

DSM 5007- DENETİMLİ İSTATİSTİKSEL ÖĞRENME- 1.ÖDEV

Buse Nur Baltacıoğlu, Eray Kaan Güngörür, Halil Durmaz, Yunus Sur.

22 12 2020

1 - Veri setini tanıtınız.

Ot çeşidine göre mera kira yapısını araştırmak için Minnesota'daki çeşitli ilçelerden aşağıdaki veriler toplandı.

<https://extension.umn.edu/pasture-based-dairy/grazing-and-pasture-management-cattle>

```
library(readr)
library(ggplot2)
library(dplyr)
library(broom)
library(ggpubr)
library(corrplot)
library(ISLR)
library(car)
library(olsrr)
library(lmtest)
library(GGally)
library(caret)
library(tidyverse)
library(AppliedPredictiveModeling)
library(pls)
library(elasticnet)
library(broom)
library(glmnet)
library(MASS)
library(PerformanceAnalytics)
library(funModeling)
library(Matrix)
library(forecast)
```

x_1 : ekilebilir arazi dönümü başına kira (\$)

x_2 : mil^2 başına süt ineği oranı

x_3 : otlak ve ekilebilir arazi arasındaki fark

y : çim çeşiti için dönüm başına kira bedeli

```
arazi <- read_table2("https://raw.githubusercontent.com/halil-durmaz/dataset/main/arazi.txt", col_names =
```

```
## Parsed with column specification:
## cols(
##   X1 = col_double(),
##   X2 = col_double(),
##   X3 = col_double(),
##   X4 = col_double(),
##   X5 = col_double()
## )
```

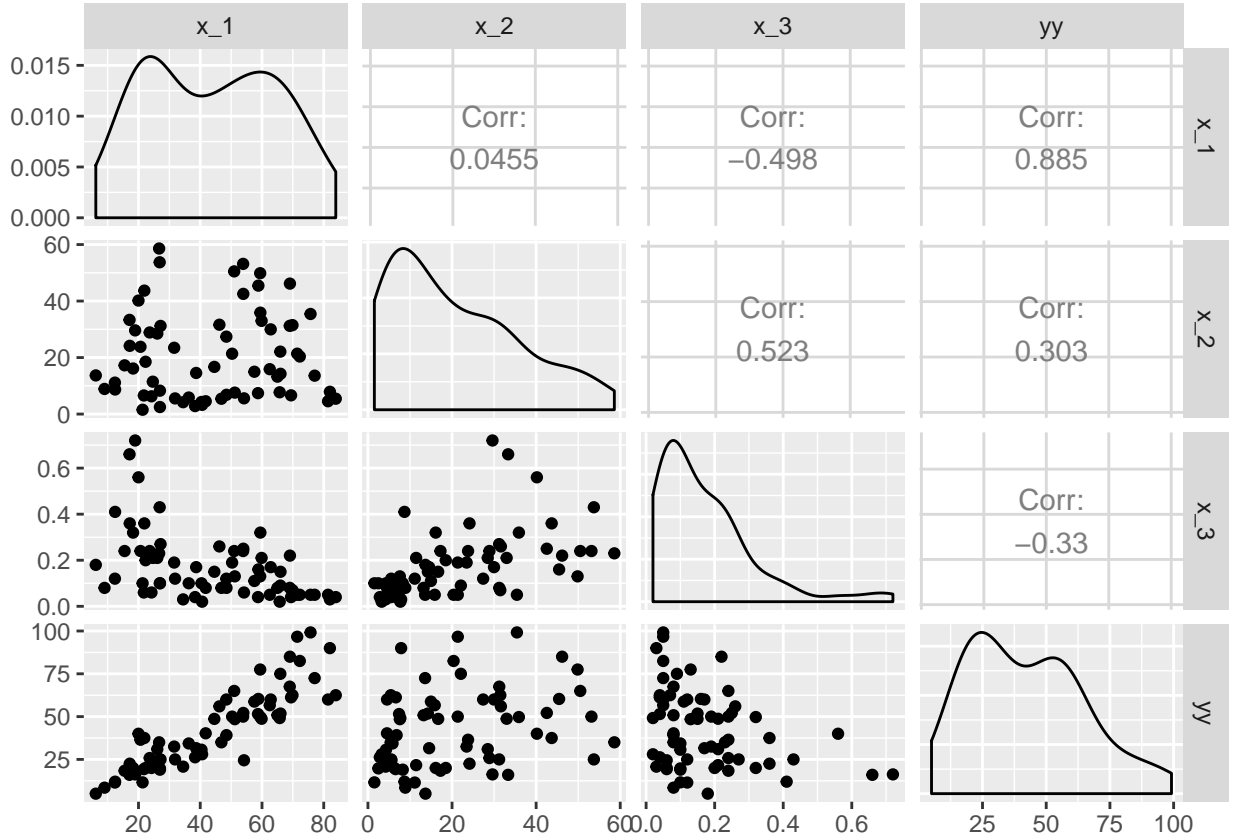
```
head(arazi)
```

```
## # A tibble: 6 x 5
##       X1      X2      X3      X4      X5
##   <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     1  15.5  17.2  0.24  18.4
## 2     2  22.3  18.5  0.2   20
## 3     3  12.4  11.1  0.12  11.5
## 4     4  31.8   5.54  0.12  25
## 5     5  83.9   5.44  0.04  62.5
## 6     6  72.2  20.4   0.05  82.5
```

```
df<-arazi[,-1]
colnames(df)<-c("x_1", "x_2", "x_3", "yy")
head(df)
```

```
## # A tibble: 6 x 4
##       x_1    x_2    x_3    yy
##   <dbl> <dbl> <dbl> <dbl>
## 1  15.5  17.2   0.24  18.4
## 2  22.3  18.5   0.2   20
## 3  12.4  11.1   0.12  11.5
## 4  31.8   5.54   0.12  25
## 5  83.9   5.44   0.04  62.5
## 6  72.2  20.4   0.05  82.5
```

```
attach(df)
ggpairs(df)
```



H_0 : Yy değişkeni normal dağılır.

H_1 : Yy değişkeni normal dağılmaz.

```
shapiro.test(yy)
```

```
##
## Shapiro-Wilk normality test
##
## data: yy
## W = 0.95554, p-value = 0.01758
```

(p-value = 0.01758) < 0.05 olduğu için H_0 red çim çeşiti için dönüm başına kira bedeli değişkeni normal dağılmaz. Grafikten ve normallik testiyle kontrol ettiğimiz bağımlı değişkenimiz doğrusal regresyon varsayımlarından biri olan bağımlı değişkenin normal dağılması varsayımını sağlayamamaktadır. Bu sebepten bağımlı değişkenimize karekök dönüşümü yaptık.

```
y1<-sqrt(yy)
y1
```

```
## [1] 4.287190 4.472136 3.391165 5.000000 7.905694 9.082951 5.000000 5.538050
## [9] 3.464102 7.826238 7.745967 8.215838 5.567764 7.745967 8.514693 7.767239
## [17] 7.053368 2.915476 6.041523 7.745967 4.031129 7.071068 3.405877 5.916080
## [25] 8.660254 5.617829 6.964194 8.803408 4.655105 4.444097 7.483315 5.000000
## [33] 6.324555 7.527948 7.196527 9.832090 7.129516 5.859181 6.982120 5.079370
## [41] 4.472136 4.000000 6.976389 4.558509 5.700877 4.358899 7.176350 7.012132
## [49] 9.219544 7.665507 4.396590 2.236068 8.062258 4.472136 7.905694 5.916080
## [57] 9.958414 6.344289 6.258594 6.123724 5.123475 7.220803 4.743416 9.486833
## [65] 5.291503 7.071068 4.949747
```

```
df1<-data.frame(y1,x_1,x_2,x_3)
head(df1)
```

```
##           y1    x_1    x_2    x_3
## 1 4.287190 15.50 17.25 0.24
## 2 4.472136 22.29 18.51 0.20
## 3 3.391165 12.36 11.13 0.12
## 4 5.000000 31.84  5.54 0.12
## 5 7.905694 83.90  5.44 0.04
## 6 9.082951 72.25 20.37 0.05
```

```
attach(df1)
```

```
## The following object is masked _by_ .GlobalEnv:
##
##      y1
##
## The following objects are masked from df:
##
##      x_1, x_2, x_3
```

1.1 Eğitim ve test verisi

```
sample <- floor(0.95 * nrow(df1))
set.seed(2468)
train_ind <- sample(nrow(df1), size = sample, replace = FALSE)
train <- df1[train_ind, ]
test <- df1[-train_ind, ]
```

```
dim(train)
```

```
## [1] 63  4
```

```
dim(test)
```

```
## [1] 4 4
```

```
str(test)
```

```
## 'data.frame':  4 obs. of  4 variables:
## $ y1 : num  8.66 6.32 5.7 4.36
## $ x_1: num  65.9 20 31.6 26.9
## $ x_2: num  22.1 40.18 23.47 8.28
## $ x_3: num  0.09 0.56 0.19 0.1
```

```
attach(train)
```

```
## The following object is masked _by_ .GlobalEnv:
##
##      y1
##
## The following objects are masked from df1:
##
##      x_1, x_2, x_3, y1
##
## The following objects are masked from df:
##
##      x_1, x_2, x_3
```

```
y<-train$y1
x1<-train$x_1
x2<-train$x_2
x3<-train$x_3
```

Veri setimiz 67 gözlem ve 4 değişkenden oluşmaktadır. Veri setimizi eğitim ve test verisi olarak ayırdık eğitim verisi 63-4 ve test verisi 4-4 optimal modele karar verdikten sonra test verisini tahminlemek için kullanıcaz.

2 - Tanımlayıcı istatistikleri grafiklerle destekleyerek elde ediniz ve yorumlayınız.

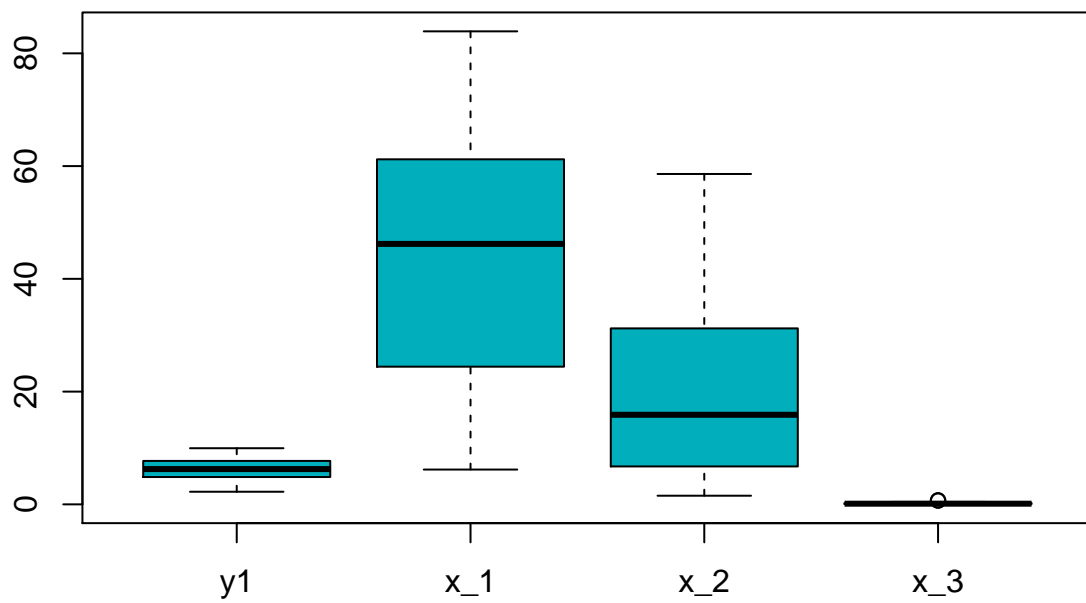
```
summary(train)
```

```
##           y1           x_1           x_2           x_3
##  Min.      :2.236   Min.      : 6.17   Min.      : 1.530   Min.      :0.0200
## 1st Qu.:4.847   1st Qu.:24.42   1st Qu.: 6.725   1st Qu.:0.0600
##  Median :6.259   Median :46.20   Median :15.890   Median :0.1200
##  Mean    :6.269   Mean    :44.28   Mean    :20.376   Mean    :0.1656
## 3rd Qu.:7.706   3rd Qu.:61.20   3rd Qu.:31.215   3rd Qu.:0.2350
##  Max.    :9.958   Max.    :83.90   Max.    :58.600   Max.    :0.7200
```

