Homework Chapter 7

Jinhong Du 15338039

7.3 Refer to Brand preference Problem 6.5.

a. Obtain the analysis of variance table that decomposes the regression sum of squares into extra sums of squares associated with X_1 and with X_2, given X_1.

```
data1 <- read.table("CH06PR05.txt",head=FALSE,col.names = c('Y','X1','X2'))
Y <- matrix(data1$Y)
n <- length(Y)
X1 <- data1$X1
X2 <- data1$X2

fit = lm(Y~X1+X2);
fit.aov <- anova(fit)
tab <- as.table(cbind(
  'SS' = c("SSR(X1, X2)"     = sum(fit.aov[1:2, 2]),
          "SSR(X1)"          = fit.aov[1, 2],
          "SSR(X2|X1)"       = fit.aov[2, 2],
          "SSE(X1,X2)"       = fit.aov[3, 2],
          "Total"            = sum(fit.aov[, 2])),

  'Df' = c(                    sum(fit.aov[1:2, 1]),
                               fit.aov[1, 1],
                               fit.aov[2, 1],
                               fit.aov[3, 1],
                               sum(fit.aov$Df)),

  'MS' = c(                    sum(fit.aov[1:2, 2]) / sum(fit.aov[1:2, 1]),
                               fit.aov[1, 3],
                               fit.aov[2, 3],
                               fit.aov[3, 3],
                               NA)
))

round(tab, 2)
```

```
##                   SS      Df      MS
## SSR(X1, X2) 1872.70    2.00  936.35
## SSR(X1)     1566.45    1.00 1566.45
## SSR(X2|X1)   306.25    1.00  306.25
## SSE(X1,X2)    94.30   13.00    7.25
## Total       1967.00   15.00
```

b. Test whether X_2 can be dropped from the regression model given that X_1 is retained. Use the F^* test statistic and level of significance .01. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

```
anova(update(fit, . ~ . - X2), fit, test='F')
```

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

```
##
## Model 1: Y ~ X1
## Model 2: Y ~ X1 + X2
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     14 400.55
## 2     13  94.30  1    306.25 42.219 2.011e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

or

```r
drop1(fit, test = "F")
```

```
## Single term deletions
##
## Model:
## Y ~ X1 + X2
##        Df Sum of Sq     RSS    AIC F value    Pr(>F)
## <none>                94.30 34.382
## X1      1   1566.45 1660.75 78.279 215.947 1.778e-09 ***
## X2      1    306.25  400.55 55.524  42.219 2.011e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$H_0 : \beta_2 = 0 \qquad H_a : \beta_2 \neq 0$$

$$F^* = \frac{\frac{SSR(X_2|X_1)}{1}}{\frac{SSE(X_1, X_2)}{13}} = 42.219 \overset{H_0}{\sim} F(1, 13)$$

The decision rule is

If $F^* \leqslant F(0.99, 1, 13) = 9.07$, then conclude H_0;

If $F^* > F(0.99, 1, 13) = 9.07$, then conclude H_a;

Since $F^* = 42.219 > 9.07$, conclude H_a.

7.10 Refer to Commercial properties Problem 6.18. Test whether $\beta_1 = -.1$ and $\beta_2 = .4$; Use $\alpha = .01$. State the alternatives, full and reduced models, decision rule, and conclusion.

```r
data2 <- read.table("CH06PR18.txt",head=FALSE,col.names = c('Y','X1','X2','X3','X4'))
Y <- data2$Y
n <- length(Y)
X1 <- data2$X1
X2 <- data2$X2
X3 <- data2$X3
X4 <- data2$X4
fit2 = lm(Y~X1+X2+X3+X4);
Yr <- Y+0.1*X1-0.4*X2
fit2reduce = lm(Yr~X3+X4)
SSEF <- sum(fit2$residuals^2)
SSER <- sum(fit2reduce$residuals^2)
dfF <- fit2$df.residual
dfR <- fit2reduce$df.residual
```

```r
F <- ((SSEF-SSER)/(dfF-dfR))/ (SSEF/dfF)
print(sprintf('F* is %f',F))
```

```
## [1] "F* is 4.607640"
```

```r
print(sprintf('F(0.99,%d,%d) is %f',dfR-dfF,dfF,qf(p=0.99,df1=(dfR-dfF),df2=dfF)))
```

```
## [1] "F(0.99,2,76) is 4.895840"
```

