduğunda ; X3 bağımsız değişkenindeki 1 birimlik artış bağımlı değişkeni ortalama 1826.74 birim artırır.

X4 Katsayısının Yorumu :

H0: SI1 değişkeni ile VS2 değişkeni arasında fark yoktur.

HS: SI1 değişkeni ile VS2 değişkeni arasında fark vardır.

Tabloya bakıldığında p value değeri=0.0456 < $\alpha = 0.05$ olduğu için H0 hipotezi reddedilir. SI1 değişkeni ile VS2 değişkeni arasında fark olduğu %95 güvenle söylenebilir.

Uyum Kestirimi

```
> #uyum kestirimi
> predict(sonuc6,data.frame(x1=0.32,x2=63.3,x3=4.36,x4="VS2"), interval='confidence')
      fit      lwr      upr
1 646.8018 566.2622 727.3415
```

Veri setinin içinden 9. satırdaki X1,X2,X3,X4 değişkenleri koda girilmiştir , tahmin değeri 646.8018 olarak bulunmuştur. X1, X2, X3, X4 değerleri için bağımlı değişkenin ortalama 52.26374 ile 53.33968 aralığında olduğu söylenebilir.

Ön kestirim

```
> predict(sonuc6,data.frame(x1=1,x2=62,x3=5,x4="SI1"), interval='confidence')
      fit      lwr      upr
1 -2413.277 -6878.84 2052.286
```

Veri setinde bulunmayan X1, X2, X3, X4 değerleri girilmiştir, tahmin değeri -2413.277 olarak bulunmuştur. X1,X2,X3,X4 değişkenleri için bağımlı değişkenin ortalama -6878.84 ile 2052.286 arasında olduğu söylenebilir.

Regresyon Katsayıları İçin Güven Aralıkları

```
> confint(sonuc6,level=.99)
              0.5 %      99.5 %
(Intercept) -18383.14708 2933.8187
x1           -16708.59095 4658.3471
x2            -63.15692  134.2158
x3           -319.91713 3973.3907
x4vs2        -26.61429  198.4528
```

Sabit Terim Güven Aralığı Yorumu: X1,X2,X3 bağımsız değişkenleri 0 olduğunda bağımlı değişkenin ortalama -18383.14 ile 2933.8187 arasında olduğu %99 güvenle söylenir.

X1 için Güven Aralığı Yorumu : X2,X3 bağımsız değişkenleri sabit tutulduğunda X1 bağımsız değişkeni 1 birim arttığında bağımlı değişken ortalama bağımlı değişkenin ortalama -18383.14 ile 2933.8187 arasında olduğu %99 güvenle söylenir.

X2 için Güven Aralığı Yorumu : X1,X3 bağımsız değişkenleri sabit tutulduğunda X2 bağımsız değişkeni 1 birim arttığında bağımlı değişken ortalama bağımlı değişkenin ortalama -63.15692 ile 134.2158 arasında olduğu %99 güvenle söylenir.

X3 için Güven Aralığı Yorumu: X1,X2 bağımsız değişkenleri sabit tutulduğunda X3 bağımsız değişkeni 1 birim arttığında bağımlı değişken ortalama bağımlı değişkenin ortalama -319.91713 ile 3973.3907 arasında olduğu %99 güvenle söylenir.

Değişken Seçimi Yöntemleri

İleriye Doğru Seçim Yöntemi:

```
> library(stats)
> #değişken seçimi
> #ileriye doğru
> library(stats)
> lm.null <- lm(veri6$y ~ 1)
> forward <- stepAIC(lm.null, veri6$y~ x1+x2+x3+x4, direction = "forward")
Start: AIC=455.22
veri6$y ~ 1

      Df Sum of Sq  RSS   AIC
+ x3    1   155701 568652 445.84
+ x1    1   147287 577066 446.53
+ x4    1    37912 686441 454.69
<none>                 724353 455.22
+ x2    1     9407 714947 456.60

Step: AIC=445.84
veri6$y ~ x3

      Df Sum of Sq  RSS   AIC
+ x4    1    74678 493974 441.22
+ x2    1   33343 535309 445.00
<none>                 568652 445.84
+ x1    1      61 568591 447.84

Step: AIC=441.22
veri6$y ~ x3 + x4

      Df Sum of Sq  RSS   AIC
<none>                 493974 441.22
+ x1    1   15738 478237 441.70
+ x2    1     461 493513 443.18
```

