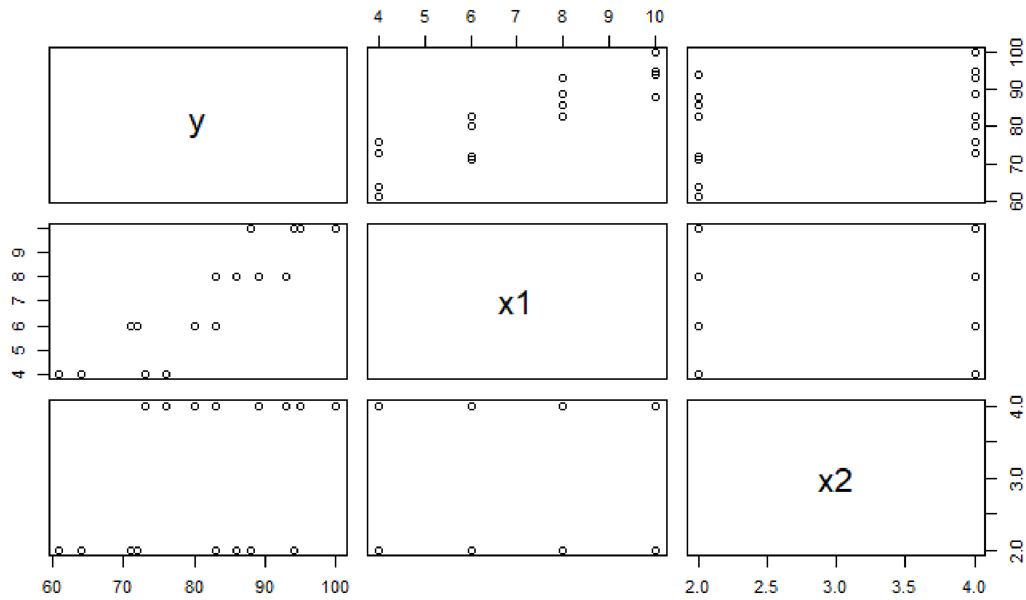
Homework 4 (50 pts)
due 10/12

1.(15) In a small scale experimental study of the relation between degree of brand liking (Y) and moisture content (X1) and sweetness (X2) of the product, data is in brand.csv. Sample size is 16. Use R to

R code is shown in Appendix.

a). (2) draw a scatter plot and the correlation matrix, describe what you see.

Figure 1.



The relationship between y and x1 looks linear and highly correlated, but the relationship between y and x2 is uncorrelated. Plus, x1 and x2 are not correlated.

Figure 2.

```
      y   x1   x2
y  1.00 0.89 0.39
x1 0.89 1.00 0.00
x2 0.39 0.00 1.00

n= 16


P
   y      x1     x2
y         0.0000 0.1304
x1 0.0000        1.0000
x2 0.1304 1.0000
```

The correlation between y and x1 is 0.89, which means they are highly correlated, while the correlation between y and x2 is 0.39, which means they are not very correlated. Plus, the correlation between x1 and x2 is 0, which means they are uncorrelated. The conclusion is consistent with that of the Figure 1.

