Module 4: Advanced Topics for MLR

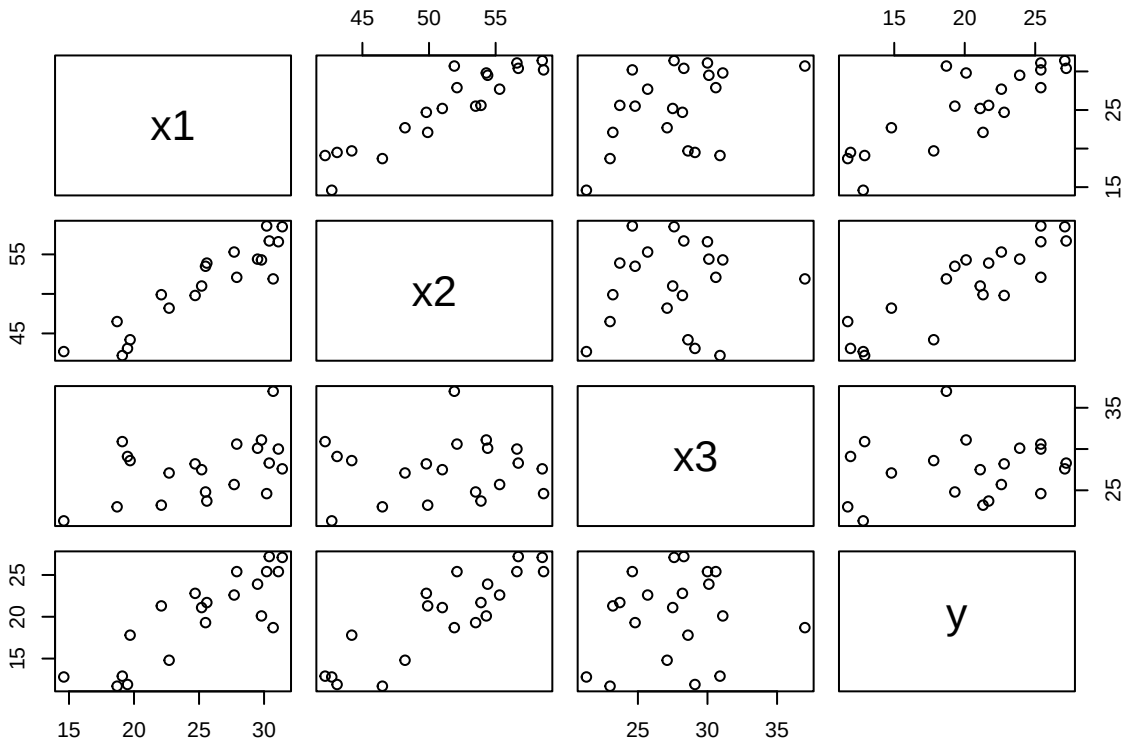
```
# Only need to run on first use and then can comment out

# install.packages('lm.beta')
```

```
library(lm.beta)
library(car)
```

```
bodyfat<-read.csv("../../datasets/bodyfat.csv")
colnames(bodyfat)<-c("x1","x2","x3","y")

plot(bodyfat)
```




```
model1<-lm(y~x1, bodyfat)
summary(model1)
```

```
##
## Call:
## lm(formula = y ~ x1, data = bodyfat)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -6.1195 -2.1904  0.6735  1.9383  3.8523
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.4961     3.3192  -0.451    0.658
## x1            0.8572     0.1288   6.656 3.02e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.82 on 18 degrees of freedom
## Multiple R-squared:  0.7111, Adjusted R-squared:  0.695
## F-statistic:  44.3 on 1 and 18 DF,  p-value: 3.024e-06
```

```
#Anova(model1,type=2)
model2<-lm(y~x2, bodyfat)
#summary(model2)
#anova(model2)



model3<-lm(y~x1+x2, bodyfat)
summary(model3)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, data = bodyfat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9469 -1.8807  0.1678  1.3367  4.0147
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -19.1742     8.3606  -2.293   0.0348 *
## x1            0.2224     0.3034   0.733   0.4737
## x2            0.6594     0.2912   2.265   0.0369 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.543 on 17 degrees of freedom
## Multiple R-squared:  0.7781, Adjusted R-squared:  0.7519
## F-statistic:  29.8 on 2 and 17 DF,  p-value: 2.774e-06
```

```
#anova(model3)
model4<-lm(y~x1+x2+x3, bodyfat)
summary(model4)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3, data = bodyfat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7263 -1.6111  0.3923  1.4656  4.1277
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  117.085     99.782   1.173    0.258
## x1             4.334      3.016   1.437    0.170
## x2            -2.857      2.582  -1.106    0.285
## x3            -2.186      1.595  -1.370    0.190
##
## Residual standard error: 2.48 on 16 degrees of freedom
## Multiple R-squared:  0.8014, Adjusted R-squared:  0.7641
## F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06
```

```
anova(model4)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x1         1 352.27  352.27 57.2768 1.131e-06 ***
## x2         1  33.17   33.17  5.3931   0.03373 *
## x3         1  11.55   11.55  1.8773   0.18956
## Residuals 16  98.40    6.15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

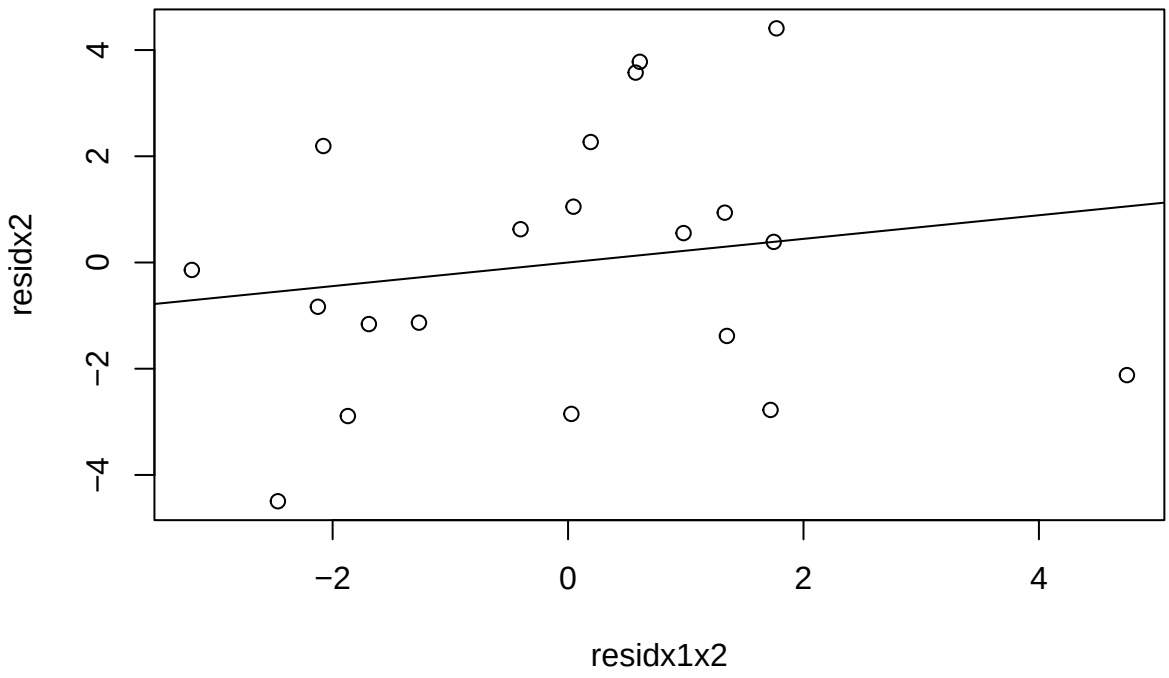
```
#anova(model4)
anova(model3, model4)  #reduced model: model3, full model model4)
```