Birinci adımda modele X3 değişkeni girmiştir. Daha sonra X4 değişkeni modele girmiştir. Sonuç olarak X3 Ve X4 'ün olduğu modelin anlamlı model olduğu görülmektedir.

```
> summary(forward)

Call:
lm(formula = veri6$y ~ x3 + x4)

Residuals:
    Min       1Q   Median       3Q      Max
-176.66  -82.59   -0.19   87.37  190.51

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2231.75     679.62  -3.284 0.002013 **
x3           635.26     153.43   4.140 0.000154 ***
x4VS2        85.63      33.20   2.579 0.013325 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 106 on 44 degrees of freedom
Multiple R-squared:  0.318,    Adjusted R-squared:  0.2871
F-statistic: 10.26 on 2 and 44 DF,  p-value: 0.0002201
```

En iyi model: $Y_i = -2231.75 + 635.26 \cdot X_3 + 85.63 \cdot X_{4VS2} \pm 106$

(679.62) (153.43) (33.20)

p değerinden görüldüğü üzere modeldeki değişkenler %95 güven düzeyinde anlamlıdır.

H0: Model anlamsızdır.

HS: Model anlamlıdır.

$p = 0.0002201 < 0.05$ olduğundan modelin de anlamlı olduğu görülmektedir.

Geriye Doğru Seçim Yöntemi:

```
> backward<-step(sonuc6,direction="backward")
Start: AIC=442.66
y ~ x1 + x2 + x3 + x4

   Df Sum of Sq  RSS   AIC
- x2    1    10508 478237 441.70
<none>                 467729 442.66
- x1    1    25784 493513 443.18
- x4    1    47257 514986 445.18
- x3    1    58705 526434 446.22

Step: AIC=441.7
y ~ x1 + x3 + x4

   Df Sum of Sq  RSS   AIC
- x1    1    15738 493974 441.22
<none>                 478237 441.70
- x3    1    49209 527446 444.31
- x4    1    90355 568591 447.84

Step: AIC=441.22
y ~ x3 + x4

   Df Sum of Sq  RSS   AIC
<none>                 493974 441.22
- x4    1    74678 568652 445.84
- x3    1   192467 686441 454.69
```

Geriye doğru seçim yönteminin özelliğinden tüm değişkenler modelde olarak başlar. İlk model tüm bağımsız değişkenlerin modelde bulunduğu durumdur. İkinci modelde X2 değişkeni modelden çıkmış, üçüncü modelde ise X1 değişkeni modelden çıkmıştır. Son modelde yalnızca X3 ve X4 değişkenleri modelde kalmıştır.

```
> summary(backward)

Call:
lm(formula = y ~ x3 + x4, data = veri6)

Residuals:
    Min       1Q   Median       3Q      Max
-176.66  -82.59   -0.19   87.37  190.51

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2231.75     679.62  -3.284 0.002013 **
x3           635.26     153.43   4.140 0.000154 ***
x4VS2        85.63      33.20   2.579 0.013325 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 106 on 44 degrees of freedom
Multiple R-squared:  0.318,    Adjusted R-squared:  0.2871
F-statistic: 10.26 on 2 and 44 DF,  p-value: 0.0002201
```

En iyi model: $Y_i = -2231.75 + 635.26 \cdot X_3 + 85.63 \cdot X_{4VS2} \pm 106$

(679.62) (153.43) (33.20)

Aynı bağımsız değişkenler ile model kurulduğundan ileri ve geriye doğru seçim yöntemleri aynı sonuçları vermiştir.

H_0 : Model anlamsızdır.

H_S : Model anlamlıdır.

$p = 0.0002201 < 0.05$ olduğundan modelin anlamlı olduğu %95 güvenle söylenir. Ayrıca katsayıların p value değerlerine bakıldığında katsayıların da anlamlı olduğu görülmektedir.