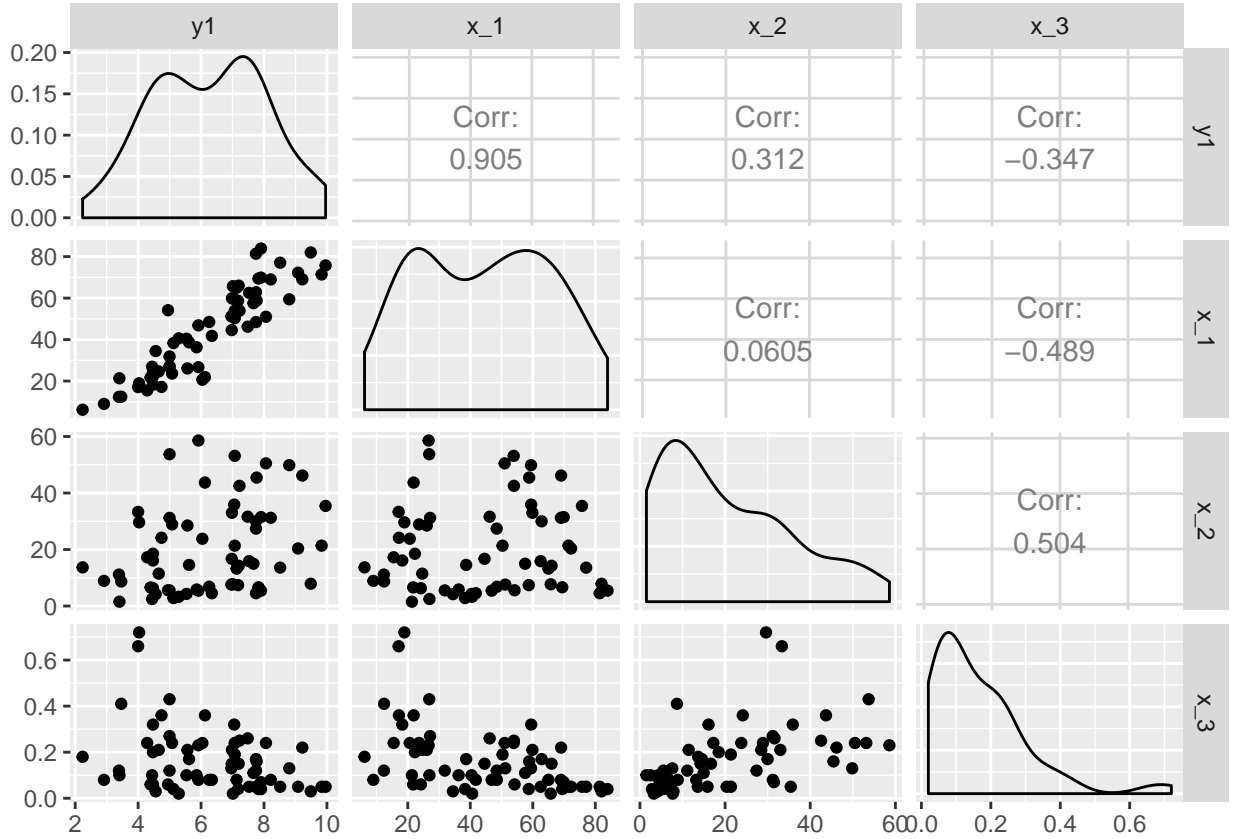
```
apply(train, 2, sd)
```

```
##           y1           x_1           x_2           x_3
## 1.8100869 21.3893132 15.5409567 0.1398707
```

```
boxplot(train, col = "#00AFBB")
```



```
ggpairs(train)
```



x_1 : Ekilebilir arazinin dönümü başına kira 6.17 dolar ile 83.90 dolar arasında 44.28 ortalama ve 21.39 standart sapmayla değişmektedir. İki tepeli bir dağılım olduğunu söyleyebiliriz.

x_2 : Mil^2 başına süt ineği oranı 1.53 ile 58.60 arasında 20.38 ortalama ve 15.54 standart sapmayla değişmektedir. Bu değişkenimiz sağa çarpık bir dağılıma sahiptir.

x_3 : Otlak ve ekilebilir arazi arasındaki fark 0.02 ile 0.72 arasında 0.17 ortalama ve 0.14 standart sapmayla değişmektedir. Bu değişkenimiz de sağa çarpık ve ağır kuyruklu olduğunu söyleyebiliriz.

y: Çim çeşidi için dönüm başına kira bedeli 2.24 ile 9.96 arasında 6.27 ortalama ve 1.81 standart sapmayla değişmektedir.

H_0 : Y değişkeni normal dağılır.

H_1 : Y değişkeni normal dağılmaz.

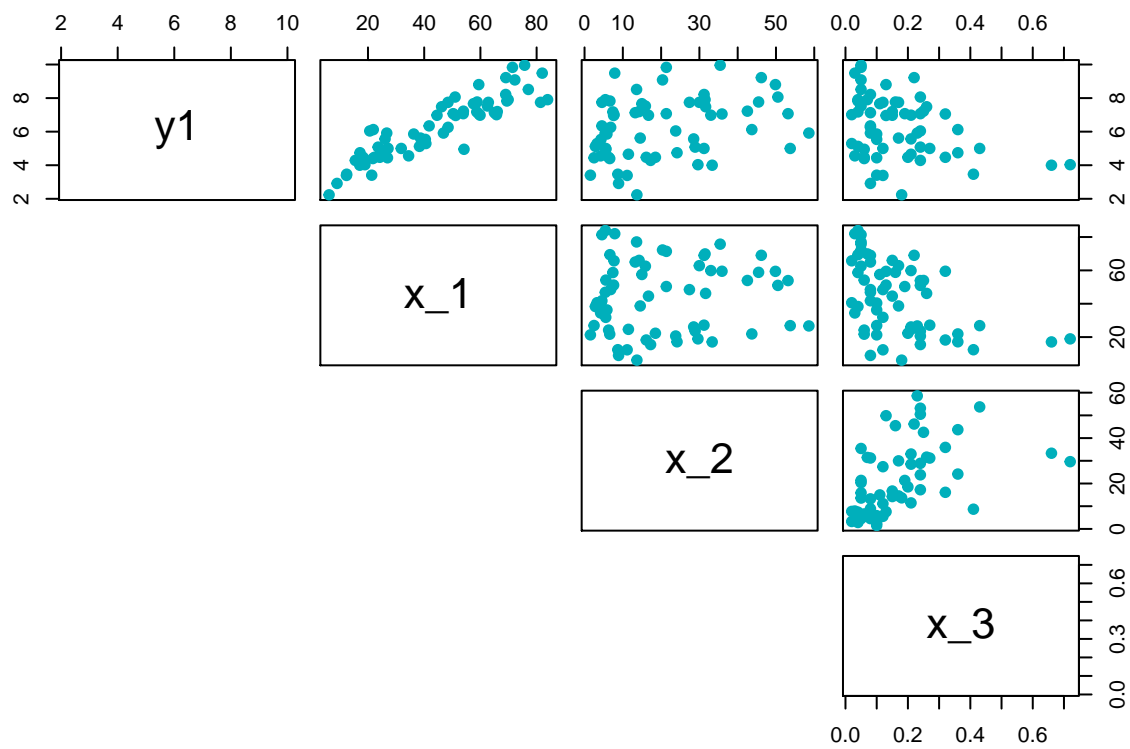
```
shapiro.test(y)
```

```
##
## Shapiro-Wilk normality test
##
## data: y
## W = 0.97919, p-value = 0.3629
```

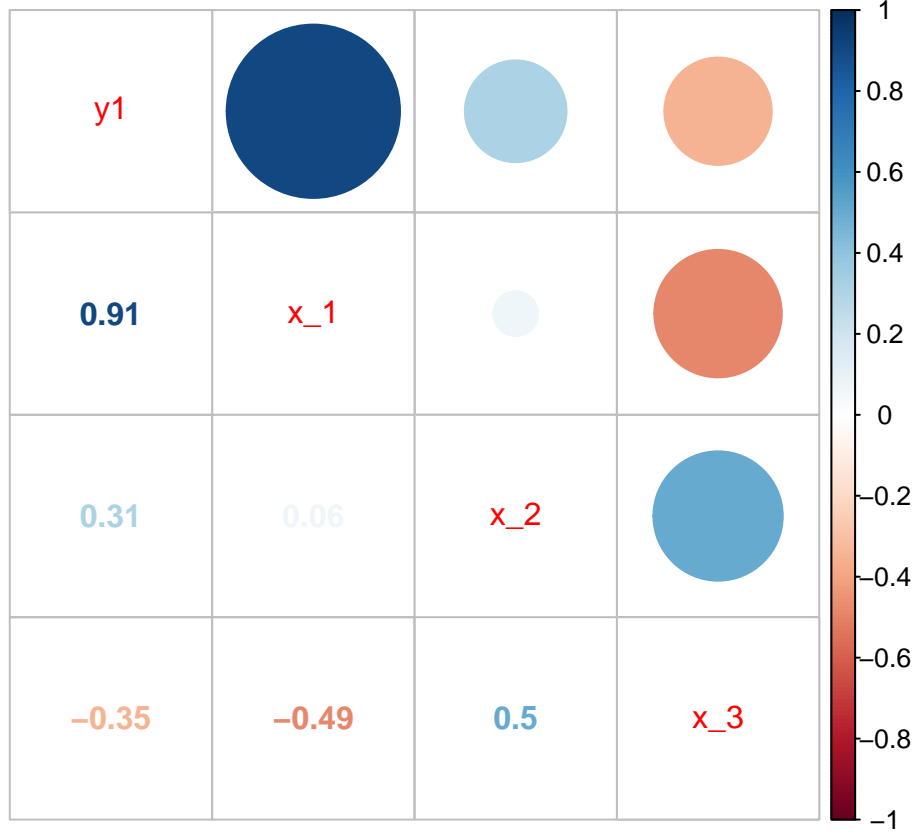
(p-value = 0.3629) > 0.05 olduğu için H_0 reddedilemez y bağımlı değişkeni normal dağılır.

3 - Korelasyon matrisini ve ilgili görselleri oluşturarak, yorumlayınız.

```
pairs(train[,1:4], pch=19, col="#00AFBB", lower.panel = NULL)
```



```
cor_df<-cor(train)
corrplot::corrplot.mixed(cor_df)
```

Korelasyon matrisine baktığımızda;

-Çim çeşidi için dönüm başına kira bedeli(y) ile ekilebilir arazinin dönümü başına kirası(x_1) arasında doğrusal pozitif yönlü güçlü bir ilişki bulunmaktadır.

-Çim çeşidi için dönüm başına kira bedeli(y) ile mil^2 başına süt ineği oranı(x_2) arasında doğrusal pozitif yönlü zayıf bir ilişki bulunmaktadır.

-Çim çeşidi için dönüm başına kira bedeli(y) ile otlak ve ekilebilir arazi arasındaki fark(x_3) değişkeni arasında doğrusal negatif yönlü zayıf bir ilişki bulunmaktadır.

-Ekilebilir arazinin dönümü başına kirası(x_1) ile mil^2 başına süt ineği oranı(x_2) arasında doğrusal bir ilişki bulunmamaktadır.

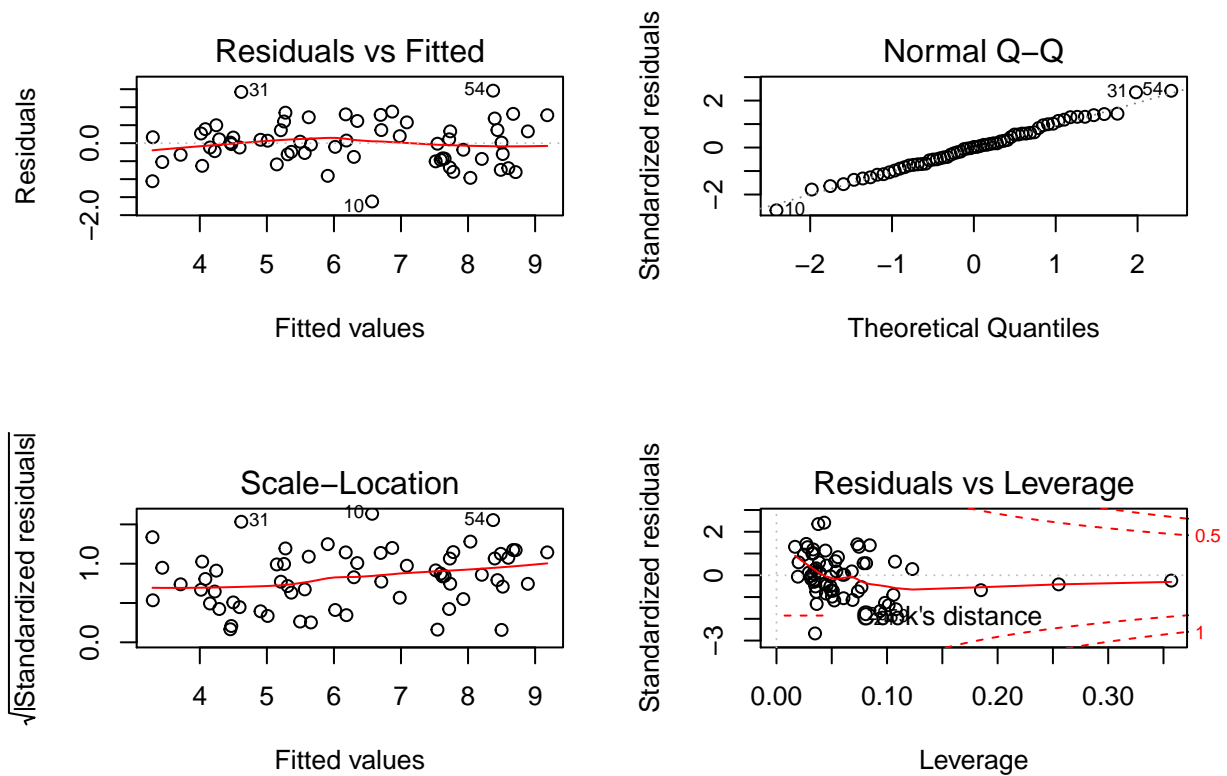
-Ekilebilir arazinin dönümü başına kirası(x_1) ile otlak ve ekilebilir arazi arasındaki fark(x_3) değişkeni arasında doğrusal negatif yönlü bir ilişki bulunmaktadır.

- Mil^2 başına süt ineği oranı(x_2) ile otlak ve ekilebilir arazi arasındaki fark(x_3) değişkeni arasında doğrusal pozitif yönlü bir ilişki bulunmaktadır.

4 - Çoklu doğrusal regresyon modelini elde ediniz.

```
model<-lm(y~x1+x2+x3)
```

```
par(mfrow=c(2,2))
plot(model)
```



```
par(mfrow=c(1,1))
```

4.a - Anlamlı katsayıları belirleyiniz.