```
## Analysis of Variance Table
##
## Model 1: y ~ x1 + x2
## Model 2: y ~ x1 + x2 + x3
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1     17 109.951
## 2     16  98.405  1    11.546 1.8773 0.1896
```

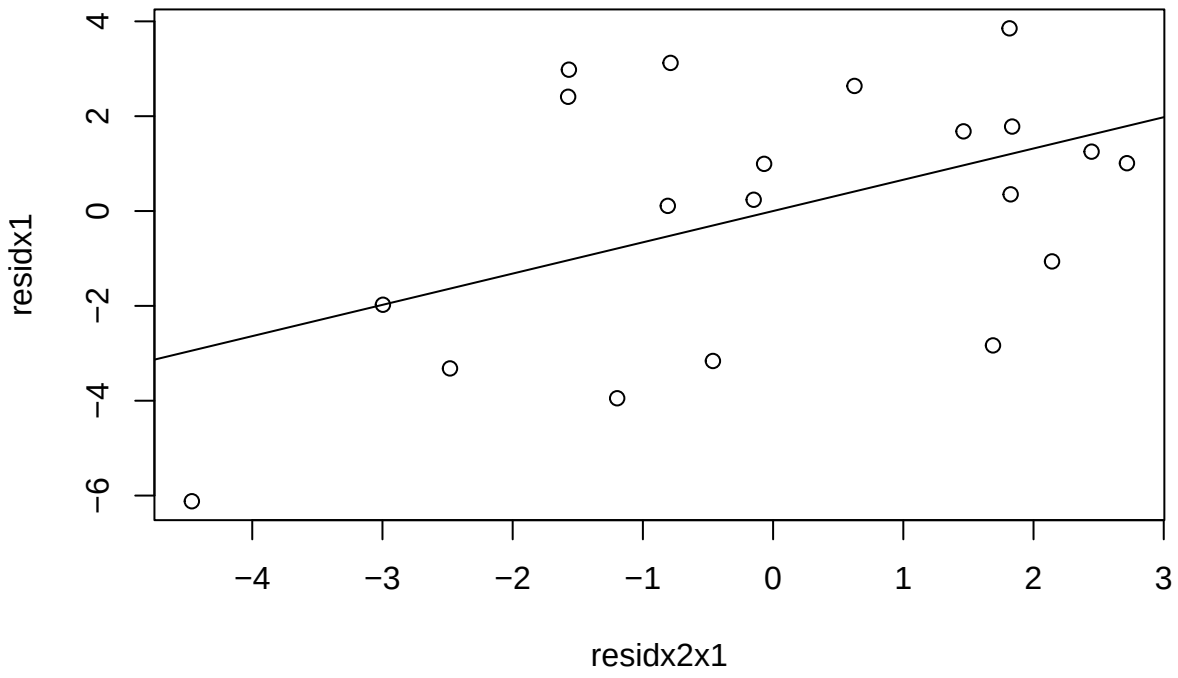
```
#anova(model1, model4) #reduced modeo: model1, full model model4
#summary(lm(y~x1+x2+x3, bodyfat))

#plot(model1)
#qf(0.95,3,16)
```

```
residx2<-model2$residuals
modelx1x2<-lm(x1~x2, bodyfat)
residx1x2<-modelx1x2$residuals
partialx1x2<-lm(residx2~residx1x2)
plot(residx1x2, residx2)
abline(partialx1x2)
```




```
residx1<-model1$residuals
modelx2x1<-lm(x2~x1, bodyfat)
residx2x1<-modelx2x1$residuals
partialx2x1<-lm(residx1~residx2x1)
plot(residx2x1, residx1)
abline(partialx2x1)
```




```
modelx2x3<-lm(y~x2+x3, bodyfat)
summary(modelx2x3)
```

```
##
## Call:
## lm(formula = y ~ x2 + x3, data = bodyfat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0777 -1.8296  0.1893  1.3545  4.1275
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -25.99695    6.99732  -3.715  0.00172 **
## x2            0.85088    0.11245   7.567 7.72e-07 ***
## x3            0.09603    0.16139   0.595  0.55968
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.557 on 17 degrees of freedom
## Multiple R-squared:  0.7757, Adjusted R-squared:  0.7493
## F-statistic:  29.4 on 2 and 17 DF,  p-value: 3.033e-06
```

```
anova(modelx2x3)
```

```
## Analysis of Variance Table
##
```

```
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x2         1 381.97  381.97  58.441 6.737e-07 ***
## x3         1   2.31    2.31   0.354    0.5597
## Residuals 17 111.11    6.54
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(model4)  #type I
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x1         1 352.27  352.27 57.2768 1.131e-06 ***
## x2         1  33.17   33.17  5.3931   0.03373 *
## x3         1  11.55   11.55  1.8773   0.18956
## Residuals 16  98.40    6.15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(model4, type="II")  #type II
```

