HW5

Zeqiu.Yu

2022-11-01

Homework 5

Problem 7.5

```
input1 <- read.table("./BookDataSets/Chapter  6 Data Sets/CH06PR15.txt")
names(input1) <- c("Y","X1", "X2","X3")
```

(a.)

```
X2.fit <- lm(Y~X2, data = input1)
X1X2.fit <- lm(Y~X2+X1, data = input1)
X1X2X3.fit <- lm(Y~X2+X1+X3, data = input1)
anova(X2.fit)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value   Pr(>F)
## X2         1 4860.3  4860.3  25.132 9.23e-06 ***
## Residuals 44 8509.0   193.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(X1X2.fit)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## X2         1 4860.3  4860.3  45.305 3.161e-08 ***
## X1         1 3896.0  3896.0  36.317 3.348e-07 ***
## Residuals 43 4613.0   107.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(X1X2X3.fit)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## X2         1 4860.3  4860.3 48.0439 1.822e-08 ***
## X1         1 3896.0  3896.0 38.5126 2.008e-07 ***
## X3         1  364.2   364.2  3.5997   0.06468 .
## Residuals 42 4248.8   101.2
```

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

From the tabel above, $SSR(X_2) = 4860.3$, $SSR(X_1|X_2) = 3896.0$, $SSR(X_3|X_1, X_2) = 364.2$

(b.) $H_0 : \beta_3 = 0$

$H_a : \beta_3 \neq 0$ $F^* = \frac{\frac{SSE(R)-SSE(F)}{df_R-df_F}}{\frac{SSE(F)}{df_F}}$, If $F^* > F(1-0.025, 1, n-4)$, we will reject the null hypothesis. Otherwise, we accept it.

```r
DropX3 <- lm(Y~X1+X2, data = input1)
Full <- lm(Y~X1+X2+X3, data = input1)
anova(DropX3, Full)
```

```
## 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     43 4613.0
## 2     42 4248.8  1    364.16 3.5997 0.06468 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We find that p-value is 0.06468>0.025, which means $F^* > F(1-0.025, 1, 42)$. Hence, we accept the null Hypothesis.

Problem 7.9

$H_0 : \beta_1 = -1,\ \beta_2 = 0$

$H_a : \beta_1 = -1,\ \beta_2 = 0$, at least one of the equalities will not hold. $F^* = \frac{\frac{SSE(R)-SSE(F)}{df_R-df_F}}{\frac{SSE(F)}{df_F}}$. If the p-value is larger than 0.025, we will reject the null hypothesis. Otherwise, we accept it.

```r
ReducedModel <- lm(Y~X3, offset = -X1+0*X2, data = input1)
anova(ReducedModel, Full)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X3
## Model 2: Y ~ X1 + X2 + X3
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     44 4427.7
## 2     42 4248.8  2    178.81 0.8838 0.4208
```

The p-value is 0.4208 $>$ 0.025, we do not reject the null Hypothesis.

Problem 8.16

```r
input2.sub1 <- read.table("./BookDataSets/Chapter  1 Data Sets/CH01PR19.txt")
input2.sub2 <- read.table("./BookDataSets/Chapter  8 Data Sets/CH08PR16.txt")
input2 <- cbind(input2.sub1, input2.sub2)
names(input2) <- c("Y","X1", "X2")
```

(a.)
The parameter β_2 measures the differential effect of whether the field is decided or not. β_1 measures the change of the response variable w.r.t 1 unit change in entrance test score. Given fixed amount of entrance exam, if the student has a major field of concentration, $\beta 2$ measures the average GPA increase. β_0 measures

average GPA when the entrance exam is 0 and the major is not decided.
(b.)