$H_0: \beta_j=0$

H_{1j} : En az bir β_j farklıdır, $j=1,2,3$

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: y
##      Df Sum Sq Mean Sq F value    Pr(>F)
## x1      1 166.478  166.478  437.4445 < 2.2e-16 ***
## x2      1  13.449   13.449   35.3403 1.597e-07 ***
## x3      1   0.756    0.756    1.9875  0.1639
## Residuals 59  22.454    0.381
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3)
##
## Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -1.61964 -0.42605  0.01484  0.36458  1.45896
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.573482   0.269312   9.556 1.37e-13 ***
## x1           0.071380   0.004600  15.518 < 2e-16 ***
## x2           0.035550   0.006393   5.561 6.83e-07 ***
## x3          -1.146124   0.812981  -1.410   0.164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6169 on 59 degrees of freedom
## Multiple R-squared:  0.8895, Adjusted R-squared:  0.8838
## F-statistic: 158.3 on 3 and 59 DF,  p-value: < 2.2e-16
```

-F istatistiğine karşılık gelen $2.2e-16 < 0.05$ olduğu için %95 güven düzeyinde H_0 red edilir; en az bir katsayı model için anlamlıdır. Model geçerlidir.

Bu modelde β_1 ve β_2 katsayıları %95 güvenle anlamlı çıkmıştır.

$-\beta_0$: Bağımsız değişkenlerin değerleri 0 olduğunda çim çeşidi için dönüm başına kira bedelinin aldığı ortalama değer 2.57 dolardır.

$-\beta_1$: Diğer değişkenler modelde ve sabitken ekilebilir arazi dönümü başına kira(x_1), 1 \$ arttığında çim çeşidi için dönüm başına kira bedelini ortalama 0.07 dolar artırır.

$-\beta_2$: Diğer değişkenler modelde ve sabitken mil^2 başına süt ineği oranı(x_2), 1 birim arttığında çim çeşidi için dönüm başına kira bedelini ortalama 0.04 dolar artırır.

$-\beta_3$: Diğer değişkenler modelde ve sabitken otlak ve ekilebilir arazi arasındaki fark(x_3), 1 birim arttığında çim çeşidi için dönüm başına kira bedelini ortalama 1.15 dolar azaltır.

4.b - R^2 ve adj- R^2 katsayılarını yorumlayınız.

```
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -1.61964 -0.42605  0.01484  0.36458  1.45896
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.573482   0.269312   9.556 1.37e-13 ***
## x1           0.071380   0.004600  15.518 < 2e-16 ***
## x2           0.035550   0.006393   5.561 6.83e-07 ***
## x3          -1.146124   0.812981  -1.410   0.164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6169 on 59 degrees of freedom
## Multiple R-squared:  0.8895, Adjusted R-squared:  0.8838
## F-statistic: 158.3 on 3 and 59 DF,  p-value: < 2.2e-16
```

-Çim çeşidi için dönüm başına kira bedelindeki değişim, bağımsız değişkenler(x_1, x_2, x_3) tarafından %88.95 oranında açıklanmaktadır.

-R-Sq (adj) değeri %88.38 olarak elde edilmiş R-Sq %88.95 değerine çok yakın olduğu yani modelde bulunan değişkenlerin model için anlamlı olduğu söylenebilir.

4.c - VIF değerlerini hesaplatarak yorumlayınız.

```
vif(model)
```

```
##          x1          x2          x3
## 1.577060 1.608063 2.106548
```

-Bağımsız değişkenler birbirinden bağımsız olmalıdır. Bu modelde vif değerleri 5'ten küçük çıkmıştır. Korelasyon matrisinde de görüldüğü gibi aralarında doğrusal bir bağlantı yoktur.

5 - Değişken seçim yöntemlerini kullanarak (Değişken seçim yöntemleri ile ana etkenlere karar verilmektedir. Veri seti içinde karesel, kübik, vb. etkisi olduğunu düşündüğünüz ya da etkileşim etkisini katabileceğiniz şekilde modeli geliştirebilirsiniz) ve alternatif modeller için tahmin performanslarını karşılaştırarak final modelinize karar veriniz.

5.1 Değişken Seçim Yöntemleri

5.1.1 Olası En İyi Alt Küme Seçimi

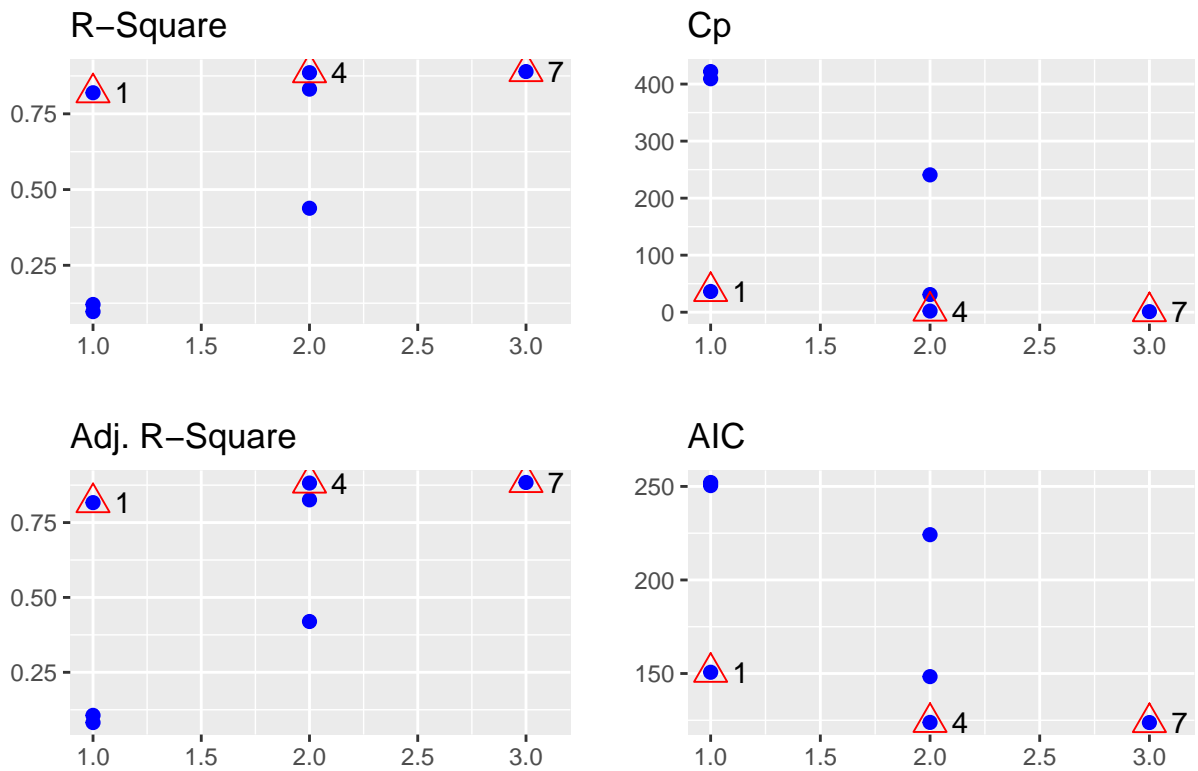
```
a<-ols_step_all_possible(model)
summary(a)
```

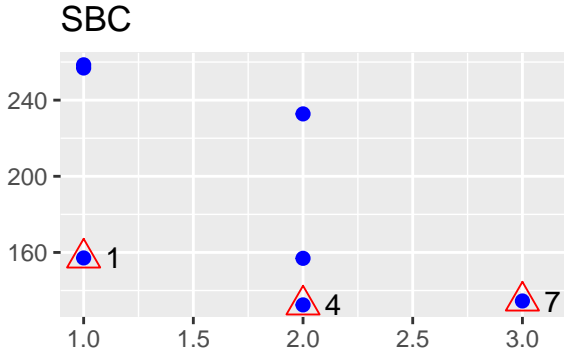
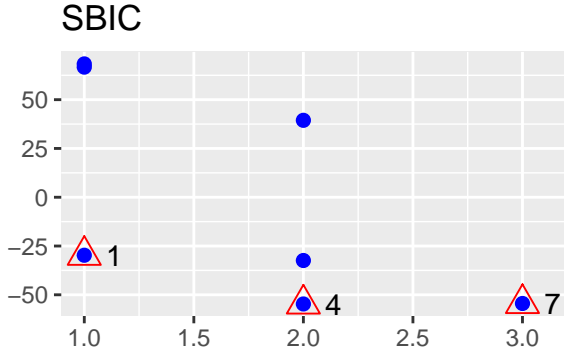
```
##          mindex          n          predictors          rsquare
## Min.      :1.0      Min.      :1.000      Length:7      Min.      :0.0971
## 1st Qu.:2.5      1st Qu.:1.000      Class :character  1st Qu.:0.2795
## Median :4.0      Median :2.000      Mode  :character  Median :0.8195
## Mean      :4.0      Mean      :1.714                      Mean      :0.5832
## 3rd Qu.:5.5      3rd Qu.:2.000                      3rd Qu.:0.8586
## Max.      :7.0      Max.      :3.000                      Max.      :0.8895
##          adjr          predrsq          cp          aic
## Min.      :0.0823      Min.      :0.04123      Min.      : 3.987      Min.      :123.8
## 1st Qu.:0.2630      1st Qu.:0.22318      1st Qu.: 18.462      1st Qu.:136.1
## Median :0.8166      Median :0.80622      Median : 37.328      Median :150.7
## Mean      :0.5738      Mean      :0.55033      Mean      :164.903      Mean      :181.9
## 3rd Qu.:0.8539      3rd Qu.:0.84150      3rd Qu.:326.570      3rd Qu.:237.3
## Max.      :0.8838      Max.      :0.87548      Max.      :422.944      Max.      :252.1
##          sbic          sbc          msep          fpe
## Min.      : -54.7129      Min.      :132.5      Min.      : 23.99      Min.      :0.4047
## 1st Qu.: -43.4426      1st Qu.:145.7      1st Qu.: 30.16      1st Qu.:0.5014
## Median : -29.7355      Median :157.1      Median : 37.86      Median :0.6201
## Mean      : 0.4456      Mean      :189.9      Mean      : 87.99      Mean      :1.4490
## 3rd Qu.: 53.0596      3rd Qu.:244.8      3rd Qu.:152.16      3rd Qu.:2.5066
## Max.      : 68.3338      Max.      :258.5      Max.      :189.43      Max.      :3.1022
##          apc          hsp
```

```
## Min.      :0.1255    Min.      :0.006556
## 1st Qu.:0.1555    1st Qu.:0.008114
## Median :0.1923    Median :0.010016
## Mean      :0.4494    Mean      :0.023421
## 3rd Qu.:0.7774    3rd Qu.:0.040516
## Max.      :0.9621    Max.      :0.050113
```

```
plot(a)
```

page 1 of 2





-ilk elemeyi C_p üzerinden yapılırsa modeldeki yarıllığı ortadan kaldırabiliriz. Modeldeki parametre sayısına eşit çıkması istenir. Bu modelde 4. ($x_1 + x_2$) biraz yanıllı ama göze alınabilecek bir yanıllık ve 7. ($x_1 + x_2 + x_3$) modelde ise yanıllık bulunmamaktadır.

-Düzeltilmiş R^2 üzerinden karşılaştırma yapmak daha uygun ve ikisininde büyük ve yakın çıkması (modele alınan değişkenlerin anlamlı olduğu anlamına gelir) bu sebepten 4. ve 7. model en yüksek açıklamaya sahiptir.

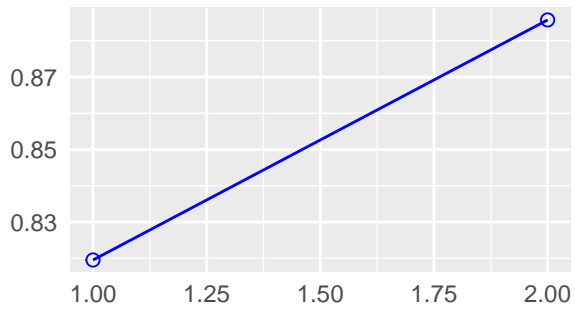
-Akaike Bilgi Kriteri (AIC) ve Bayesian Bilgi Kriteri (BIC, SBIC) kriterleri için de değerlerin küçük çıkması istenir. Bu kriterler 4. ve 7. modelleri önerir.

-Bunlarla birlikte ne kadar az değişken o kadar iyi olduğu için R^2 'ler ve C_p açısından da 4. model alternatif bir modeldir.

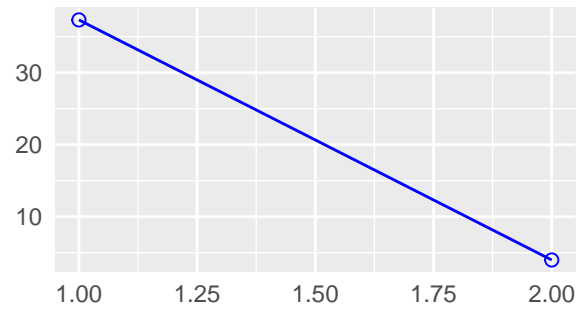
5.1.2 Adımsal Regresyon