```
## Anova Table (Type II tests)
##
## Response: y
##           Sum Sq Df F value Pr(>F)
## x1        12.705  1  2.0657 0.1699
## x2         7.529  1  1.2242 0.2849
## x3        11.546  1  1.8773 0.1896
## Residuals 98.405 16
```

```
model7<-lm(y~x3+x2+x1, bodyfat)
anova(model7)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x3         1  10.05   10.05  1.6343    0.2193
## x2         1 374.23  374.23 60.8471 7.684e-07 ***
## x1         1  12.70   12.70  2.0657    0.1699
## Residuals 16  98.40    6.15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model7)
```

```
##
## Call:
## lm(formula = y ~ x3 + x2 + x1, data = bodyfat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7263 -1.6111  0.3923  1.4656  4.1277
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  117.085     99.782   1.173    0.258
## x3            -2.186      1.595  -1.370    0.190
## x2            -2.857      2.582  -1.106    0.285
## x1             4.334      3.016   1.437    0.170
##
## Residual standard error: 2.48 on 16 degrees of freedom
## Multiple R-squared:  0.8014, Adjusted R-squared:  0.7641
## F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06
```

```
model1<-lm(y~x1, bodyfat)
#anova(model1)
#summary(model1)
typeIrs<-summary(model1)$r.squared
#adjustedtypeIrs<-summary(model1)$adj.r.squared
#typeIrs
#adjustedtypeIrs


model2<-lm(y~x2, bodyfat)

#anova(model2)
#summary(model2)

model3<-lm(y~x1+x2, bodyfat)
#anova(model3)
#summary(model3)

model4<-lm(y~x1+x2+x3, bodyfat)
anova(model4)
```

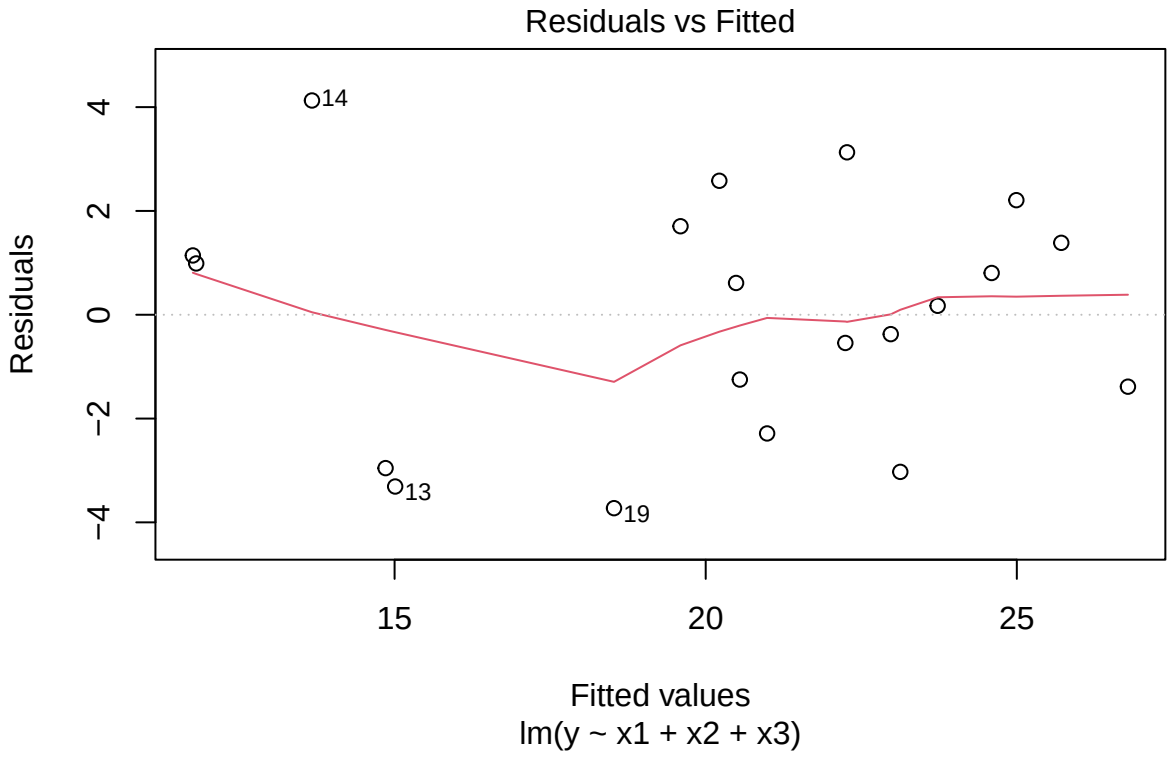
```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x1         1 352.27  352.27 57.2768 1.131e-06 ***
## x2         1  33.17   33.17  5.3931   0.03373 *
## x3         1  11.55   11.55  1.8773   0.18956
## Residuals 16  98.40    6.15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