Adımsal Regresyon Yöntemi:

```
> #adımsal
> library(MASS)
> step.model <- stepAIC(sonuc6, direction = "both", trace = FALSE)
> step.model

Call:
lm(formula = y ~ x3 + x4, data = veri6)

Coefficients:
            x3          x4VS2
(Intercept) -2231.75      635.26      85.63

> summary(step.model)

Call:
lm(formula = y ~ x3 + x4, data = veri6)

Residuals:
    Min       1Q   Median       3Q      Max
-176.66  -82.59   -0.19   87.37  190.51

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2231.75     679.62  -3.284 0.002013 **
x3           635.26     153.43   4.140 0.000154 ***
x4VS2        85.63      33.20   2.579 0.013325 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 106 on 44 degrees of freedom
Multiple R-squared:  0.318,    Adjusted R-squared:  0.2871
F-statistic: 10.26 on 2 and 44 DF,  p-value: 0.0002201
```

İlk modele X3 ve X4 değişkenleri alınarak başlanmıştır. Diğer bağımsız değişkenler anlamlı olmadığından yöntem tamamen ileri doğru seçim yöntemi gibi sonra ermiştir. En iyi model X3 ve X4 değişkenlerinin olduğu modeldir.

En iyi model: $Y_i = -2231.75 + 635.26 \cdot X_3 + 85.63 \cdot X_4 \pm 106$

(679.62) (153.43) (33.20)

Ridge Regresyonu

```
> ridge <- lm.ridge(veri6$y~x1+x2+x3 ,lambda = seq(0,1,0.05),data=veri6)
> matplot(ridge$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
x1	-107.39871	-99.73173	-92.86681	-86.68465	-81.08847	-75.99898	-71.35056	-67.08846	-63.16663
x2	45.00028	43.64530	42.42903	41.33083	40.33399	39.42479	38.59189	37.82584	37.11866
x3	170.79074	162.91829	155.86563	149.51084	143.75501	138.51709	133.73002	129.33783	125.29346

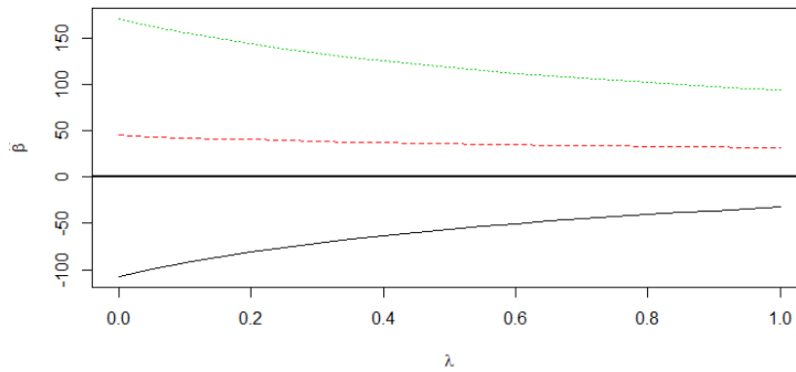
```
> ridge$coef
```

	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85
x1	-59.54605	-56.19345	-53.08030	-50.18196	-47.47707	-44.94701	-42.57546	-40.34809	-38.25224
x2	36.46361	35.85496	35.28775	34.75773	34.26122	33.79499	33.35622	32.94243	32.55144
x3	121.55704	118.09456	114.87685	111.87874	109.07838	106.45676	103.99720	101.68504	99.50734

```
> ridge$coef
```

	0.90	0.95	1.00
x1	-36.27669	-34.41145	-32.64762
x2	32.18130	31.83029	31.49688
x3	97.45265	95.51074	93.67253

```
> select(ridge)
modified HKB estimator is 0.2802879
modified L-W estimator is 2.688535
smallest value of GCV at 0.5
> ridge$coef[,ridge$lam == 0.4]
      x1      x2      x3
-63.16663  37.11866 125.29346
```



Lambda parametresine göre katsayılar değişmektedir. Hızlı azalış ve artışların bittiği yerde bir lambda parametresi belirlenerek katsayı kestirimleri elde edilir ve model kurulur ancak grafikte birden artış ya da birden azalış görülmemektedir dolayısıyla çoklu bağlantı sorunu yoktur ayrıca lamdanın 0,4 olarak alındığı durumun sonuçlarına göre katsayı kestirimleri yukarıdaki çıktıda elde edilmiştir. Grafığe göre tüm değişkenler önemli çıkmıştır. Ridge fonksiyonu çoklu bağlantıyı yakalayamamıştır.