```
s<-ols_step_both_p(model)
plot(s)
```

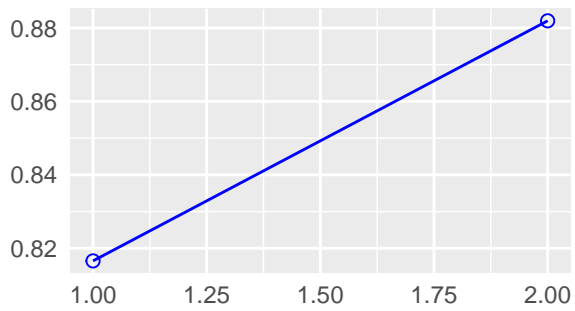
R-Square



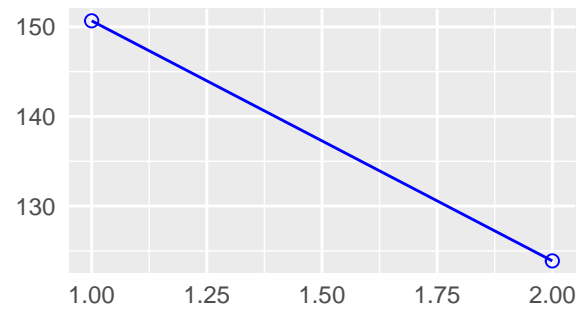
C(p)



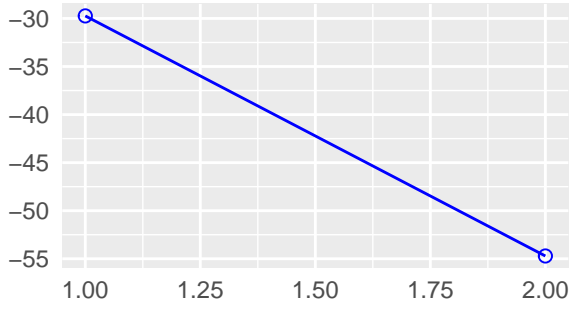
Adj. R-Square



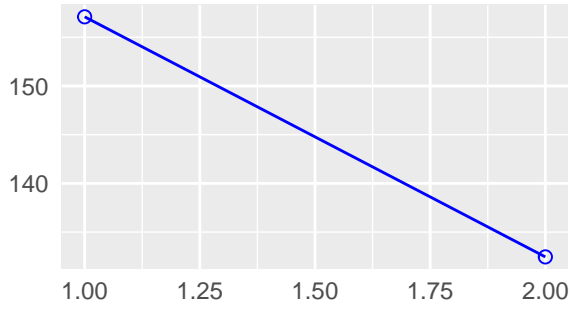
AIC



SBIC



SBC



```
s$model
```

```
##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = l)
##
## Coefficients:
## (Intercept)          x1          x2
##    2.32319    0.07529    0.03002
```

```
s
```

```
##
##                               Stepwise Selection Summary
## -----
##                               Added/      Adj.
## Step  Variable  Removed  R-Square  R-Square  C(p)    AIC    RMSE
## -----
##    1     x1      addition    0.820    0.817   37.3280 150.6741 0.7752
##    2     x2      addition    0.886    0.882   3.9870 123.8775 0.6220
## -----
```

-Adımsal regresyon modeline göre 4. model seçilmiştir. x_1 ve x_2 nin olduğu model

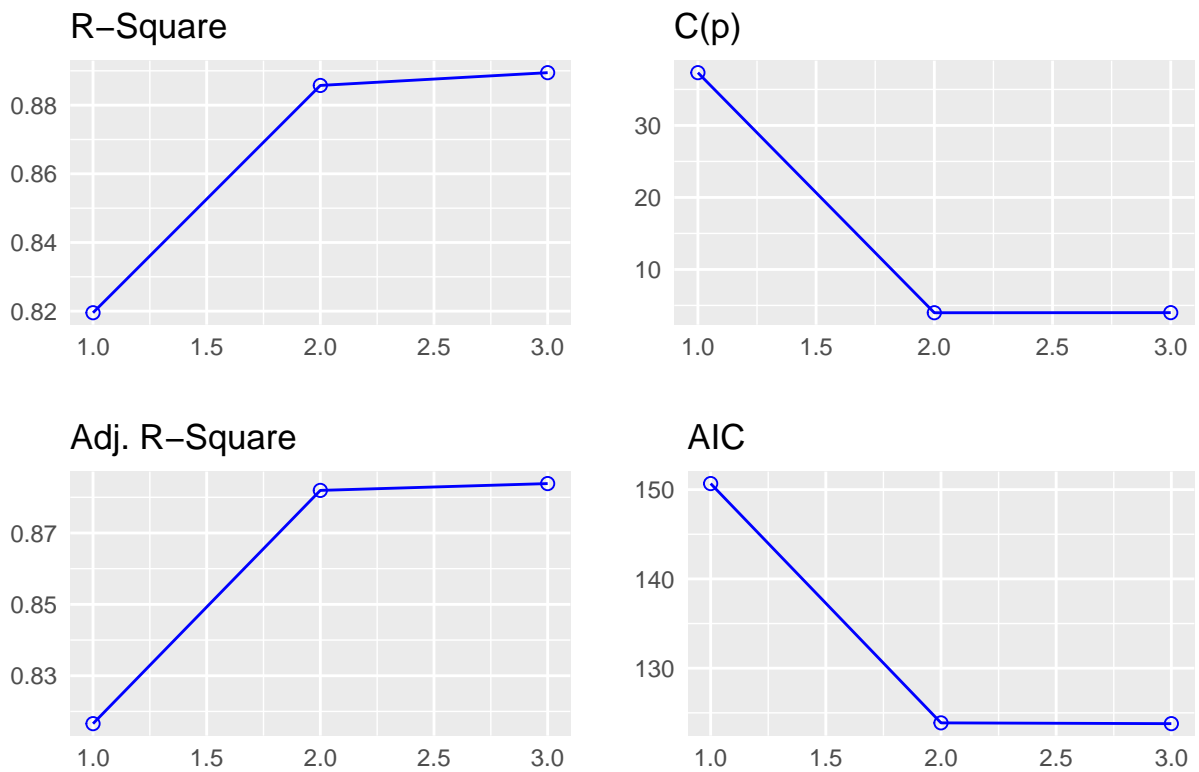
5.1.3 İleriye Doğru Seçim

```
f=ols_step_forward_p(model)
f$model
```

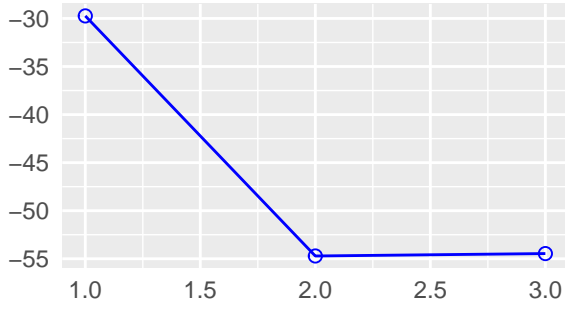
```
##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = 1)
##
## Coefficients:
## (Intercept)          x1          x2          x3
##    2.57348    0.07138    0.03555   -1.14612
```

```
plot(f)
```

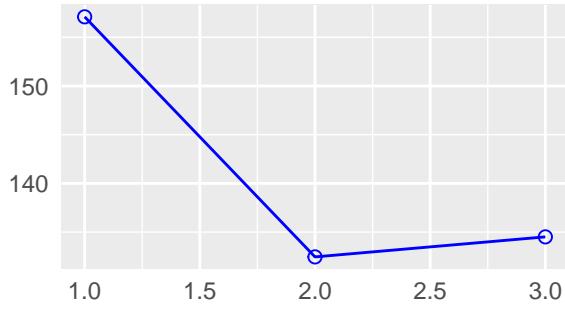
page 1 of 2



SBIC



SBC



f

```
##
##                               Selection Summary
## -----
##      Variable      Adj.
## Step  Entered  R-Square R-Square  C(p)    AIC    RMSE
## -----
##    1    x1      0.8195   0.8166   37.3278  150.6741  0.7752
##    2    x2      0.8857   0.8819    3.9875  123.8775  0.6220
##    3    x3      0.8895   0.8838    4.0000  123.7903  0.6169
## -----
```

-İleriye doğru seçim 7. modeli yani full modeli seçmiştir.

5.1.4 Geriye Doğru Eleme

```
b=ols_step_backward_p(model)
b$model

##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = 1)
##
## Coefficients:
## (Intercept)          x1          x2          x3
```

```
##      2.57348      0.07138      0.03555      -1.14612
```

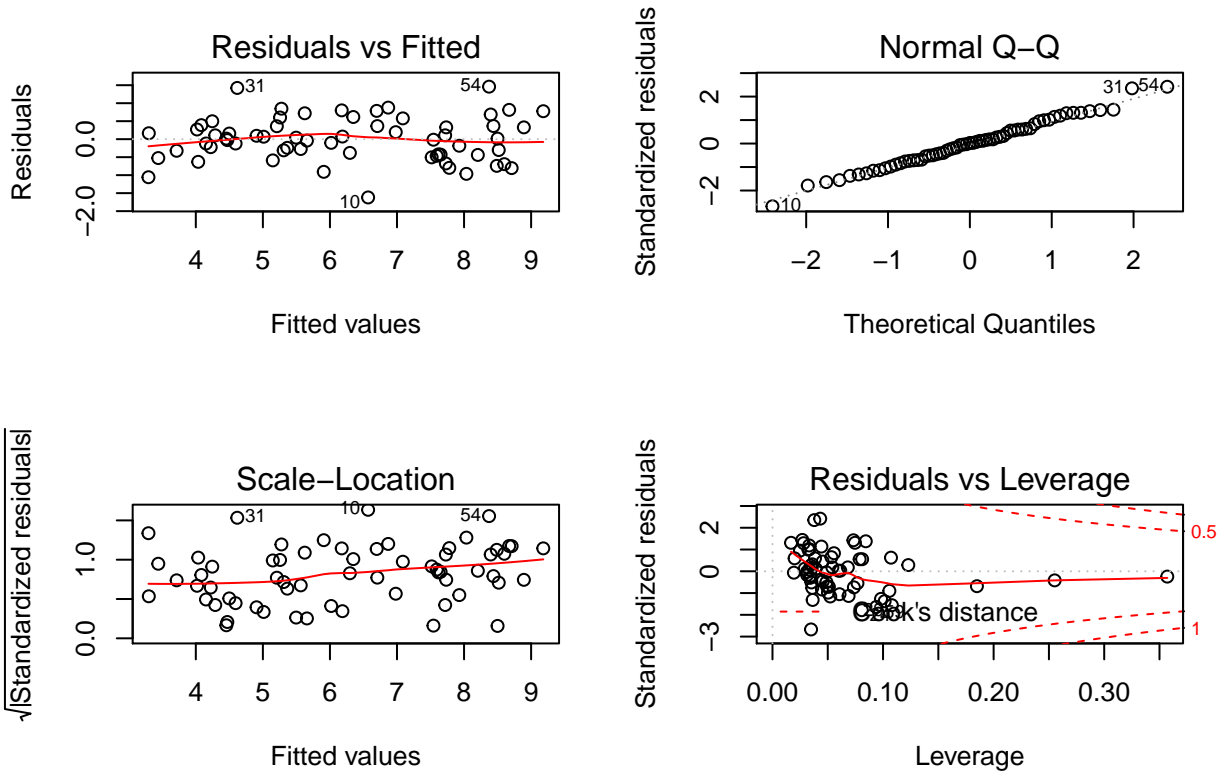
-Geriye doğru eleme yöntemi 7. modeli seçmiştir.

Tüm değişken seçim yöntemleri x_1+x_2 modeli ile full modeli önermiştir.

5.1.5 Full Model ($x_1+x_2+x_3$)

```
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.61964 -0.42605  0.01484  0.36458  1.45896
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.573482   0.269312   9.556 1.37e-13 ***
## x1           0.071380   0.004600  15.518 < 2e-16 ***
## x2           0.035550   0.006393   5.561 6.83e-07 ***
## x3          -1.146124   0.812981  -1.410   0.164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6169 on 59 degrees of freedom
## Multiple R-squared:  0.8895, Adjusted R-squared:  0.8838
## F-statistic: 158.3 on 3 and 59 DF,  p-value: < 2.2e-16
par(mfrow=c(2,2))
plot(model)
```



```
par(mfrow=c(1,1))
```

H_0 : Artıkların varyansı homojendir.

H_1 : Artıkların varyansı heterojendir.

```
bptest(model)
```

```
##
```

```
## studentized Breusch-Pagan test
```

```
##
```

```
## data: model
```

```
## BP = 1.5677, df = 3, p-value = 0.6667
```

$0.6667 > 0.05 \rightarrow H_0$ reddedilemez artıklar sabitdir.

H_0 : Artıklar normal dağılır.

H_1 : Artıklar normal dağılmaz.