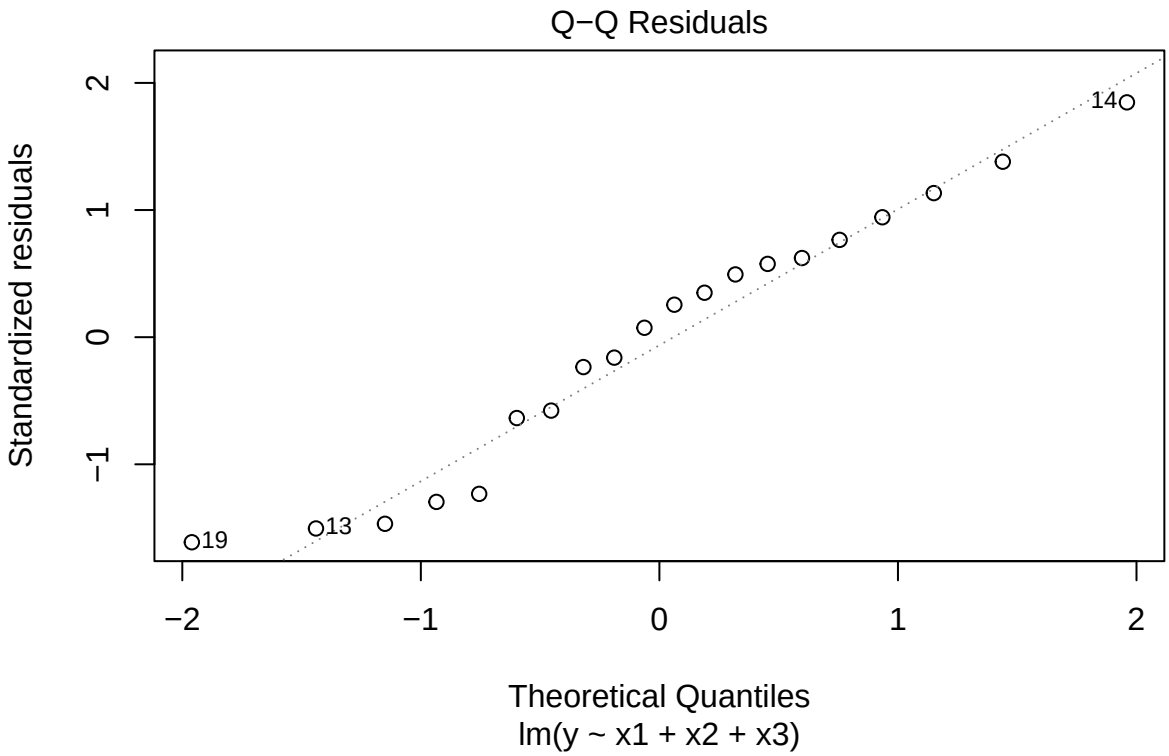
```
summary(model4)
```

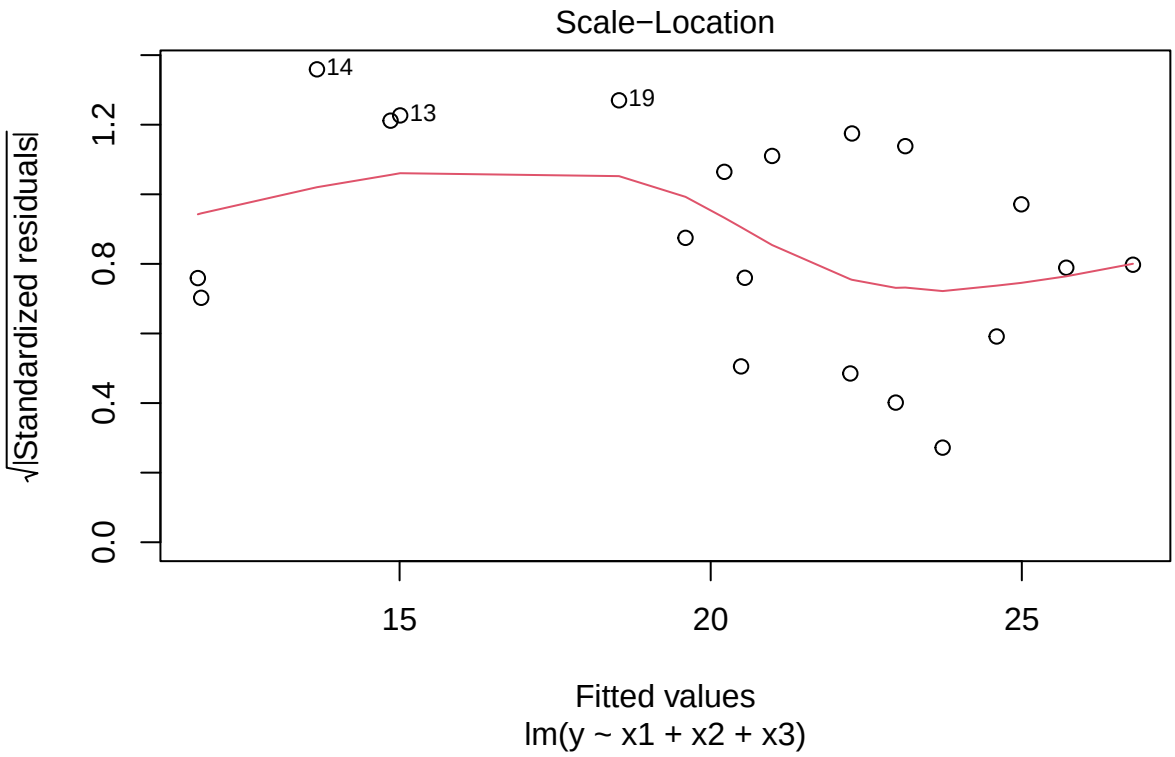
```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3, data = bodyfat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7263 -1.6111  0.3923  1.4656  4.1277
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  117.085     99.782   1.173    0.258
```

