Tut3_ST5202

Name:Zhu Xu
User ID:E0337988
Student ID:A0131944H

26/Mar/2019

Q1:

a)

$$\begin{aligned} SSR(x_1, x_2, \cdots, x_{p-1}) &= SSR(x_1) + SSR(x_2|x_1) + SSR(x_3|x_1, x_2) + \cdots + SSR(x_{p-1}|x_1, \cdots, x_{p-2}) \\ &= SSR(x_1, x_2) + SSR(x_3|x_1, x_2) + \cdots + SSR(X_{p-1}|x_1, \cdots, x_{p-2}) \\ &= SSR(x_1, x_2, x_3) + SSR(x_4|x_1, x_2, x_3) + \cdots + SSR(x_{p-1}|x_1, \cdots, x_{p-2}) \\ &= \cdots \\ &= SSR(x_1, x_2, \cdots, x_{p-1}) \end{aligned}$$

b)

Similarly,

$$\begin{aligned} SSR &= SSR(x_1, x_2, \cdots, x_{p-1}) \\ &= SSR(x_1, x_2, x_3) + SSR(x_4|x_1, x_2, x_3) + \cdots + SSR(x_{p-1}|x_1, \cdots, x_{p-2}) \\ &= \cdots \\ &= SSR(x_p) + SSR(x_{p-1}|x_p) + SSR(x_{p-2}|x_{p-1}, x_p) + \cdots + SSR(x_1|x_p, \cdots, x_2) \end{aligned}$$

5.3

$$Y = X\beta + \varepsilon \rightarrow \quad \varepsilon = Y - X\beta$$

Then minimize,

$$\varepsilon^T\varepsilon = (Y - X\beta)^T(Y - X\beta)$$

$$2X^TX\beta - 2X^TY = 0$$

$$\hat{\beta} = (X^TX)^{-1}XY$$

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} - \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \hat{Y}_3 \\ \hat{Y}_4 \end{bmatrix}, \qquad X_{ij} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}_{p\times 1}$$

5.17

a)

$$\begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

b)

$$E(W) = \begin{bmatrix} E(Y_1) + E(Y_2) + E(Y_3) \\ E(Y_1) - E(Y_2) \\ E(Y_1) - E(Y_2) - E(Y_3) \end{bmatrix}$$

c)

$$cov(W) = \begin{bmatrix} Var(Y_1) + Var(Y_2) + Var(Y_3) & Var(Y_1) - Var(Y_2) & Var(Y_1) - Var(Y_2) - Var(Y_3) \\ Var(Y_1) - Var(Y_2) & Var(Y_1) + Var(Y_2) & Var(Y_1) + Var(Y_2) \\ Var(Y_1) - Var(Y_2) - Var(Y_3) & Var(Y_1) + Var(Y_2) & Var(Y_1) + Var(Y_2) + Var(Y_3) \end{bmatrix}$$

6.22

a)

It's a general linear regression model.

b)

No, $logY_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}^2 + \varepsilon$

c)

No, $Y_i = log\beta_1 + logX_{i1} + \beta_2 X_{i2} + \varepsilon$

d)

No, it has no transformation

e)

No, $log(\frac{1}{Y_i} - 1) = \beta_0 + \beta_1 X_{i1}\varepsilon$

6.10

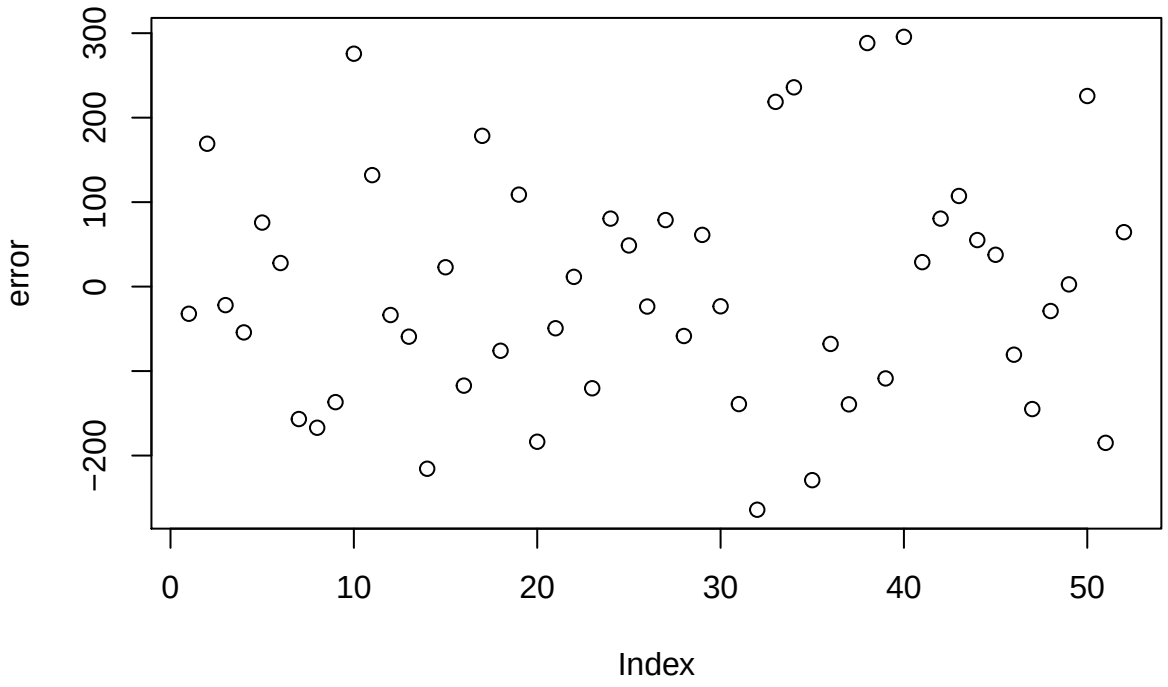
a)

```
data_1 = read.csv(
  '/Users/xuzhu/Desktop/Notes/Sem2/ST5202-Applied_Regression_Analysis/Tut/grocery_retailer.txt',
  sep='', header=F)
y = data_1[,1]
x1 = data_1[,2]
x2 = data_1[,3]
x3 = data_1[,4]
fit = lm(y~x1+x2+x3, data=data_1)
# fit = lm(v1~v2+v3+v4, data=data_1)
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3, data = data_1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -264.05 -110.73  -22.52   79.29  295.75
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.150e+03  1.956e+02  21.220  < 2e-16 ***
## x1           7.871e-04  3.646e-04   2.159   0.0359 *
## x2          -1.317e+01  2.309e+01  -0.570   0.5712
## x3           6.236e+02  6.264e+01   9.954 2.94e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 143.3 on 48 degrees of freedom
## Multiple R-squared:  0.6883, Adjusted R-squared:  0.6689
## F-statistic: 35.34 on 3 and 48 DF,  p-value: 3.316e-12
```