```
shapiro.test(model$residuals)
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: model$residuals
```

```
## W = 0.99141, p-value = 0.9408
```

$0.9408 > 0.05 \rightarrow H_0$ reddedilemez artıklar normal dağılır.

Artıkların otokorelasyon kontrolü

```
dwtest(model)
```

```
##  
## Durbin-Watson test  
##  
## data: model  
## DW = 1.8487, p-value = 0.2683  
## alternative hypothesis: true autocorrelation is greater than 0  
(DW = 1.8487)<2 olduğu için zayıf pozitif yönlü otokorelasyon bulunmaktadır.
```

```
vif(model)
```

```
##          x1          x2          x3  
## 1.577060 1.608063 2.106548
```

```
layout(matrix(c(1,2,3,4),2,2))
```

```
e<-resid(model)
```

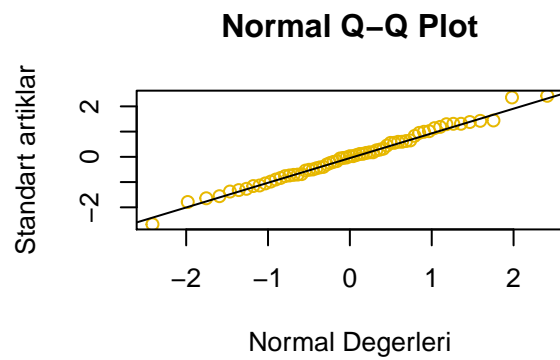
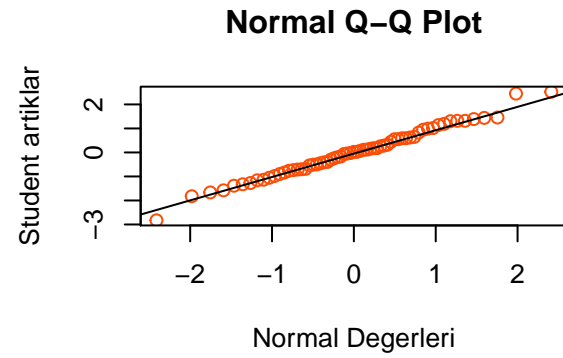
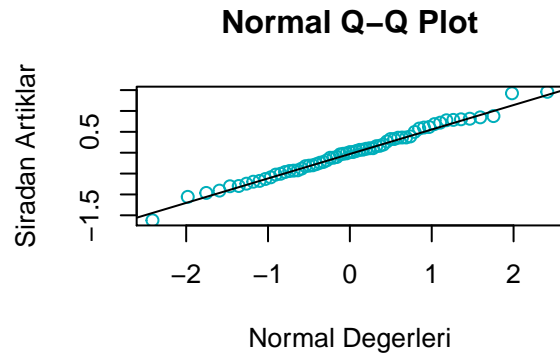
```
qqnorm(e, ylab = "Sıradan Artıklar", xlab = "Normal Değerleri", col="#00AFBB")  
qqline(e)
```

```
d<-rstandard(model)
```

```
qqnorm(d, ylab = "Standart artıklar", xlab = "Normal Değerleri", col="#E7B800")  
qqline(d)
```

```
r<-rstudent(model)
```

```
qqnorm(r, ylab = "Student artıklar", xlab = "Normal Değerleri", col="#FC4E07")  
qqline(r)
```



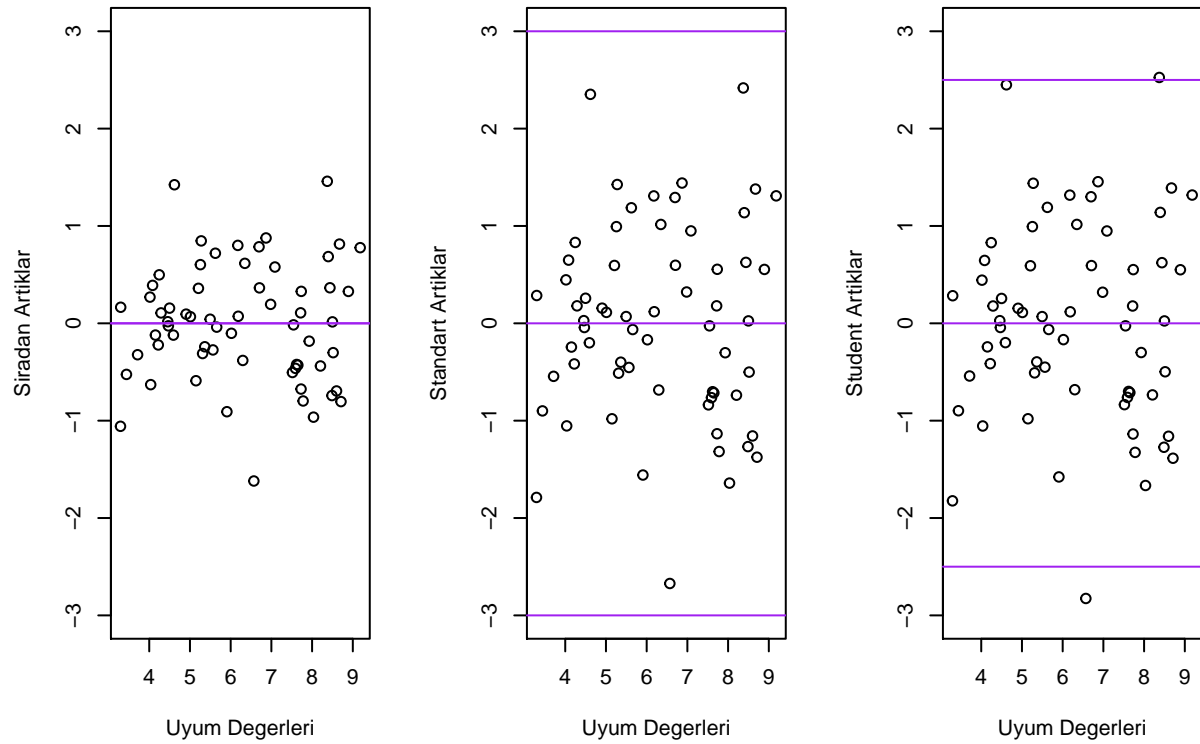
```
uyum.değerleri<-model$fitted.values

layout(matrix(c(1,2,3),1,3))

plot(uyum.değerleri, e, ylab = "Sıradan Artıklar", xlab = "Uyum Değerleri", ylim = c(-3,3))
abline(h=c(0,0), col="purple")

plot(uyum.değerleri,d, ylab = "Standart Artıklar", xlab = "Uyum Değerleri", ylim = c(-3,3))
abline(h=c(-3,0,3),col="purple")

plot(uyum.değerleri,r, ylab = "Student Artıklar", xlab = "Uyum Değerleri", ylim = c(-3,3))
abline(h=c(-2.5,0,2.5),col="purple")
```



```
which.min(r)
```

```
## 10
## 10
```

```
which.max(r)
```

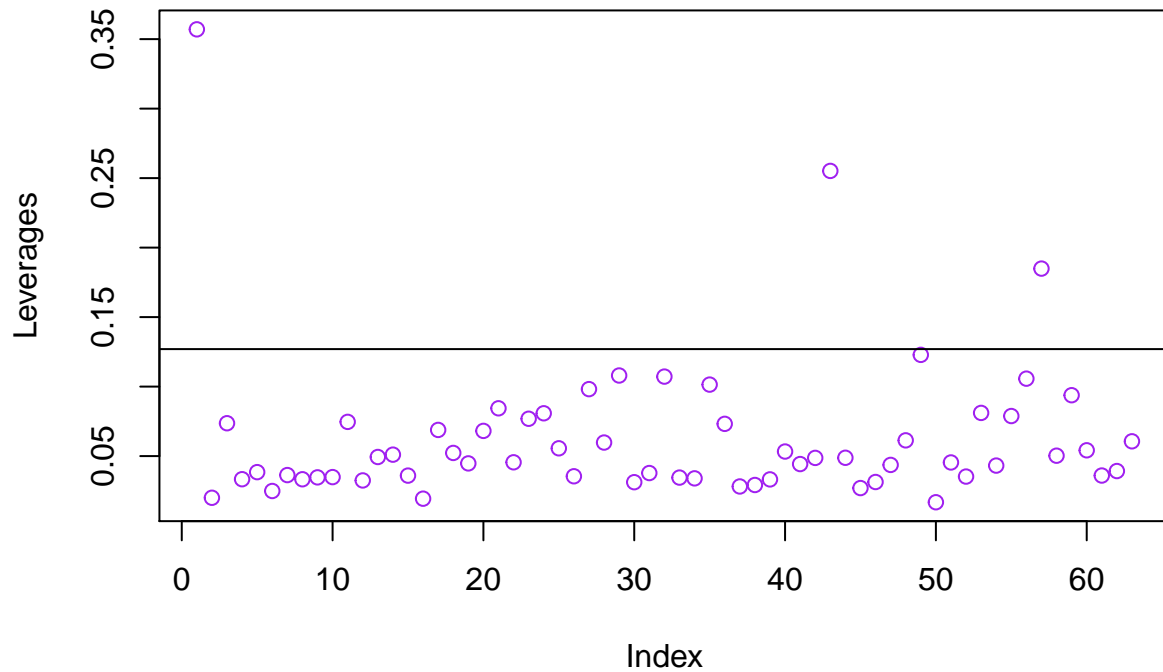
```
## 54
## 54
```

student artıklara göre 10. ve 54. gözlem uç değer çıkmıştır.

Şapka Matrisindeki Köşegen Değerleri

```
h_v<-hatvalues(model)
plot(h_v,ylab="Leverages",main="Index plot of Leverages", col="purple")
abline(h=2*mean(h_v))
```

Index plot of Leverages



```
sum(h_v)
```

```
## [1] 4
```

```
h_v[h_v > 2*mean(h_v)]
```

```
##          1          43          57
## 0.3570276 0.2551823 0.1849034
```

hat değerlerine baktığımızda 1., 43., ve 57., gözlemler etkin gözlemlerdir ve incelenmelidir.

Cook Uzaklığı, verilerden belirli bir gözlem atıldığında $\hat{\beta}$ vektöründeki değişimi ölçmek için tasarlanmıştır.

```
c<-cooks.distance(model)
```

```
cooks.distance(model)>(4/model$df.residual)
```

```
##      1      2      3      4      5      6      7      8      9     10     11     12     13
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##     14     15     16     17     18     19     20     21     22     23     24     25     26
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE  FALSE FALSE
##     27     28     29     30     31     32     33     34     35     36     37     38     39
## FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##     40     41     42     43     44     45     46     47     48     49     50     51     52
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##     53     54     55     56     57     58     59     60     61     62     63
## FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
```

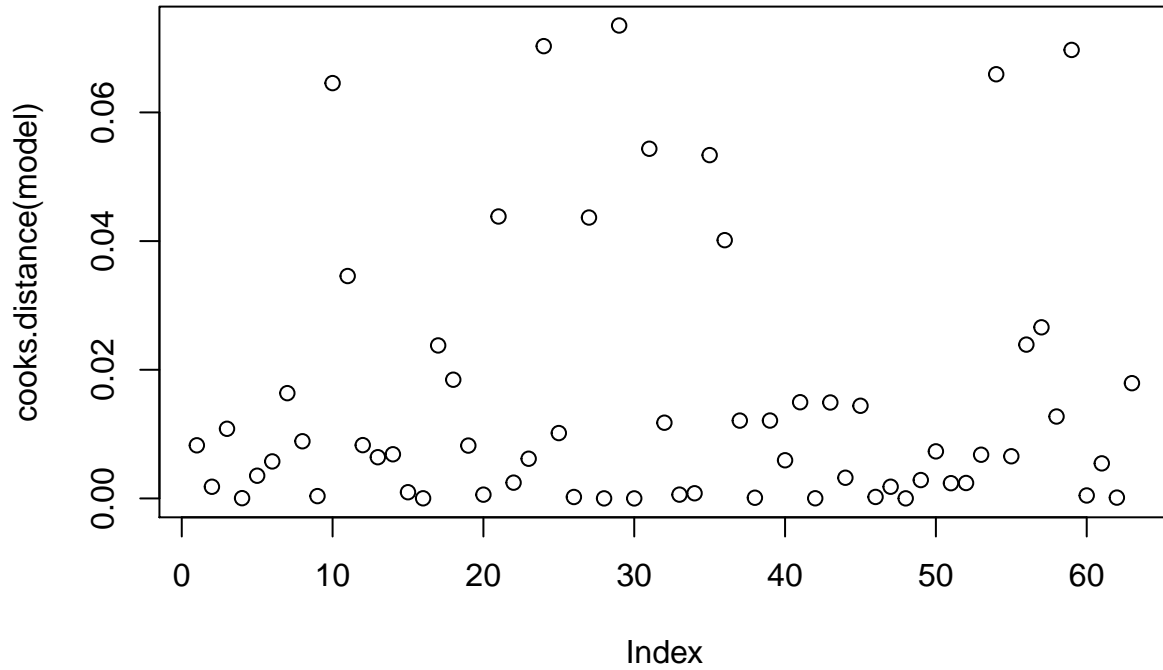
```
which(cooks.distance(model)>4/model$df.residual)
```

```
## 24 29 59
```



```
## 24 29 59
```

```
plot(cooks.distance(model))  
ci<-4/model$df.residual  
identify(c, pch = 1, frame = FALSE)
```



```
## integer(0)
```

```
which(c>ci)
```

```
## 24 29 59
```

```
## 24 29 59
```

24., 29. ve 59. gözlemler B(şapka) üzerinde yüksek etkili bir gözlemdir.

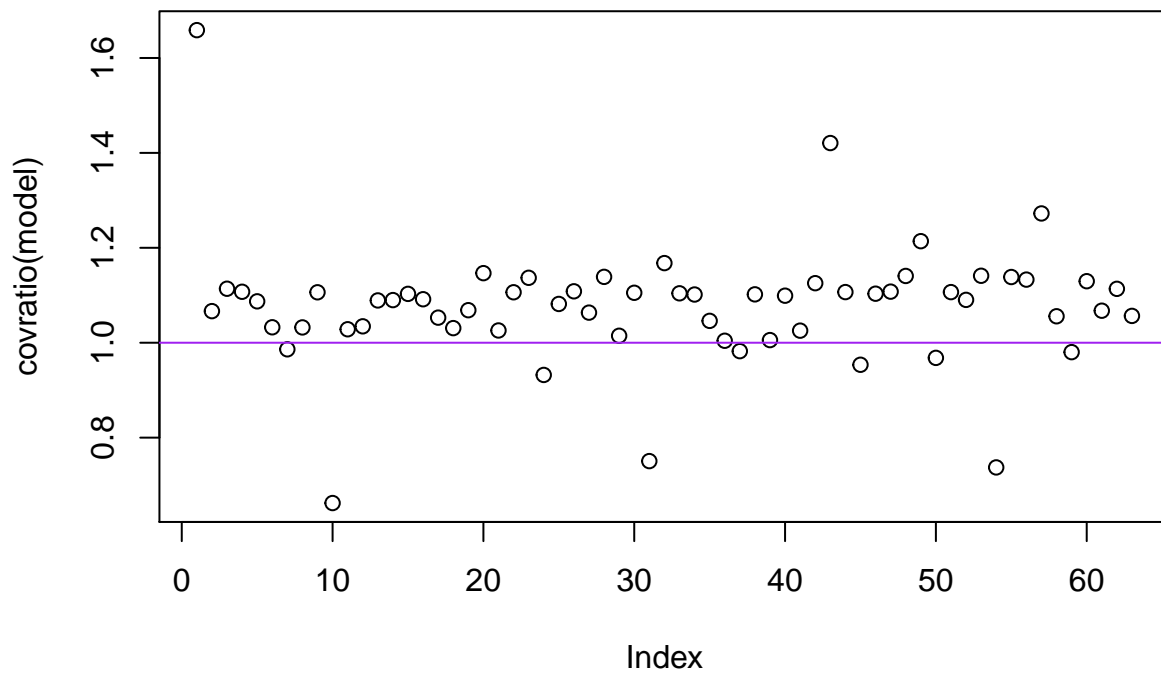
Covratio istatistiği bir gözlemin silinmesi durumunda genel varyansın nasıl etkilendiğini gösterir.