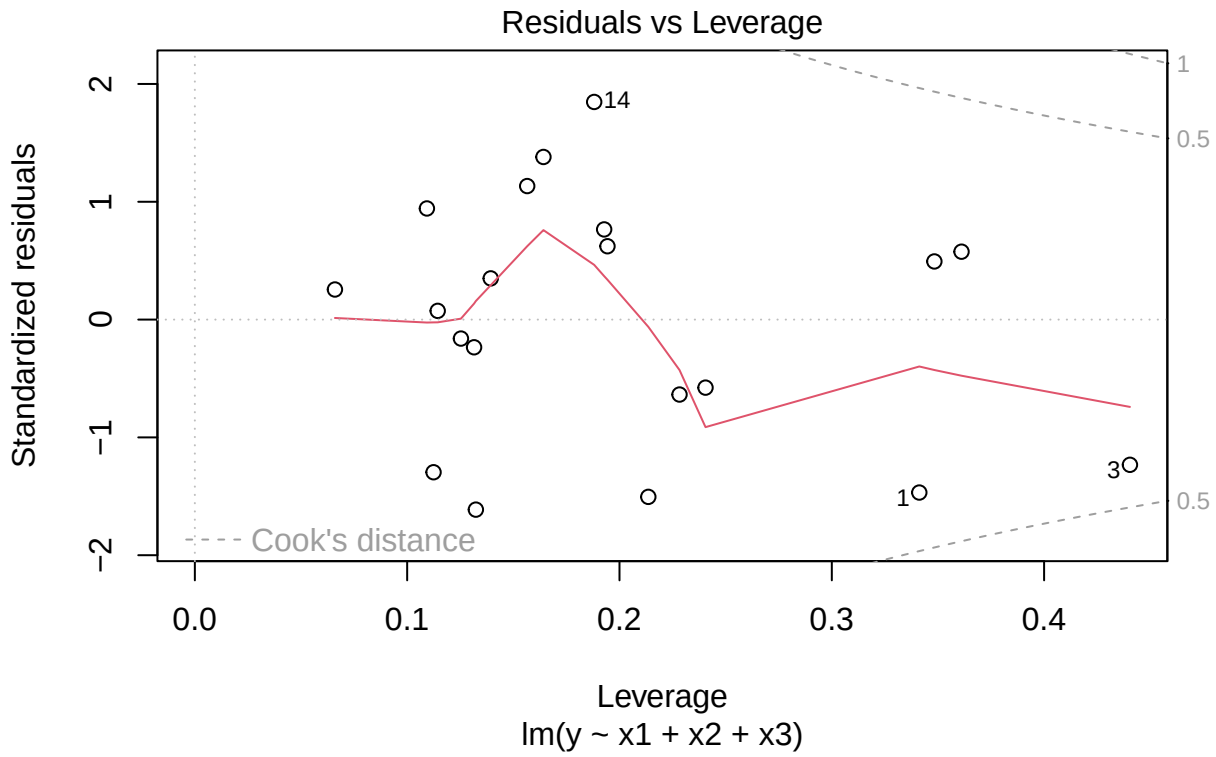
```
## x1             4.334      3.016   1.437    0.170
## x2            -2.857      2.582  -1.106    0.285
## x3            -2.186      1.595  -1.370    0.190
##
## Residual standard error: 2.48 on 16 degrees of freedom
## Multiple R-squared:  0.8014, Adjusted R-squared:  0.7641
## F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06
```

```
plot(model4)
```




Q−Q Residuals
Standardized residuals
2
1
0
−1
19
13
14
−2
−1
0
1
2
Theoretical Quantiles
lm(y ~ x1 + x2 + x3)


Scale−Location
√|Standardized residuals|
0.0
0.4
0.8
1.2
14
13
19
15
20
25
Fitted values
lm(y ~ x1 + x2 + x3)




```
model5<-lm(y~x3+x2, bodyfat)
#tail(anova(model1)$"Sum Sq",1)  #use the tail function to get just the SSE or residual which is the la
#anova(model3)
#anova(model3)$"Sum Sq"
#Anova(model3)

model6<-lm(y~x3, bodyfat)
#anova(model6)
#summary(model6)

model7<-lm(y~x3+x2+x1, bodyfat)
anova(model7)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x3         1  10.05   10.05  1.6343    0.2193
## x2         1 374.23  374.23 60.8471 7.684e-07 ***
## x1         1  12.70   12.70  2.0657    0.1699
## Residuals 16  98.40    6.15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model7)
```

```
##
```

```
## Call:
## lm(formula = y ~ x3 + x2 + x1, data = bodyfat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7263 -1.6111  0.3923  1.4656  4.1277
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  117.085     99.782   1.173    0.258
## x3            -2.186      1.595  -1.370    0.190
## x2            -2.857      2.582  -1.106    0.285
## x1             4.334      3.016   1.437    0.170
##
## Residual standard error: 2.48 on 16 degrees of freedom
## Multiple R-squared:  0.8014, Adjusted R-squared:  0.7641
## F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06
```

```
anova(model5)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x3         1  10.05   10.05  1.5379    0.2318
## x2         1 374.23  374.23 57.2576 7.722e-07 ***
## Residuals 17 111.11    6.54
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model5)
```

```
##
## Call:
## lm(formula = y ~ x3 + x2, data = bodyfat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0777 -1.8296  0.1893  1.3545  4.1275
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -25.99695    6.99732  -3.715  0.00172 **
## x3            0.09603    0.16139   0.595  0.55968
## x2            0.85088    0.11245   7.567 7.72e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.557 on 17 degrees of freedom
## Multiple R-squared:  0.7757, Adjusted R-squared:  0.7493
## F-statistic:  29.4 on 2 and 17 DF,  p-value: 3.033e-06
```

```
PartialR2<-function(model.full, model.reduced){
  anova.full<-anova(model.full)
  anova.reduced<-anova(model.reduced)
```

```
  sse.full<-tail(anova.full$"Sum Sq", 1)
  sse.reduced<-tail(anova.reduced$"Sum Sq", 1)

  pR2<-(sse.reduced-sse.full)/sse.reduced

  return(pR2)
}

#PartialR2(model4,model5) #computing X2 enterring X1 X3

model8<-lm(y~x1+x3+x2,bodyfat)
model9<-lm(y~x1+x3,bodyfat)
PartialR2(model8,model9)
```

```
## [1] 0.07107507
```

```
#PartialR2(model4,model5) #computing X3 enterring X1 X2
model10<-lm(y~x1+x2+x3,bodyfat)
model11<-lm(y~x1+x2,bodyfat)
PartialR2(model10,model11)
```

```
## [1] 0.1050097
```

```
model12<-lm(y~x1+x3, bodyfat)
anova(model12)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x1         1 352.27  352.27 56.5312 8.406e-07 ***
## x3         1  37.19   37.19  5.9674   0.02579 *
## Residuals 17 105.93    6.23
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model1<-lm(y~x2, bodyfat)
summary(model1)
```

```
##
## Call:
## lm(formula = y ~ x2, data = bodyfat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4949 -1.5671  0.1241  1.3362  4.4084
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -23.6345     5.6574  -4.178 0.000566 ***
## x2            0.8565     0.1100   7.786  3.6e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.51 on 18 degrees of freedom
## Multiple R-squared:  0.771,  Adjusted R-squared:  0.7583
```

```
## F-statistic: 60.62 on 1 and 18 DF,  p-value: 3.6e-07
```

```
anova(model1)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value  Pr(>F)
## x2         1 381.97  381.97  60.617 3.6e-07 ***
## Residuals 18 113.42    6.30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model2<-lm(y~x1+x2, bodyfat)
PartialR2(model2, model1)  # computer the coefficient of partial determination between Y and x1, when x
```

```
## [1] 0.03061875
```

```
model1<-lm(y~x1,bodyfat)
model2<-lm(y~x2,bodyfat)
model3<-lm(y~x1+x2, bodyfat)
model31<-lm(y~x2+x1, bodyfat)
model4<-lm(y~x1+x2+x3, bodyfat)
model5<-lm(y~x3+x2+x1,bodyfat)
model6<-lm(y~x2+x1+x3, bodyfat)

#anova(model1)
#summary(model1)
#anova(model2)
#anova(model3)
#anova(model31)
#Anova(model3)

#anova(model4)
#anova(model5)
#anova(model6)
Anova(model4)
```

```
## Anova Table (Type II tests)
##
## Response: y
##           Sum Sq Df F value Pr(>F)
## x1        12.705  1  2.0657 0.1699
## x2         7.529  1  1.2242 0.2849
## x3        11.546  1  1.8773 0.1896
## Residuals 98.405 16
```

```
#Anova(model4)
```

```
model1<-lm(y~x1+x2+x3, bodyfat)
modle2<-lm(y~x1+x2+x3+(x1+x2), bodyfat)
anova(model1)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## x1         1 352.27  352.27 57.2768 1.131e-06 ***
## x2         1  33.17   33.17  5.3931   0.03373 *
## x3         1  11.55   11.55  1.8773   0.18956
## Residuals 16  98.40    6.15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model1)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3, data = bodyfat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7263 -1.6111  0.3923  1.4656  4.1277
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  117.085     99.782   1.173    0.258
## x1             4.334      3.016   1.437    0.170
## x2            -2.857      2.582  -1.106    0.285
## x3            -2.186      1.595  -1.370    0.190
##
## Residual standard error: 2.48 on 16 degrees of freedom
## Multiple R-squared:  0.8014, Adjusted R-squared:  0.7641
## F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06
```

```
anova(model2)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value  Pr(>F)
## x2         1 381.97  381.97  60.617 3.6e-07 ***
## Residuals 18 113.42    6.30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model2)
```

```
##
## Call:
## lm(formula = y ~ x2, data = bodyfat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4949 -1.5671  0.1241  1.3362  4.4084
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -23.6345     5.6574  -4.178 0.000566 ***
## x2            0.8565     0.1100   7.786  3.6e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 2.51 on 18 degrees of freedom
## Multiple R-squared:  0.771,  Adjusted R-squared:  0.7583
## F-statistic: 60.62 on 1 and 18 DF,  p-value: 3.6e-07
```