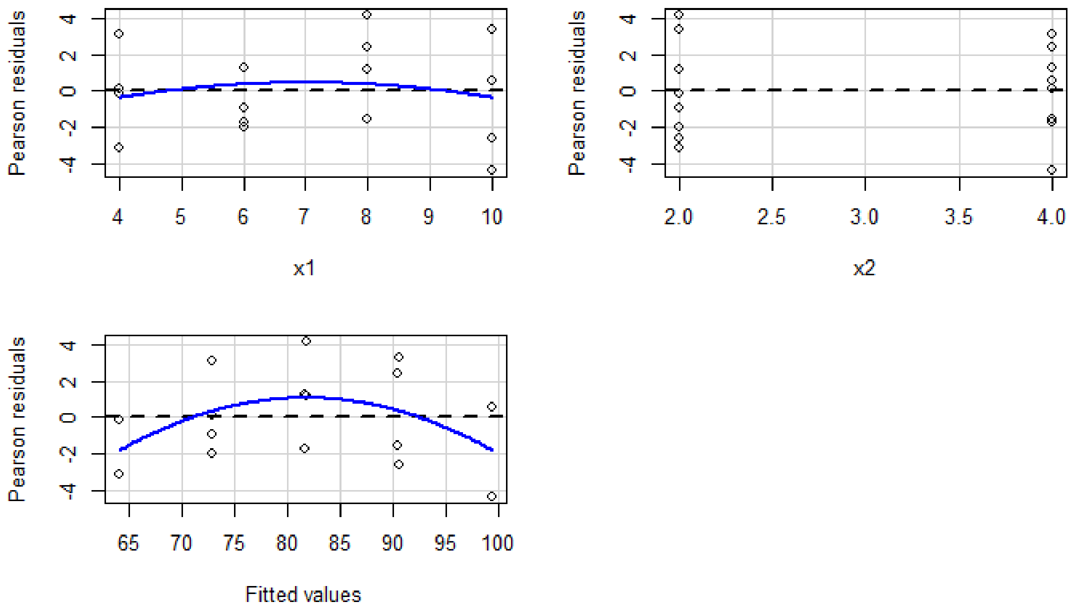
b). (1) fit regression model to the data without interaction, $\widehat{Y} = \beta_1 X_1 + \beta_2 X2$
Figure 3.

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

Residuals:
   Min     1Q Median     3Q    Max 
-4.400 -1.762  0.025  1.587  4.200 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  37.6500     2.9961  12.566 1.20e-08 ***
x1            4.4250     0.3011  14.695 1.78e-09 ***
x2            4.3750     0.6733   6.498 2.01e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.693 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
```

`

$\widehat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 = 37.65 + 4.425X_1 + 4.375X_2$

c). (2) Perform a test to see if the residuals are Normal.
Figure 4.



From Figure 4, the distribution of residuals vs. X1 has a random pattern around 0 line, which means the relationship is linear is reasonable and the residuals have constant variances. There no unusual points. However, the distribution of residuals vs. x2 has a random pattern, which means the error terms are non-normal and have non-constant variances. Moreover, the distribution of residuals vs fitted value has a certain pattern, which means residuals may not be normally distributed.

H_0: Data follows normal distribution

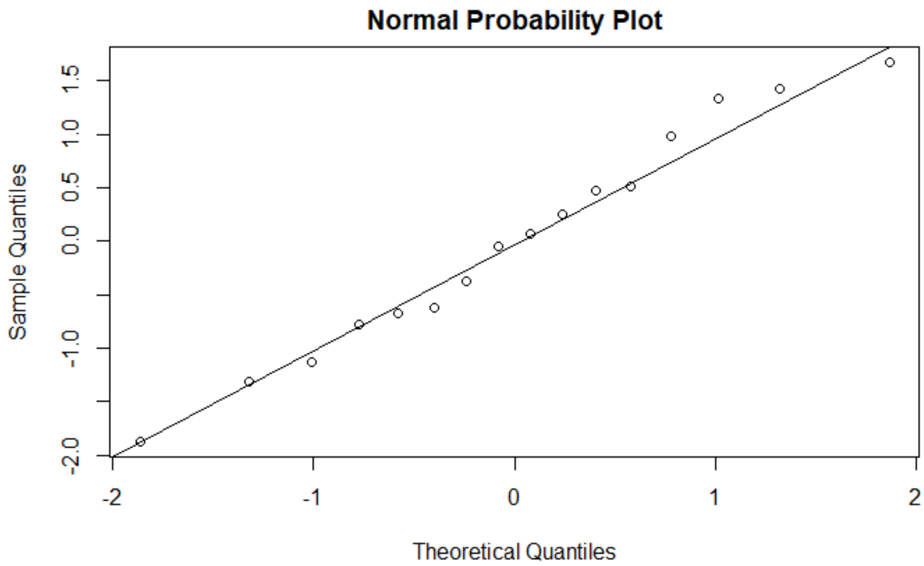
H_a: Data violates normal distribution

Refer to Figure 5., P-value = 0.9111> significance level $\alpha = 0.05$, and it is much bigger than the significance level, so the distribution of the residuals is a normal distribution. Also, from the Normal probability plot, the residual points fit the line well, and they have a normal distribution.

Figure 5.