```
covratio(model)
```

```
##      1      2      3      4      5      6      7      8  
## 1.6586381 1.0665399 1.1137845 1.1073856 1.0871812 1.0325297 0.9861052 1.0322503  
##      9     10     11     12     13     14     15     16  
## 1.1062576 0.6621729 1.0283493 1.0345557 1.0891085 1.0898109 1.1029365 1.0916269  
##     17     18     19     20     21     22     23     24  
## 1.0528387 1.0307699 1.0685533 1.1465628 1.0258064 1.1062840 1.1368027 0.9320727  
##     25     26     27     28     29     30     31     32  
## 1.0816900 1.1083370 1.0634080 1.1387790 1.0146211 1.1052308 0.7505416 1.1677789  
##     33     34     35     36     37     38     39     40  
## 1.1040982 1.1015886 1.0460730 1.0041369 0.9822219 1.1019354 1.0055356 1.0990813  
##     41     42     43     44     45     46     47     48
```

```
## 1.0254287 1.1254580 1.4207316 1.1066652 0.9535432 1.1033056 1.1076143 1.1406788
##          49          50          51          52          53          54          55          56
## 1.2139684 0.9680386 1.1066412 1.0903031 1.1411001 0.7372503 1.1384312 1.1330046
##          57          58          59          60          61          62          63
## 1.2723961 1.0557132 0.9801518 1.1296160 1.0674842 1.1135646 1.0565038
```

```
plot(covratio(model))
abline(h=c(1),col="purple")
```



1,10,31,43,49,54,57. gözlemler parametre tahminlerinin var-cov matrisleri üzerinde etkili gözlemler olarak dikkate alınır.

```
outlierTest(model)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##      rstudent unadjusted p-value Bonferroni p
## 10 -2.826316          0.0064494      0.40631
```

5.1.6 Model $\rightarrow x_1 + x_2$

```
model_a<-lm(y~x1+x2)
summary(model_a)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.61915 -0.46330  0.01813  0.39256  1.49080
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.323189    0.204164  11.379  < 2e-16 ***
## x1           0.075290    0.003700   20.350  < 2e-16 ***
## x2           0.030024    0.005092    5.896 1.83e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.622 on 60 degrees of freedom
## Multiple R-squared:  0.8857, Adjusted R-squared:  0.8819
## F-statistic: 232.6 on 2 and 60 DF,  p-value: < 2.2e-16
```

H_0 : Artıkların varyansı homojendir.

H_1 : Artıkların varyansı heterojendir.

```
bptest(model_a)
```

```
##
## studentized Breusch-Pagan test
##
## data: model_a
## BP = 2.0579, df = 2, p-value = 0.3574
```

$0.3574 > 0.05 \rightarrow H_0$ reddedilemez artıkların varyansı sabitdir.

H_0 : Artıklar normal dağılır.

H_1 : Artıklar normal dağılmaz.

```
shapiro.test(model_a$residuals)
```

```
##
## Shapiro-Wilk normality test
##
## data: model_a$residuals
## W = 0.99034, p-value = 0.9045
```

$0.9045 > 0.05 \rightarrow H_0$ reddedilemez artıklar normal dağılır.

Artıkların otokorelasyon kontrolü

```
dwtest(model_a)
```

```
##
## Durbin-Watson test
##
## data: model_a
## DW = 1.8766, p-value = 0.3154
## alternative hypothesis: true autocorrelation is greater than 0
```

$(DW = 1.8766) < 2$ olduğu için zayıf pozitif yönlü otokorelasyon bulunmaktadır.

```
vif(model_a)
```

```
##      x1      x2
## 1.003673 1.003673
```

```

outlierTest(model_a)

## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##      rstudent unadjusted p-value Bonferroni p
## 10 -2.796521      0.0069627      0.43865

e1<-resid(model_a)
d1<-rstandard(model_a)
r1<-rstudent(model_a)

uyum.değerleri_a<-model_a$fitted.values

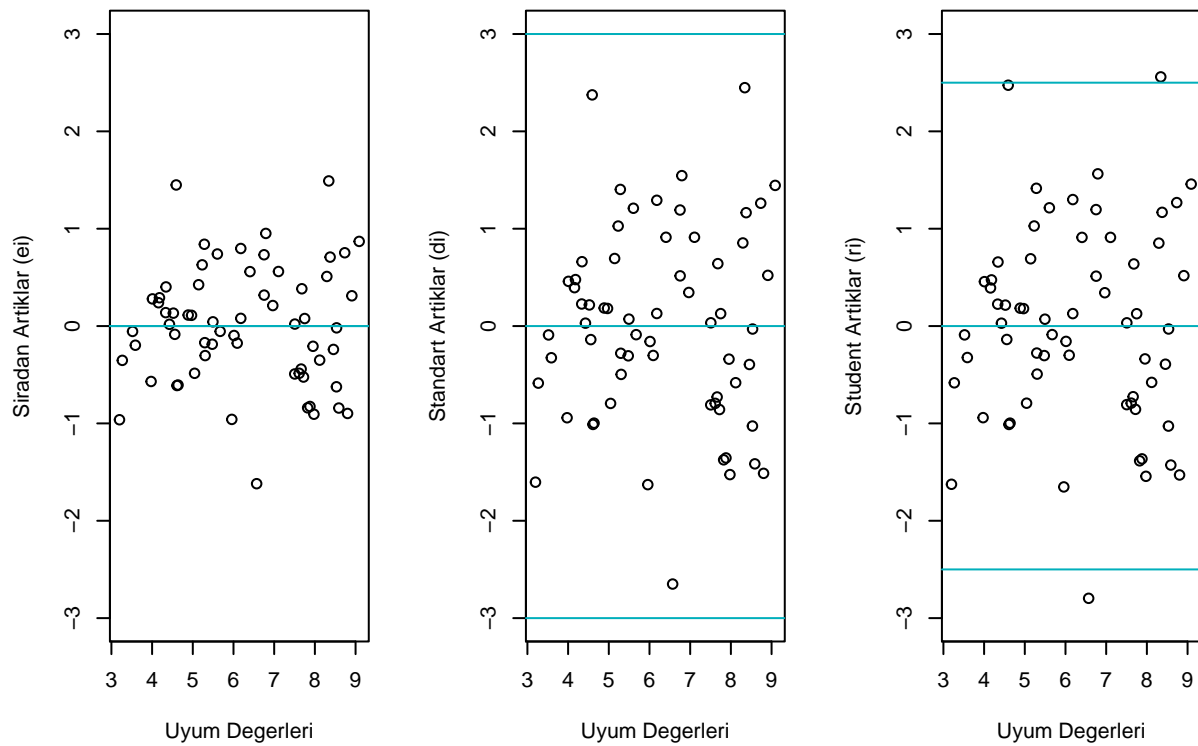
layout(matrix(c(1,2,3),1,3))

plot(uyum.değerleri_a, e1, ylab = "Sıradan Artıklar (ei)", xlab = "Uyum Değerleri", ylim = c(-3,3))
abline(h=c(0,0), col="#00AFBB")

plot(uyum.değerleri_a,d1, ylab = "Standart Artıklar (di)", xlab = "Uyum Değerleri", ylim = c(-3,3))
abline(h=c(-3,0,3),col="#00AFBB")

plot(uyum.değerleri_a,r1, ylab = "Student Artıklar (ri)", xlab = "Uyum Değerleri", ylim = c(-3,3))
abline(h=c(-2.5,0,2.5),col="#00AFBB")

```



```
which.min(r1)
```

```
## 10
```

```
## 10
```

```
which.max(r1)
```

```
## 54
```

```
## 54
```

$|d_i| > 3$ kriteri ile standart artıklara baktığımızda 3 tane potansiyel uç değer vardır ve $|r_i| > 2.5$ kriterimize baktığımızda 2 tanede ki bunlar 10. ve 54. gözlemler uç değer çıkmıştır.

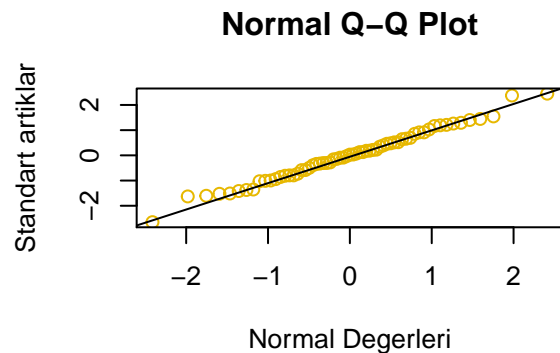
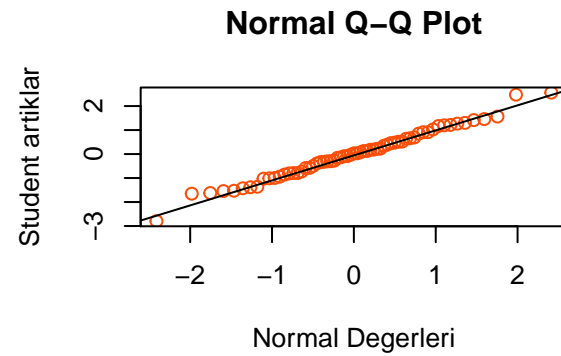
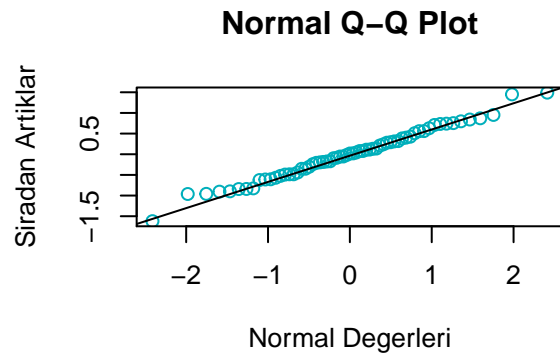
5.a - Hataların normal dağıldığı varsayımını grafikte ve uygun istatistiksel test ile kontrol ediniz.

```
layout(matrix(c(1,2,3,4),2,2))
```

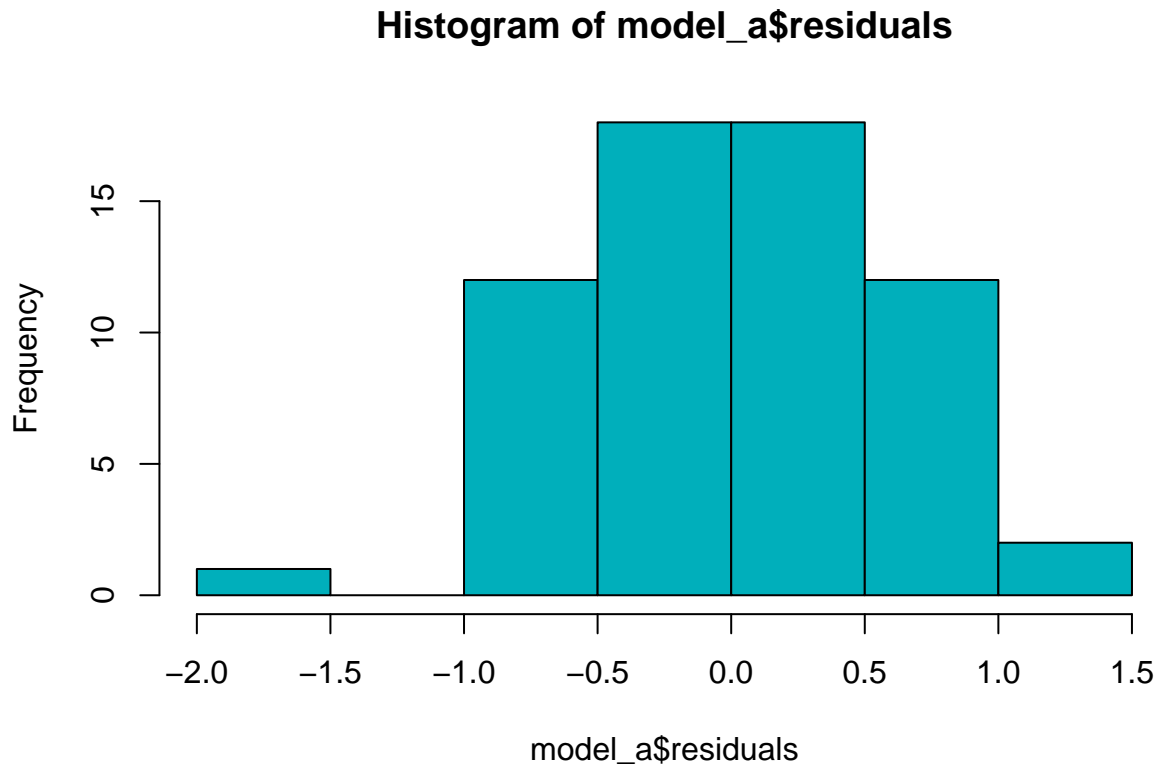
```
qqnorm(e1, ylab = "Sıradan Artıklar", xlab = "Normal Değerleri", col="#00AFBB")  
qqline(e1)
```

```
qqnorm(d1, ylab = "Standart artıklar", xlab = "Normal Değerleri", col="#E7B800")  
qqline(d1)
```

```
qqnorm(r1, ylab = "Student artıklar", xlab = "Normal Değerleri", col="#FC4E07")  
qqline(r1)
```