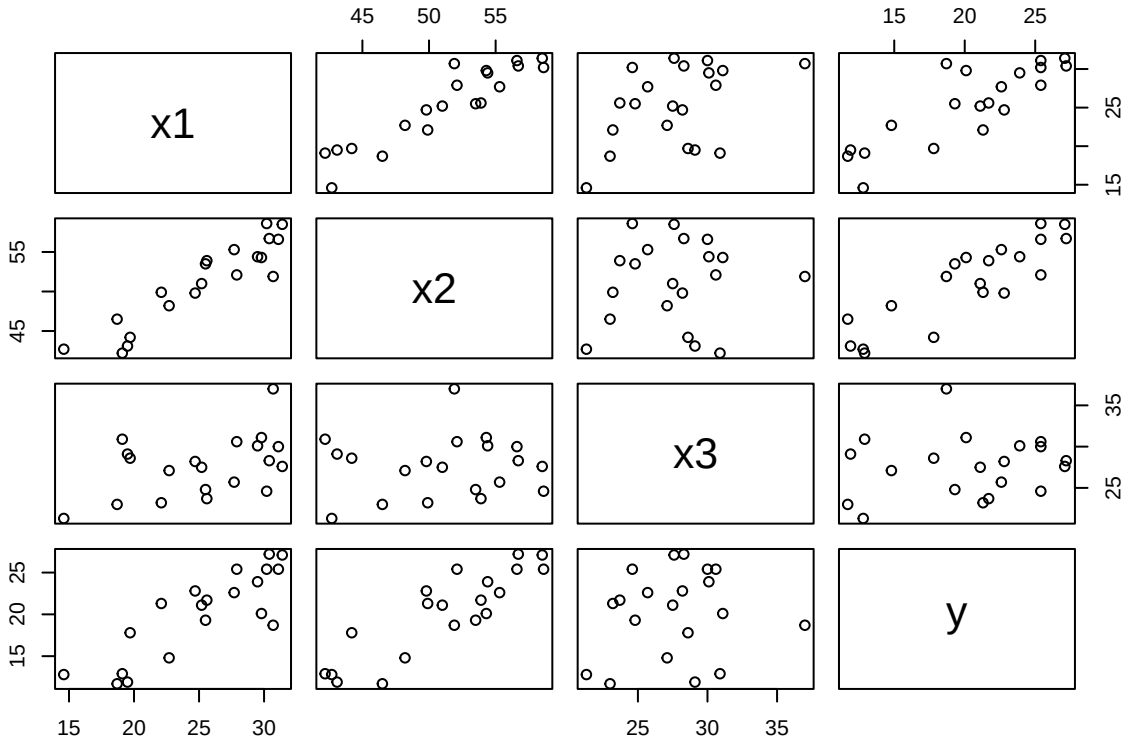
```
model1<-lm(y~x1+x2+x3, bodyfat)
stdmodel1<-lm.beta(model1)
summary(stdmodel1)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3, data = bodyfat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7263 -1.6111  0.3923  1.4656  4.1277
##
## Coefficients:
##             Estimate Standardized Std. Error t value Pr(>|t|)
## (Intercept)  117.085           NA     99.782   1.173    0.258
## x1             4.334        4.264      3.016   1.437    0.170
## x2            -2.857       -2.929      2.582  -1.106    0.285
## x3            -2.186       -1.561      1.595  -1.370    0.190
##
## Residual standard error: 2.48 on 16 degrees of freedom
## Multiple R-squared:  0.8014, Adjusted R-squared:  0.7641
## F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06
```

```
summary(model1)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3, data = bodyfat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7263 -1.6111  0.3923  1.4656  4.1277
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  117.085     99.782   1.173    0.258
## x1             4.334      3.016   1.437    0.170
## x2            -2.857      2.582  -1.106    0.285
## x3            -2.186      1.595  -1.370    0.190
##
## Residual standard error: 2.48 on 16 degrees of freedom
## Multiple R-squared:  0.8014, Adjusted R-squared:  0.7641
## F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06
```

```
plot(bodyfat)
```




```
cor(bodyfat)
```

```
##           x1        x2        x3         y
## x1 1.0000000 0.9238425 0.4577772 0.8432654
## x2 0.9238425 1.0000000 0.0846675 0.8780896
## x3 0.4577772 0.0846675 1.0000000 0.1424440
## y  0.8432654 0.8780896 0.1424440 1.0000000
```

```
model<-lm(x3~x1+x2, bodyfat)
summary(model)
```

```
##
## Call:
## lm(formula = x3 ~ x1 + x2, data = bodyfat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.58200 -0.30625  0.02592  0.29526  0.56102
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 62.33083    1.23934   50.29   <2e-16 ***
## x1           1.88089    0.04498   41.82   <2e-16 ***
## x2          -1.60850    0.04316  -37.26   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.377 on 17 degrees of freedom
## Multiple R-squared:  0.9904, Adjusted R-squared:  0.9893
## F-statistic: 880.7 on 2 and 17 DF,  p-value: < 2.2e-16
```