$$H_0 : \beta_1 = -0.1, \beta_2 = 0.4 \qquad H_a : \beta_1 \neq -0.1 \text{ or } \beta_2 \neq 0.4$$

$$F^* = \frac{\frac{SSE_F - SSE_R}{df_F - df_R}}{\frac{SSE_F}{df_F}} \overset{H_0}{\sim} F(df_R - df_F, df_F)$$

The decision rule is

If $F^* \leqslant F(0.99, 2, 76) = 4.895840$, then conclude H_0;

If $F^* > F(0.99, 2, 76) = 4.895840$, then conclude H_a;

Since $F^* = 4.607640 < 4.895840$, conclude H_0.

7.12 Refer to Brand preference Problem 6.5. Calculate R^2_{Y1}, R^2_{Y2}, R^2_{12}, $R^2_{Y1|2}$, $R^2_{Y2|1}$ and R^2. Explain what each coefficient measures and interpret your results.

```r
data1 <- read.table("CH06PR05.txt",head=FALSE,col.names = c('Y','X1','X2'))
Y <- data1$Y
n <- length(Y)
X1 <- data1$X1
X2 <- data1$X2

fit = lm(Y~X1+X2);
fit.aov <- anova(fit)
tab <- as.table(cbind(
  'R2' = c(
          "Y1"               = cor(Y,X1)^2,
          "Y2"               = cor(Y,X2)^2,
          "12"               = cor(X1,X2)^2,
          "Y1|2"             = fit.aov[1, 2]/(fit.aov[1, 2]+fit.aov[3,2]),
          "Y2|1"             = fit.aov[2, 2]/(fit.aov[2, 2]+fit.aov[3,2]),
          "R2"               = sum(fit.aov[1:2,2])/sum(fit.aov[, 2])
  )
))

round(tab, 4)
```

```
##          R2
## Y1   0.7964
## Y2   0.1557
## 12   0.0000
## Y1|2 0.9432
## Y2|1 0.7646
## R2   0.9521
```

$$
\begin{aligned}
R^2_{Y1|2} &= \frac{SSE(X_2) - SSE(X_1, X_2)}{SSR(X_2)} \\
&= \frac{SSR(X_1|X_2)}{SSE(X_2)} \\
R^2_{Y2|1} &= \frac{SSE(X_1) - SSE(X_1, X_2)}{SSR(X_1)} \\
&= \frac{SSR(X_2|X_1)}{SSE(X_1)} \\
R^2 &= \frac{SSR(X_1, X_2)}{SSTO}
\end{aligned}
$$

7.16. Refer to Brand preference Problem 6.5.

a. Transform the variables by means of the correlation transformation (7.44) and fit the standardized regression model (7.45).

```
zY <- (Y-mean(Y))/sqrt(var(Y)*(n-1))
zX1 <- (X1- mean(X1))/sqrt(var(X1)*(n-1))
zX2 <- (X2- mean(X2))/sqrt(var(X2)*(n-1))
fitz <- lm(zY~zX1+zX2)
summary(fitz)
```

```
##
## Call:
## lm(formula = zY ~ zX1 + zX2)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.099209 -0.039740  0.000564  0.035794  0.094699
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.238e-17  1.518e-02   0.000        1
## zX1          8.924e-01  6.073e-02  14.695 1.78e-09 ***
## zX2          3.946e-01  6.073e-02   6.498 2.01e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06073 on 13 degrees of freedom
## Multiple R-squared:  0.9521, Adjusted R-squared:  0.9447
## F-statistic: 129.1 on 2 and 13 DF,  p-value: 2.658e-09
```

$$
\begin{aligned}
Y^* &= \frac{Y - \overline{Y}}{\sqrt{SSE(Y)}} \\
X_i^* &= \frac{X_i - \overline{X_i}}{\sqrt{SSE(X_i)}}
\end{aligned}
$$

b. Interpret the standardized regression coefficient b_1^*.