```
hist(model_a$residuals, col = "#00AFBB")
```



H_0 : Artıklar normal dağılır.

H_1 : Artıklar normal dağılmaz.

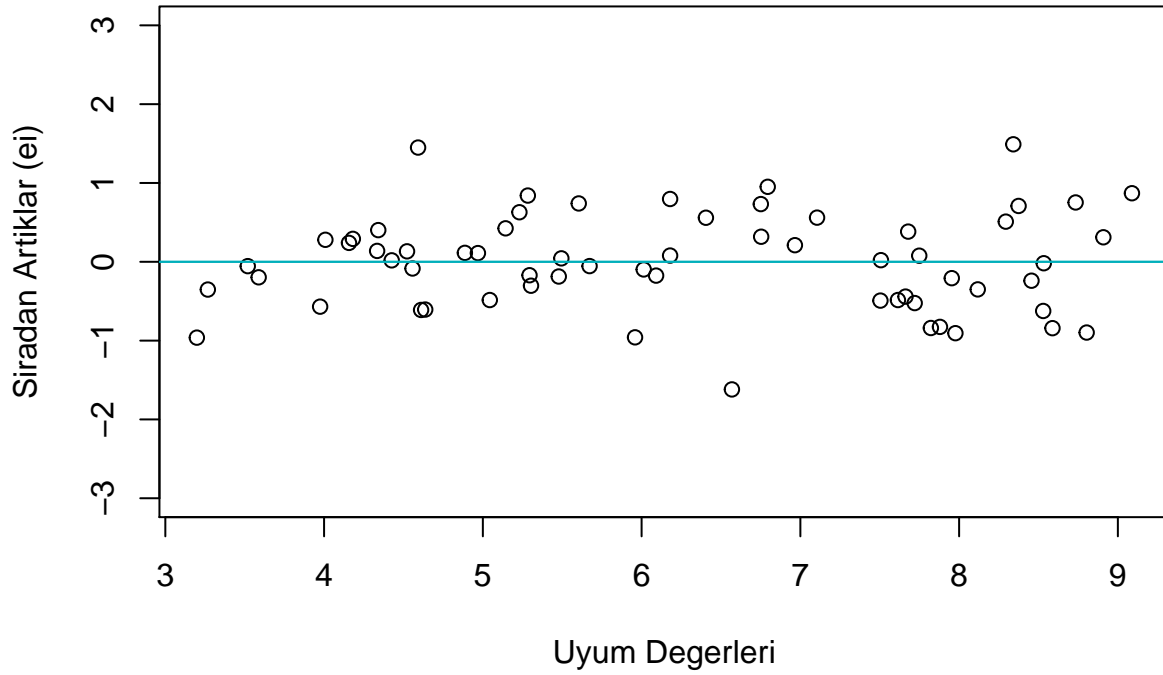
```
shapiro.test(model_a$residuals)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: model_a$residuals  
## W = 0.99034, p-value = 0.9045
```

$0.9045 > 0.05 \rightarrow H_0$ reddedilemez artıklar normal dağılır.

5.b - Hataların sabit varyanslı olup olmadığını grafikte ve uygun istatistiksel test ile kontrol ediniz.

```
plot(uyum.değerleri_a, model_a$residuals, ylab = "Sıradan Artıklar (ei)", xlab = "Uyum Değerleri",  
     ylim = c(-3,3))  
abline(h=c(0,0), col="#00AFBB")
```



H_0 : Artıkların varyansı homojendir.

H_1 : Artıkların varyansı heterojendir.

```
bptest(model_a)
```

```
##
## studentized Breusch-Pagan test
##
## data: model_a
## BP = 2.0579, df = 2, p-value = 0.3574
```

$0.3574 > 0.05 \rightarrow H_0$ reddedilemez artıkların varyansı sabittir.

5.c - Uç değer ve etkin gözlem olup olmadığını grafiklerle ve ilgili değerlerle belirleyiniz.

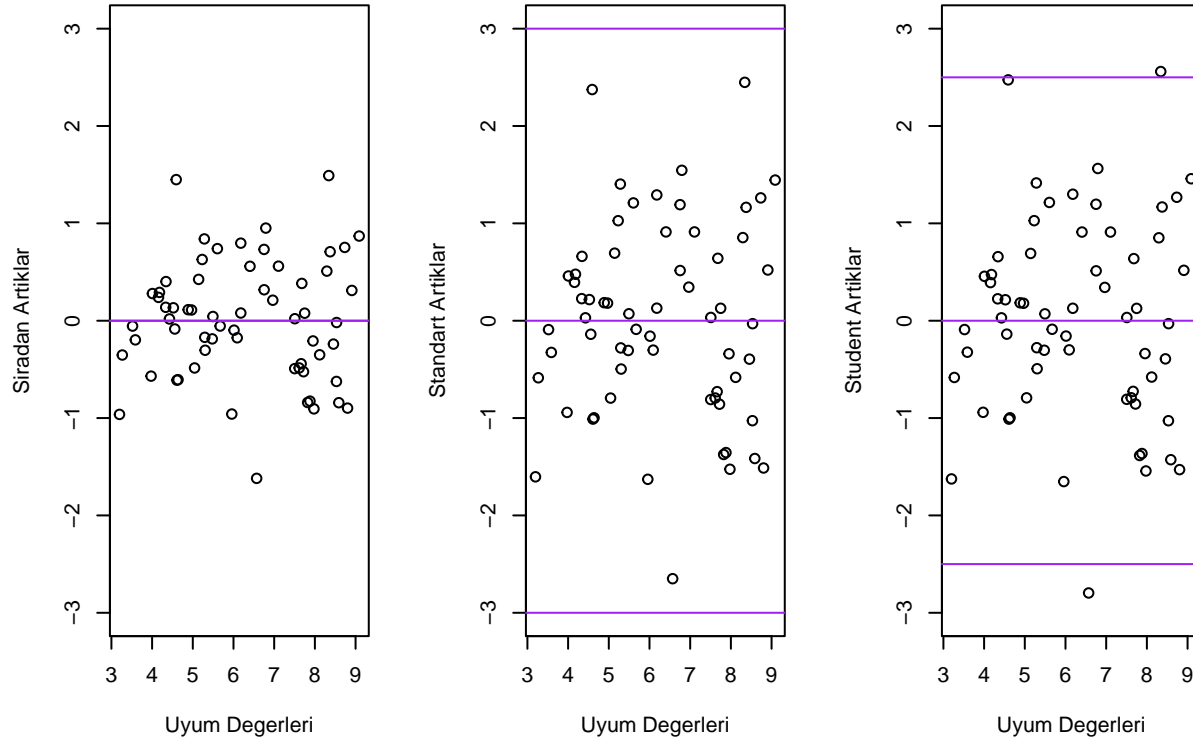
```
uyum.değerleri1<-model_a$fitted.values

layout(matrix(c(1,2,3),1,3))

plot(uyum.değerleri1, e1, ylab = "Sıradan Artıklar", xlab = "Uyum Değerleri", ylim = c(-3,3))
abline(h=c(0,0), col="purple")

plot(uyum.değerleri1,d1, ylab = "Standart Artıklar", xlab = "Uyum Değerleri", ylim = c(-3,3))
abline(h=c(-3,0,3),col="purple")
```

```
plot(uyum.değerleri1,r1, ylab = "Student Artıklar", xlab = "Uyum Değerleri", ylim = c(-3,3))
abline(h=c(-2.5,0,2.5),col="purple")
```



```
which.min(r1)
```

```
## 10
## 10
```

```
which.max(r1)
```

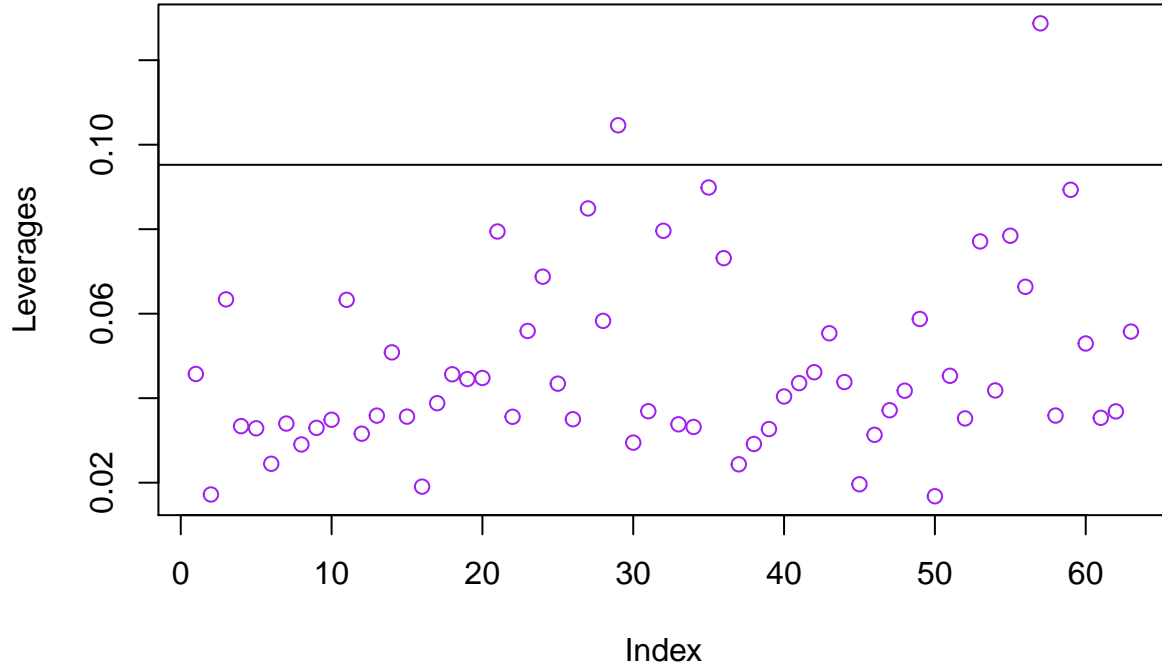
```
## 54
## 54
```

$|di| > 3$ kriteri ile standart artıklara baktığımızda 3 tane potansiyel uç değer vardır ve $|ri| > 2.5$ kriterimize baktığımızda 2 tanede ki bunlar 10. ve 54. gözlemler uç değer çıkmıştır.

Leverage noktaları Hat matrisinin köşegen elemanlarıdır. $h_{ii} > 2p/n$ olması durumunda bu gözlem etkili bir noktadır.

```
h<-hatvalues(model_a)
plot(h,ylab="Leverages",main="Index plot of Leverages", col="purple")
abline(h=2*mean(h))
```


Index plot of Leverages



```
sum(h)
```

```
## [1] 3
```

```
h[h > 2*mean(h)]
```

```
##          29          57
## 0.1045941 0.1287216
```

29. ve 57. gözlemler etkin gözlemlerdir ve incelenmelidir.

Cook uzaklığı, verilerden belirli bir gözlem atıldığında $B(\text{şapka})$ vektöründeki değişimi ölçmek için tasarlanmıştır. $Di > 1$ veya $Di > 4/(n-p)$ ise, i . gözlemin $B(\text{şapka})$ üzerinde yüksek etkili bir gözlemdir.

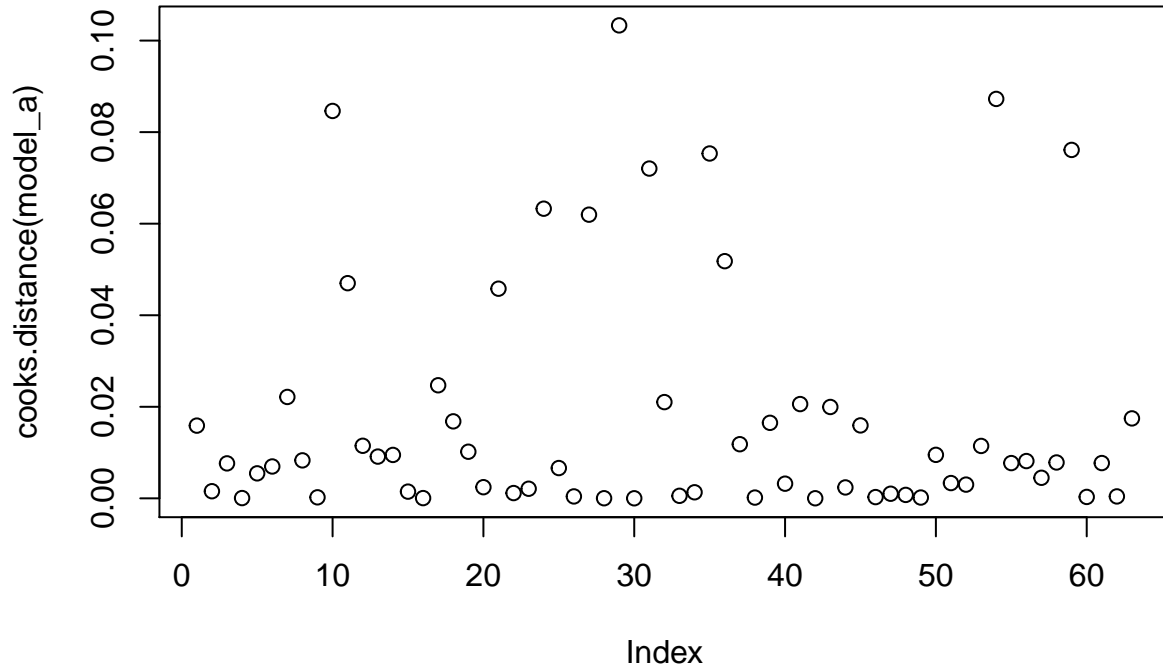
```
ca<-cooks.distance(model_a)
```

```
cooks.distance(model_a)>(4/model$df.residual)
```

```
##      1      2      3      4      5      6      7      8      9     10     11     12     13
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
##     14     15     16     17     18     19     20     21     22     23     24     25     26
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##     27     28     29     30     31     32     33     34     35     36     37     38     39
## FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
##     40     41     42     43     44     45     46     47     48     49     50     51     52
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##     53     54     55     56     57     58     59     60     61     62     63
## FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
```