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: x3
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## x1         1  52.963  52.963  372.66 5.333e-13 ***
## x2         1 197.353 197.353 1388.64 < 2.2e-16 ***
## Residuals 17   2.416   0.142
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model1<-lm(y~x1, bodyfat)
anova(model1)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x1         1 352.27  352.27  44.305 3.024e-06 ***
## Residuals 18 143.12    7.95
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#summary(model1)
model2<-lm(y~x2+x1, bodyfat)
anova(model2)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x2         1 381.97  381.97  59.057 6.281e-07 ***
## x1         1   3.47    3.47   0.537    0.4737
## Residuals 17 109.95    6.47
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#summary(model2)
model3<-lm(y~x1+x2+x3, bodyfat)
anova(model3)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x1         1 352.27  352.27 57.2768 1.131e-06 ***
## x2         1  33.17   33.17  5.3931   0.03373 *
## x3         1  11.55   11.55  1.8773   0.18956
## Residuals 16  98.40    6.15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#summary(model3)

newdata1<-data.frame(x1=25)
predict(model1,newdata1,interval="confidence", se.fit=TRUE)
```

```
## $fit
##        fit      lwr     upr
## 1 19.93356 18.60632 21.2608
##
## $se.fit
## [1] 0.6317416
##
## $df
## [1] 18
##
## $residual.scale
## [1] 2.819769
```

```r
model<-lm(y~x1+x2, data=bodyfat)
residuals.lm(model)
```

```
##             1             2             3             4             5
## -1.6827093112  3.6429311788 -3.1759701405 -3.1584651200 -0.0002886579
##             6             7             8             9            10
## -0.3608155187  0.7161991891  4.0147327554  2.6551057360 -2.4748115410
##            11            12            13            14            15
##  0.3358063798  2.2255110139 -3.9468613463  3.4474561945  0.5705871038
##            16            17            18            19            20
##  0.6422984777 -0.8509464751 -0.7829198812 -2.8572887647  1.0404487275
```

```r
lm.influence(model)$hat   #calcualte the diagonal elements of the Hat matrix
```

```
##          1          2          3          4          5          6          7
## 0.20101253 0.05889478 0.37193301 0.11094009 0.24801034 0.12861620 0.15551745
##          8          9         10         11         12         13         14
## 0.09628780 0.11463564 0.11024435 0.12033655 0.10926629 0.17838181 0.14800684
##         15         16         17         18         19         20
## 0.33321201 0.09527739 0.10559466 0.19679280 0.06695419 0.05008526
```

```r
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, data = bodyfat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9469 -1.8807  0.1678  1.3367  4.0147
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -19.1742     8.3606  -2.293   0.0348 *
## x1            0.2224     0.3034   0.733   0.4737
## x2            0.6594     0.2912   2.265   0.0369 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2.543 on 17 degrees of freedom
## Multiple R-squared:  0.7781, Adjusted R-squared:  0.7519
## F-statistic:  29.8 on 2 and 17 DF,  p-value: 2.774e-06
```

```
t<-anova(model)
t$`Sum Sq`[3]   #= calcuate the SSE
```

```
## [1] 109.9508
```

```
t$`Mean Sq`[3]
```

```
## [1] 6.467694
```

```
lm.influence(model)
```

```
## $hat
##          1          2          3          4          5          6          7
## 0.20101253 0.05889478 0.37193301 0.11094009 0.24801034 0.12861620 0.15551745
##          8          9         10         11         12         13         14
## 0.09628780 0.11463564 0.11024435 0.12033655 0.10926629 0.17838181 0.14800684
##         15         16         17         18         19         20
## 0.33321201 0.09527739 0.10559466 0.19679280 0.06695419 0.05008526
##
## $coefficients
##      (Intercept)            x1            x2
## 1  -2.5873120410 -4.045756e-02  6.851256e-02
## 2   1.3885856142  3.359106e-02 -3.996603e-02
## 3  -6.7460811783 -3.417891e-01  2.959198e-01
## 4  -0.8298018102 -8.699579e-02  5.576707e-02
## 5  -0.0005491464 -9.548311e-06  1.507863e-05
## 6   0.3417033826  1.252797e-02 -1.327784e-02
## 7  -0.6662800194 -4.869870e-03  1.625790e-02
## 8   2.0621886739  1.119746e-01 -9.133441e-02
## 9  -1.2562286444 -8.876192e-02  7.137567e-02
## 10  1.9837372912  7.407218e-02 -7.811621e-02
## 11 -0.0776962164  5.331751e-03 -7.452907e-04
## 12 -1.0957798277  6.844439e-03  2.047112e-02
## 13  0.9361849968  1.685643e-01 -1.063493e-01
## 14  3.6377705378  3.307612e-02 -8.349511e-02
## 15 -0.0258331550 -3.893438e-02  2.059515e-02
## 16  0.0800533575  1.345694e-02 -7.525586e-03
## 17  0.6827037689  1.715290e-02 -2.275412e-02
## 18  1.1340627287  2.347924e-02 -3.472625e-02
## 19 -1.0715507527 -1.221911e-03  1.855296e-02
## 20  0.0873645353  7.127872e-04 -9.895689e-04
##
## $sigma
##        1        2        3        4        5        6        7        8
## 2.578843 2.447567 2.422431 2.484075 2.621436 2.619654 2.614185 2.399419
##        9       10       11       12       13       14       15       16
## 2.524733 2.538051 2.619907 2.554290 2.384730 2.449505 2.615608 2.615994
##       17       18       19       20
## 2.611767 2.612322 2.514966 2.607815
##
## $wt.res
```