```
fit2 <- lm(Y~X1+X2, data = input2)
summary(fit2)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2, data = input2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.70304 -0.35574  0.02541  0.45747  1.25037
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.19842    0.33886   6.488 2.18e-09 ***
## X1           0.03789    0.01285   2.949  0.00385 **
## X2          -0.09430    0.11997  -0.786  0.43341
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6241 on 117 degrees of freedom
## Multiple R-squared:  0.07749,    Adjusted R-squared:  0.06172
## F-statistic: 4.914 on 2 and 117 DF,  p-value: 0.008928
```

Then $\hat{Y} = 2.19842 + 0.03789X_1 - 0.09430X_2$.
(c.)

$H_0 : \beta_2 = 0$
$H_a : \beta_2 \neq 0$ $F^* = \frac{\frac{SSE(R)-SSE(F)}{df_R-df_F}}{\frac{SSE(F)}{df_F}}$. If the p-value is larger than 0.01, we will reject the null hypothesis. Otherwise, we accept it.

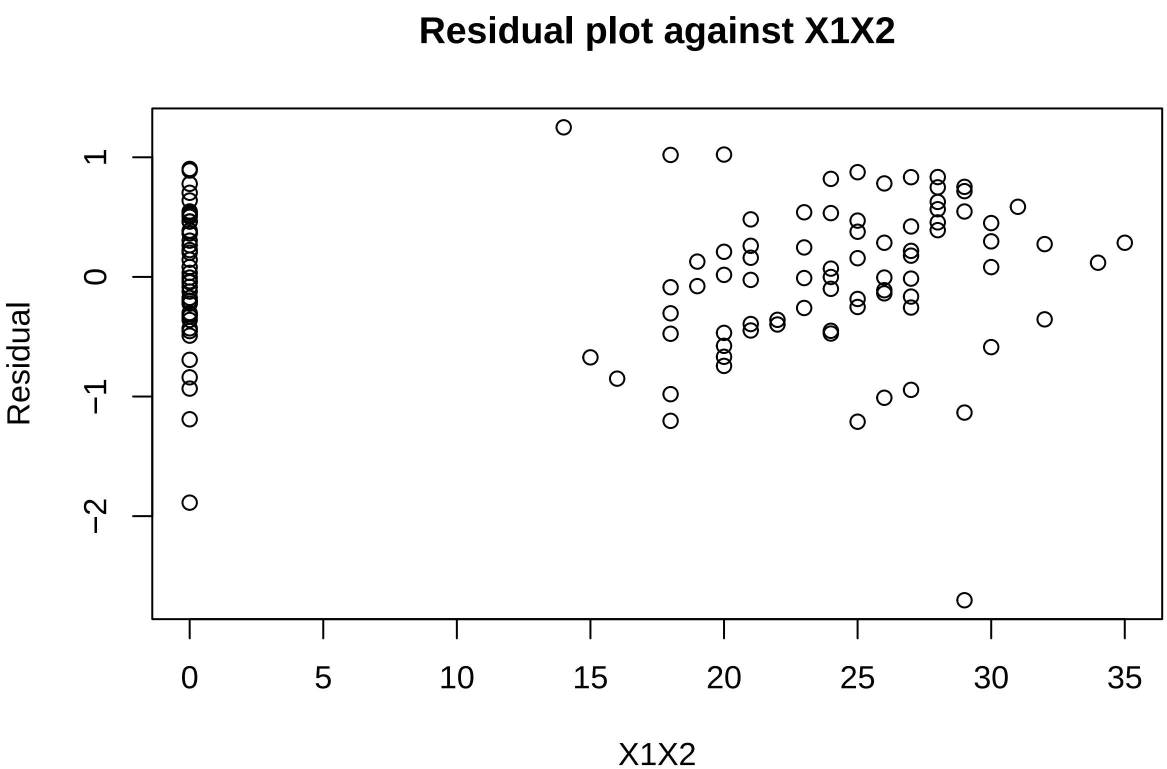
```
fit2.dropX2 <- lm(Y~X1, data = input2)
anova(fit2.dropX2,fit2)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X1
## Model 2: Y ~ X1 + X2
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    118 45.818
## 2    117 45.577  1   0.24071 0.6179 0.4334
```

Hence, we do not reject the null hypothesis.

(d.)

```
e <- fit2$residuals
par(mfrow=c(1,1))
X1X2 <- input2$X1*input2$X2
plot(e~X1X2, xlab="X1X2", ylab="Residual", main="Residual plot against X1X2")
```



The answer is no!

Problem 8.20

```
fit3 <- lm(Y~.^2, data = input2)
summary(fit3)
```

```
##
## Call:
## lm(formula = Y ~ .^2, data = input2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.80187 -0.31392  0.04451  0.44337  1.47544
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.226318   0.549428   5.872 4.18e-08 ***
## X1          -0.002757   0.021405  -0.129   0.8977
## X2          -1.649577   0.672197  -2.454   0.0156 *
## X1:X2        0.062245   0.026487   2.350   0.0205 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6124 on 116 degrees of freedom
## Multiple R-squared:  0.1194, Adjusted R-squared:  0.09664
## F-statistic: 5.244 on 3 and 116 DF,  p-value: 0.001982
```

Then $\hat{Y} = 3.226318 - 0.002757X_1 + 0.062245X_2 + 0.062245X_1X_2$.
(b.)
$H_0 : \beta_3 = 0$

$H_a : \beta_3 \neq 0$ $F^* = \frac{\frac{SSE(R)-SSE(F)}{df_R - df_F}}{\frac{SSE(F)}{df_F}}$. . If the p-value is larger than 0.05, we will reject the null hypothesis. Otherwise, we accept it.

```
anova(fit2, fit3)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X1 + X2
## Model 2: Y ~ (X1 + X2)^2
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    117 45.577
## 2    116 43.506  1    2.0713 5.5226 0.02046 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-value is smaller than 0.05, we reject the null hypothesis and accept the alternative hypothesis.
If $X_2 = 0$, E[Y] $= \beta_0 + \beta_1 X_1$.
If $X_2 = 1$, E[Y] $= (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X_1$
The interaction effect in the model is not additive. The average change of response with a unit change in X_1 is not the same with X_2 constant.

Problem 8.21

(a.)
Hard hat : E{Y}= $(\beta_0 + \beta_2) + \beta_1 X_1$
Bump cap : E{Y}= $(\beta_0 + \beta_3) + \beta_1 X_1$
None: E{Y} = $\beta_0 + \beta_1 X_1$

(b.)
(1) H_0: $\beta_3 = 0$; H_a: $\beta_3 < 0$ (2) H_0: $\beta_2 = \beta_3$; H_a: $\beta_2 \neq \beta_3$