```
which(cooks.distance(model_a)>4/model$df.residual)
```

```
## 10 29 31 35 54 59
## 10 29 31 35 54 59
plot(cooks.distance(model_a))
cai<-4/model_a$df.residual
identify(ca, pch = 1, frame = FALSE)
```



```
## integer(0)
which(ca>cai)
```

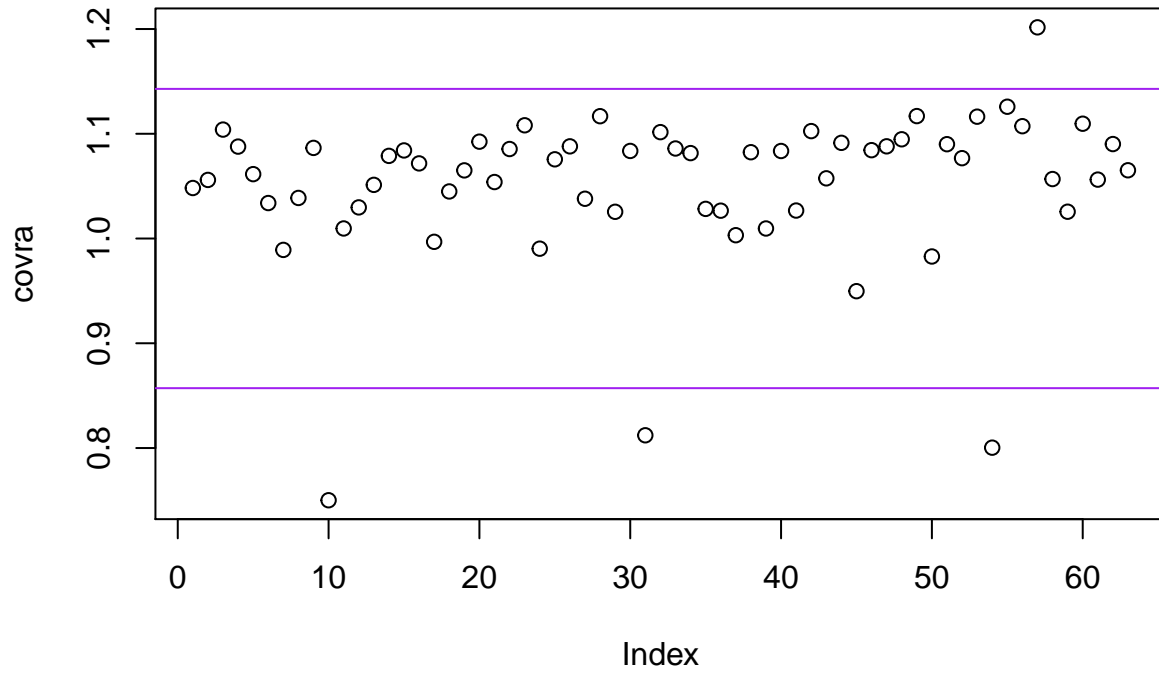
```
## 10 29 31 35 54 59
## 10 29 31 35 54 59
```

10,29,31,35,54,59. gözlemler $D_i > 4/(n-p)$ bu kriter için B(şapka) üzerinde yüksek etkili bir gözlemdir.

$D_i > 1$ kriteri ile incelenirse B(şapka) üzerinde yüksek etkili bir gözlem bulunmamıştır diyebiliriz.

COVRATIOi istatistiği bir gözlemin silinmesi durumunda genel varyansın nasıl etkilendiğini gösterir.

```
covra<-covratio(model_a)
plot(covra)
cl<-(1-3*(mean(h)))
cu<-(1+3*(mean(h)))
abline(h=c(cl,cu),col="purple")
identify(covra, pch = 1, frame = FALSE)
```



```
## integer(0)
which(covra<c1)

## 10 31 54
## 10 31 54
which(covra>cu)
```

```
## 57
## 57
```

10,31,54,57 bu gözlemler parametre tahminlerinin var-cov matrisleri üzerinde etkili gözlemler olarak dikkate alınır.

5.d - VIF değerlerine bakarak yorumlayınız.

```
vif(model_a)

##          x1          x2
## 1.003673 1.003673
```

-Bu modelde vif değerleri 5'ten küçük çıkmıştır. Aralarında çoklu doğrusal bir bağlantı yoktur.

5.e - Final modelin katsayılarını yorumlayınız.

```
model_a<-lm(y~x1+x2)
summary(model_a)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.61915 -0.46330  0.01813  0.39256  1.49080
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.323189    0.204164  11.379 < 2e-16 ***
## x1           0.075290    0.003700  20.350 < 2e-16 ***
## x2           0.030024    0.005092   5.896 1.83e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.622 on 60 degrees of freedom
## Multiple R-squared:  0.8857, Adjusted R-squared:  0.8819
## F-statistic: 232.6 on 2 and 60 DF,  p-value: < 2.2e-16
```

$H_0: \beta_j=0$

H_1 : En az bir β_j farklıdır, $j=1,2$

-F istatistiğine karşılık gelen $2.2e-16 < 0.05$ olduğu için %95 güven düzeyinde H_0 red edilir; en az bir katsayı model için anlamlıdır. Model geçerlidir.

Bu modelde tüm katsayılar %95 güvenle anlamlı çıkmıştır.

$-\beta_0$: ekilebilir arazi dönümü başına kira(x_1) ve mil^2 başına süt ineği oranı(x_2) 0 olduğunda çim çeşidi için dönüm başına kira bedelinin aldığı ortalama değer 2.32 dolar olacaktır.

$-\beta_1$: mil^2 başına süt ineği oranı(x_2) modelde ve sabitken ekilebilir arazi dönümü başına kira(x_1), 1 \$ arttığında çim çeşidi için dönüm başına kira bedelini ortalama 0.075 dolar artırır.

$-\beta_2$: ekilebilir arazi dönümü başına kira(x_1) değişkeni modelde ve sabitken mil^2 başına süt ineği oranı(x_2), 1 birim arttığında çim çeşidi için dönüm başına kira bedelini ortalama 0.03 dolar artırır.

5.f - Katsayıların %95'lik güven aralıklarını elde ederek yorumlayınız.

```
confint.lm(model_a)
```

```
##              2.5 %      97.5 %
## (Intercept) 1.91480041 2.73157703
## x1           0.06788985 0.08269082
## x2           0.01983901 0.04020987
```

$-\beta_0$: ekilebilir arazi dönümü başına kira(x_1) ve mil^2 başına süt ineği oranı(x_2) 0 olduğunda %95 güvenle çim çeşidi için dönüm başına kira bedeli 1.915 dolar ile 2.732 dolar arasındadır.

$-\beta_1$: mil^2 başına süt ineği oranı(x_2) modelde ve sabitken, ekilebilir arazi dönümü başına kira(x_1) 1 \$ arttığında %95 güvenle çim çeşidi için dönüm başına kira bedeli 0.068 dolar ile 0.083 dolar arasındadır.

$-\beta_2$: ekilebilir arazi dönümü başına kira(x_1) modelde ve sabitken, mil^2 başına süt ineği oranı(x_2) 1 birim arttığında %95 güvenle çim çeşidi için dönüm başına kira bedeli 0.02 dolar ile 0.04 dolar arasındadır.

5.g - Yeni bir gözlem değeri için %95'lik güven aralığı ve kestirim aralığını bularak yorumlayınız.

```
test
```

```
##          y1    x_1    x_2    x_3
## 25 8.660254 65.94 22.10 0.09
## 33 6.324555 20.00 40.18 0.56
## 45 5.700877 31.55 23.47 0.19
## 46 4.358899 26.94  8.28 0.10
```

```
new <- data.frame(x1 = test$x_1, x2= test$x_2)
predict(model_a, newdata = new)
```

```
##          1          2          3          4
## 7.951374 5.035377 5.403272 4.600113
```

-ekilebilir arazi dönümü başına kira(x_1) 65.94 dolar ve mil^2 başına süt ineği oranı(x_2) 22.10 değerlerini aldığımda çim çeşidi için dönüm başına kira bedeli ortalama 7.95 dolardır.

-ekilebilir arazi dönümü başına kira(x_1) 20 dolar ve mil^2 başına süt ineği oranı(x_2) 40.18 değerlerini aldığımda çim çeşidi için dönüm başına kira bedeli ortalama 5.04 dolardır.

-ekilebilir arazi dönümü başına kira(x_1) 31.55 dolar ve mil^2 başına süt ineği oranı(x_2) 23.47 değerlerini aldığımda çim çeşidi için dönüm başına kira bedeli ortalama 5.4 dolardır.

-ekilebilir arazi dönümü başına kira(x_1) 26.94 dolar ve mil^2 başına süt ineği oranı(x_2) 8.28 değerlerini aldığımda çim çeşidi için dönüm başına kira bedeli ortalama 4.6 dolardır.

x_1 ve x_2 için y 'nin dağılımının ortalamasının güven aralığı

```
predictnewconf=predict(model_a,new,interval="confidence")
cbind(new,predictnewconf, test$y1)
```

```
##          x1    x2      fit      lwr      upr test$y1
## 1 65.94 22.10 7.951374 7.727236 8.175511 8.660254
## 2 20.00 40.18 5.035377 4.716133 5.354622 6.324555
## 3 31.55 23.47 5.403272 5.216749 5.589796 5.700877
## 4 26.94  8.28 4.600113 4.367097 4.833129 4.358899
```

-ekilebilir arazi dönümü başına kira(x_1) 65.94 dolar ve mil^2 başına süt ineği oranı(x_2) 22.10 iken %95 güvenle çim çeşidi için dönüm başına kira bedeli ortalama 7.73 dolar ile 8.18 dolar arasında değişir.

-ekilebilir arazi dönümü başına kira(x_1) 20 dolar ve mil^2 başına süt ineği oranı(x_2) 40.18 iken %95 güvenle çim çeşidi için dönüm başına kira bedeli ortalama 4.7 dolar ile 5.35 dolar arasında değişir.

-ekilebilir arazi dönümü başına kira(x_1) 31.55 dolar ve mil^2 başına süt ineği oranı(x_2) 23.47 iken %95 güvenle çim çeşidi için dönüm başına kira bedeli ortalama 5.22 dolar ile 5.59 dolar arasında değişir.

-ekilebilir arazi dönümü başına kira(x_1) 26.94 dolar ve mil^2 başına süt ineği oranı(x_2) 8.28 iken %95 güvenle çim çeşidi için dönüm başına kira bedeli ortalama 4.37 dolar ile 4.83 dolar arasında değişir.

x_1 ve x_2 için y 'a ait kestirim aralığı

```
predictnewpred<-predict(model_a,new,interval="prediction")
cbind(new,predictnewpred, test$y1)
```

```
##          x1    x2      fit      lwr      upr test$y1
## 1 65.94 22.10 7.951374 6.687241 9.215507 8.660254
## 2 20.00 40.18 5.035377 3.750966 6.319788 6.324555
## 3 31.55 23.47 5.403272 4.145264 6.661281 5.700877
```

```
## 4 26.94 8.28 4.600113 3.334375 5.865850 4.358899
```

-ekilebilir arazi dönümü başına kira(x_1) 65.94 dolar ve mil^2 başına süt ineği oranı(x_2) 22.10 iken %95 güvenle çim çeşidi için dönüm başına kira bedeli ortalama 6.69 dolar ile 9.22 dolar arasında değişir.

-ekilebilir arazi dönümü başına kira(x_1) 20 dolar ve mil^2 başına süt ineği oranı(x_2) 40.18 iken %95 güvenle çim çeşidi için dönüm başına kira bedeli ortalama 3.75 dolar ile 6.32 dolar arasında değişir.

-ekilebilir arazi dönümü başına kira(x_1) 31.55 dolar ve mil^2 başına süt ineği oranı(x_2) 23.47 iken %95 güvenle çim çeşidi için dönüm başına kira bedeli ortalama 4.15 dolar ile 6.66 dolar arasında değişir.

-ekilebilir arazi dönümü başına kira(x_1) 26.94 dolar ve mil^2 başına süt ineği oranı(x_2) 8.28 iken %95 güvenle çim çeşidi için dönüm başına kira bedeli ortalama 3.33 dolar ile 5.87 dolar arasında değişir.

```
predictions_a<-predict(model_a,test)
RMSE<-RMSE(predictions_a, test$y1)
RMSE
```

```
## [1] 2.198093
```

```
MAE<-MAE(predictions_a, test$y1)
MAE
```

```
## [1] 1.775054
```

```
cbind(RMSE, MAE)
```

```
##           RMSE           MAE
## [1,] 2.198093 1.775054
```

```
accuracy(model_a)
```

```
##           ME           RMSE           MAE           MPE           MAPE           MASE
## Training set 1.410151e-17 0.6069701 0.481981 -1.239469 8.178677 0.3139003
```

6 - Modeli geliştirmek üzere görüş ve önerileriniz varsa belirtiniz.

→ Bu veri seti için etkileşim etkisi, karesel ve kübik formlar denenmiş fakat anlamsız çıkmıştır. → Seçilen modeldeki uç değerler ve kötü kaldıraç noktalarını (4 gözlem) çıkartarak elde ettiğimiz modelin düzeltilmiş R^2 si 0.9042 olarak elde edilmiştir. Düzeltilmiş R^2 'leri kıyasladığımızda aralarında 0.0223 fark vardır uç değerlerin çıkarıldığı model daha fazla açıklama oranına sahiptir. Fakat 4 gözlemin bilgisi ve yeni oluşan uç değerleri göz önüne aldığımızda bu modeli tercih etmedik zaten biz bu modeli deneme amaçlı kurmuştuk çünkü cook uzaklığı ve covratio değerlerindeki göz önüne aldığımızda aslında uç değerler ve kaldıraç değerlerini bu gözlemleri kapsamaktadır.