```
##             1             2             3             4             5
## -1.6827093112  3.6429311788 -3.1759701405 -3.1584651200 -0.0002886579
##             6             7             8             9            10
## -0.3608155187  0.7161991891  4.0147327554  2.6551057360 -2.4748115410
##            11            12            13            14            15
##  0.3358063798  2.2255110139 -3.9468613463  3.4474561945  0.5705871038
##            16            17            18            19            20
##  0.6422984777 -0.8509464751 -0.7829198812 -2.8572887647  1.0404487275
```

```
cell<-read.csv("../../datasets/battery_life.csv")
colnames(cell)<-c("cycle","rate","temp")
cell<-cbind(cell, cell$rate^2, cell$temp^2)
colnames(cell)<-c("cycle","rate","temp","rate2","temp2")
#cor(cell[,2:5])
#summary(cell)
mean(cell$rate)
```

```
## [1] 1.04
```

```
mean(cell$temp)
```

```
## [1] 21
```

```
sd(cell$rate)
```

```
## [1] 0.2951459
```

```
sd(cell$temp)
```

```
## [1] 7.378648
```

```
anova(lm(cycle~rate+temp+rate2+temp2+rate*temp, data=cell))
```

```
## Analysis of Variance Table
##
## Response: cycle
##           Df  Sum Sq Mean Sq F value   Pr(>F)
## rate       1 26331.6 26331.6 35.0463 0.004078 **
## temp       1 26051.2 26051.2 34.6731 0.004159 **
## rate2      1  2001.0  2001.0  2.6633 0.178026
## temp2      1   138.1   138.1  0.1838 0.690178
## rate:temp  1  2546.4  2546.4  3.3892 0.139442
## Residuals  4  3005.3   751.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(lm(cycle~rate+temp+rate2+temp2+rate*temp, data=cell))
```

```
## Anova Table (Type II tests)
##
## Response: cycle
##           Sum Sq Df F value  Pr(>F)
## rate      3856.2  1  5.1325 0.08616 .
## temp      1402.0  1  1.8660 0.24370
## rate2     3527.8  1  4.6954 0.09614 .
## temp2       20.0  1  0.0266 0.87825
## rate:temp 2546.4  1  3.3892 0.13944
## Residuals 3005.3  4
## ---
```

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

```
#summary(lm(cycle~rate+temp+rate2+temp2+rate*temp, data=cell))
```

```
celln<-cbind(cell$cycle, (cell$rate-1)/0.4,(cell$temp-20)/10)
celln<-cbind(celln[,1],celln[,2:3],celln[,2]^2,celln[,3]^2,celln[,2]*celln[,3])
colnames(celln)<-c("Y","x1","x2","x11","x22","x12")
celln<-data.frame(celln)
```

```
lm(Y~x1+x2+x11+x22+x12, data=celln)
```

```
##
## Call:
## lm(formula = Y ~ x1 + x2 + x11 + x22 + x12, data = celln)
##
## Coefficients:
## (Intercept)           x1           x2          x11          x22          x12
##     157.362      -73.188       58.145       41.096        3.096       37.532
```

```
anova(lm(Y~x1+x2+x11+x22+x12, data=celln))
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df  Sum Sq Mean Sq F value   Pr(>F)
## x1         1 26331.6 26331.6 35.0463 0.004078 **
## x2         1 26051.2 26051.2 34.6731 0.004159 **
## x11        1  2001.0  2001.0  2.6633 0.178026
## x22        1   138.1   138.1  0.1838 0.690178
## x12        1  2546.4  2546.4  3.3892 0.139442
## Residuals  4  3005.3   751.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm(Y~x1+x2+x11+x22+x12, data=celln))
```

```
##
## Call:
## lm(formula = Y ~ x1 + x2 + x11 + x22 + x12, data = celln)
##
## Residuals:
##         1         2         3         4         5         6         7         8
## -16.31206  16.31206  16.35461  -0.36170 -26.36170  26.63830 -16.26950 -16.35461
##         9        10
##  16.39716  -0.04255
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  157.362     14.416  10.916  0.00040 ***
## x1          -73.188     15.049  -4.863  0.00826 **
## x2           58.145     15.049   3.864  0.01809 *
## x11          41.096     18.965   2.167  0.09614 .
## x22           3.096     18.965   0.163  0.87825
## x12          37.532     20.387   1.841  0.13944
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 27.41 on 4 degrees of freedom
## Multiple R-squared:   0.95,  Adjusted R-squared:  0.8874
## F-statistic: 15.19 on 5 and 4 DF,  p-value: 0.01041
```

```
Anova(lm(Y~x1+x2+x11+x22+x12, data=celln), type="II")
```

```
## Anova Table (Type II tests)
##
## Response: Y
##            Sum Sq Df F value   Pr(>F)
## x1        17770.9  1 23.6524 0.008259 **
## x2        11216.6  1 14.9288 0.018089 *
## x11        3527.8  1  4.6954 0.096144 .
## x22          20.0  1  0.0266 0.878251
## x12        2546.4  1  3.3892 0.139442
## Residuals  3005.3  4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
reducedModel<-lm(Y~x1+x2+x11+x22+x12, data=celln)
fullModel<-lm(Y~factor(x1)*factor(x2)*factor(x11)*factor(x22)*factor(x12), data=celln)
anova(reducedModel, fullModel)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ x1 + x2 + x11 + x22 + x12
## Model 2: Y ~ factor(x1) * factor(x2) * factor(x11) * factor(x22) * factor(x12)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1      4 3005.3
## 2      2 1404.7  2    1600.7 1.1395 0.4674
```

```
anova(lm(Y~x1+x2, data=celln),lm(Y~x1+x2+x11+x22+x12, data=celln))
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ x1 + x2
## Model 2: Y ~ x1 + x2 + x11 + x22 + x12
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1      7 7690.9
## 2      4 3005.3  3    4685.5 2.0788 0.2457
```