b)

```
error = y - 4.150e+03 - 7.871e-04*x1 + 1.317e+01*x2 - 6.236e+02*x3
plot(error)
```



6.12

```
new_x1 = c(302000, 245000, 280000, 350000, 295000)
new_x2 = c(7.20, 7.40, 6.90, 7.00, 6.70)
new_x3 = c(0, 0, 0, 0, 1)
new_data_1 = data.frame(new_x1, new_x2, new_x3)
predict_result = predict(fit, newdata=new_data_1, interval='confidence')
```

```
## Warning: 'newdata' had 5 rows but variables found have 52 rows
```

```
predict_result = data.frame(predict_result)

tvalue = qt(1-0.05/2, 50-4)
tvalue
```

```
## [1] 2.012896
```

```
bond = qt(1-0.05/(2*5), 50-4)
bond
```

```
## [1] 2.687013
```

```
width = sqrt(2*qf(1-0.05, 4, 50-4))
width
```

```
## [1] 2.268936
```

```
calfun_lwr = function(a){
  v = (predict_result$lwr - predict_result$fit)/tvalue * a + predict_result$fit
  return(v)
}

calfun_upr = function(b){
  v = (predict_result$lwr - predict_result$fit)/tvalue * b + predict_result$fit
  return(v)
}

bond_lwr = calfun_lwr(bond)
bond_upr = calfun_upr(bond)
width_lwr = calfun_lwr(width)
width_upr = calfun_upr(width)

CI = data.frame(bond_lwr, bond_upr, width_lwr, width_upr)
summary(CI)
```

```
##     bond_lwr       bond_upr      width_lwr      width_upr
##  Min.   :4125   Min.   :4125   Min.   :4142   Min.   :4142
##  1st Qu.:4186   1st Qu.:4186   1st Qu.:4198   1st Qu.:4198
##  Median :4215   Median :4215   Median :4225   Median :4225
##  Mean   :4265   Mean   :4265   Mean   :4280   Mean   :4280
##  3rd Qu.:4246   3rd Qu.:4246   3rd Qu.:4258   3rd Qu.:4258
##  Max.   :4818   Max.   :4818   Max.   :4849   Max.   :4849
```

7.4

a)

```
new_fit = lm(y~x1+x3+x2, data=data_1)
anova(new_fit)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## x1         1  136366  136366  6.6417   0.01309 *
## x3         1 2033565 2033565 99.0443 2.963e-13 ***
## x2         1    6675    6675  0.3251   0.57123
## Residuals 48  985530   20532
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

b)

$$F = \frac{SSR(X2|X1, X3)}{\frac{SSR(X1)+SSR(X3|X1)}{50-3}} = 0.1445783$$

```
F = 0.1445
if(F < qf(1-0.05,1, 47)){
    print("X2 can be dropped")
}
```

```
## [1] "X2 can be dropped"
```

c)

They are equal.

7.25

a)

```
new_fit_1 = lm(y~x1, data=data_1)
anova(new_fit_1)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq F value Pr(>F)
## x1         1  136366  136366  2.2534 0.1396
## Residuals 50 3025770   60515
```

b)

The original b_1 is 7.871×10^{-4}, the new one is 9.355×10^{-4}

c)

```
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## x1         1  136366  136366  6.6417   0.01309 *
## x2         1    5726    5726  0.2789   0.59987
## x3         1 2034514 2034514 99.0905 2.941e-13 ***
## Residuals 48  985530   20532
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
new_fit_2 = lm(y~x1+x3+x2, data=data_1)
anova(new_fit_2)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## x1         1  136366  136366  6.6417   0.01309 *
## x3         1 2033565 2033565 99.0443 2.963e-13 ***
## x2         1    6675    6675  0.3251   0.57123
## Residuals 48  985530   20532
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```