Simple scaling factors involving ratios of standard deviations. Therefore, b_1^* is proportional to b_1.

c. Transform the estimated standardized regression coefficients by means of (7.53) back to the ones for the fitted regression model in the original variables. Verify that they are the same as the ones obtained in Problem 6.5b.

```
sY <- sqrt(var(Y))
sX1 <- sqrt(var(X1))
b1_ <- sY/sX1*fitz$coefficients[2]
b1_
```

```
##   zX1
## 4.425
```

```
fit$coefficients[2]
```

```
##    X1
## 4.425
```

$$b_1 = \frac{s_Y}{s_{X_1}} b_1^*$$

7.24. Refer to Brand preference Problem 6.5.

a. Fit first-order simple linear regression model (2.1) for relating brand liking (Y) to moisture content (X_1). State the fitted regression function.

```
fit1 <- lm(Y~X1)
summary(fit1)
```

```
##
## Call:
## lm(formula = Y ~ X1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.475 -4.688 -0.100  4.638  7.525
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   50.775      4.395  11.554 1.52e-08 ***
## X1             4.425      0.598   7.399 3.36e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.349 on 14 degrees of freedom
## Multiple R-squared:  0.7964, Adjusted R-squared:  0.7818
## F-statistic: 54.75 on 1 and 14 DF,  p-value: 3.356e-06
```

$$Y = 4.425X_1 + 50.775$$

b. Compare the estimated regression coefficient for moisture content obtained in part (a) with the corresponding coefficient obtained in Problem 6.5b. What do you find?

```r
fit$coefficients
```

```
## (Intercept)          X1          X2
##      37.650       4.425       4.375
```

```r
fit1$coefficients
```

```
## (Intercept)          X1
##      50.775       4.425
```

They are the same.

c. Does $SSR(X_1)$ equal $SSR(X_1|X_2)$ here? If not, is the difference substantial?

```r
fit1 <- lm(Y~X1)
fit2 <- lm(Y~X2)
tab <- as.table(cbind(
  'SS' = c(
         "SSE(X1)"           = sum(fit1$residuals^2),
         "SSE(X2)"           = sum(fit2$residuals^2),
         "SSE(X1,X2)"        = fit.aov[3, 2],
         "SSR(X1)"           = fit.aov[1, 2],
         "SSR(X1|X2)"        = sum(fit2$residuals^2)-fit.aov[3, 2],
         "SSR(X2|X1)"        = sum(fit1$residuals^2)-fit.aov[3, 2],
         "Total"             = sum(fit.aov[, 2]))
))

round(tab, 2)
```

```
##                   SS
## SSE(X1)       400.55
## SSE(X2)      1660.75
## SSE(X1,X2)     94.30
## SSR(X1)      1566.45
## SSR(X1|X2)   1566.45
## SSR(X2|X1)    306.25
## Total        1967.00
```

$$SSR(X_1) = SSR(X_1|X_2) = 1566.45$$

d. Refer to the correlation matrix obtained in Problem 6.5a. What bearing does this have on your findings in parts (b) and (c)?

Since $r_{12} = 0$, X_2 doesn't have influence on X_1's coefficient and $SSR(X_1|X_2) = SSR(X_1)$.

7.30. Refer to Brand preference Problem 6.5.

a. Regress Y on X_2 using simple linear regression model (2.1) and obtain the residuals.

```r
fit2$residuals
```

```
##       1       2       3       4       5       6       7       8       9
## -13.375 -13.125 -16.375 -10.125  -5.375  -6.125  -6.375  -3.125   5.625
##      10      11      12      13      14      15      16
##   2.875   8.625   6.875  10.625   8.875  16.625  13.875
```

b. Regress X_1 on X_2 using simple linear regression model (2.1) and obtain the residuals.

```
fit12 <- lm(X1~X2)
fit2$residuals
```

```
##       1       2       3       4       5       6       7       8       9
## -13.375 -13.125 -16.375 -10.125  -5.375  -6.125  -6.375  -3.125   5.625
##      10      11      12      13      14      15      16
##   2.875   8.625   6.875  10.625   8.875  16.625  13.875
```

c. Calculate the coefficient of simple correlation between the two sets of residuals and show that it equals $r_{Y1|2}$.

```
cor(fit2$residuals,fit12$residuals)
```

```
## [1] 0.9711943
```

```
rY12 <- sign(fit1$coefficients[2])*sqrt((sum(fit2$residuals^2)-fit.aov[3, 2])/sum(fit2$residuals^2))
rY12
```

```
##        X1
## 0.9711943
```

7.31. The following regression model is being considered in a water resources study:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \beta_4 \sqrt{X_{i3}} + \epsilon_i$$

State the reduced models for testing whether or not:

(1) $\beta_3 = \beta_4 = 0$,

Reduce model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$$

Given α,

$$H_0 : \beta_3 = \beta_4 = 0 \qquad H_a : \beta_3 \neq 0 \text{ or } \beta_4 \neq 0$$

$$F^* = \frac{\frac{SSE_F - SSE_R}{2}}{\frac{SSE_F}{n-5}} \overset{H_0}{\sim} F(2, n-5)$$

The decision rule is

If $F^* \leqslant F(0.99, 2, n-5)$, then conclude H_0;

If $F^* > F(0.99, 2, n-5)$, then conclude H_a;

(2) $\beta_3 = 0$,

Reduce model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i$$

Given α,

$$H_0 : \beta_3 = 0 \qquad H_a : \beta_3 \neq 0$$

$$F^* = \frac{\frac{SSE_F - SSE_R}{1}}{\frac{SSE_F}{n-5}} \overset{H_0}{\sim} F(1, n-5)$$

The decision rule is

If $F^* \leqslant F(1-\alpha, 1, n-5)$, then conclude H_0;

If $F^* > F(1-\alpha, 1, n-5)$, then conclude H_a;

(3) $\beta_1 = \beta_2 = 5$**,**

Reduce model:

$$Y^{(3)} = Y_i - 5X_{i1} - 5X_{i2} = \beta_0 + \beta_3 X_{i1}X_{i2} + \beta_4\sqrt{X_{i3}} + \epsilon_i$$

Given α,

$$H_0 : \beta_1 = \beta_2 = 5 \qquad H_a : \beta_1 \neq 5 \text{ or } \beta_2 \neq 5$$

$$F^* = \frac{\frac{SSE_F - SSE_R}{2}}{\frac{SSE_F}{n-5}} \overset{H_0}{\sim} F(2, n-5)$$

The decision rule is

If $F^* \leqslant F(1-\alpha, 2, n-5)$, then conclude H_0;

If $F^* > F(1-\alpha, 2, n-5)$, then conclude H_a;

(4) $\beta_4 = 7$**.**

Reduce model:

$$Y^{(4)} = Y_i - 7\sqrt{X_{i3}} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}X_{i2} + \epsilon_i$$

Given α,

$$H_0 : \beta_4 = 7 \qquad H_a : \beta_4 \neq 7$$

$$F^* = \frac{\frac{SSE_F - SSE_R}{1}}{\frac{SSE_F}{n-5}} \overset{H_0}{\sim} F(1, n-5)$$

The decision rule is

If $F^* \leqslant F(1-\alpha, 1, n-5)$, then conclude H_0;

If $F^* > F(1-\alpha, 1, n-5)$, then conclude H_a;