```
           Test stat Pr(>|Test stat|)
x1           -0.5416          0.59801
x2            0.1994          0.84527
Tukey test   -1.7221          0.08506 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

        Shapiro-Wilk normality test

data:  concentration.stdres
W = 0.97496, p-value = 0.9111
```

Figure 6.



d). (2) Perform BF test for constancy of the residuals. You can make two groups (Y ≤81.75, or >81.75, where 81.75 is the average).

H_0: Residuals have constant variances

H_a: Residuals have non-constant variances

Refer to Figure 7., P-value = 0.4027057 > significance level $\alpha = 0.05$, so we do not reject H_0. Hence, we are 95% confident that the residuals have constant variances.

Figure 7.

```
        t.value   P.Value alpha df
[1,] 0.8629512 0.4027057  0.05 14
```

e). (4) Perform a lack of fit test of the model use a significant level of 0.01. Sate Ho/Ha, test statistic, critical value, p value and conclusion.

Figure 8.

```
Analysis of Variance Table

Model 1: y ~ x1 + x2
Model 2: y ~ factor(x1) * factor(x2)
  Res.Df  RSS Df Sum of Sq     F Pr(>F)
1     13 94.3
2      8 57.0  5      37.3 1.047  0.453
```

$H_0: E\{Y\} = \beta_0 + \beta_1 X + \beta_2 X_2$

$H_a: E\{Y\} \neq \beta_0 + \beta_1 X + \beta_2 X_2$

From Figure 8,
$F^* = 1.047$

From data, there are 8 pairs of independent variables and 3 parameters, so n=8, p=3, and there are 16 data, so N =16.

Critical value: $F(1-\alpha; n-p, N-n) = F(0.99; 5,8) = 6.63$

Since $F^* = 1.047 < F(1-\alpha; n-p, N-n) = F(0.99; 5,8) = 6.63$, do not reject H_0

P-value = 0.453 > $\alpha = 0.01$, so we do not reject H_0, so there is no lack of fit. Hence the model fits the data well.

f). (4) Find MSE, the variance-covariance matrix of estimators (i.e., $\Sigma\{\mathbf{b}\}$), variance-covariance matrix of predictors (i.e, $\Sigma\{\hat{Y}_h\}$) when ($X1 = 5, X2 = 4$)

Figure 9.

```
> MSE
         week
week 7.253846
```

From Figure 9, MSE = 7.253846.

Figure 10.

```
> sigmaBeta
           Intercept         temp1      temp2
Intercept  8.9766346 -6.347115e-01 -1.3600962
temp1     -0.6347115  9.067308e-02  0.0000000
temp2     -1.3600962  1.887513e-16  0.4533654
```

From Figure 10, the variance-covariance matrix of estimators is above.

$$\sum b = \begin{pmatrix} 8.9766346 & -0.6347115 & -1.3600962 \\ -0.6347115 & -0.09067308 & 0 \\ -1.3600962 & 1.887513 * 10^{-16} & 0.4533654 \end{pmatrix}$$

Figure 11.

```
> varYhat
         [,1]
[1,] 1.269423
```

From Figure 11, variance-covariance matrix of predictors is above.

2. (10) Refer to question 1, compute the following question by hand.

a). (2) Obtain an interval estimate of $E\{Y_h\}$ (i.e., $\hat{Y}_h$) when ($X1 = 5, X2 = 4$), with 99% confidence level.

b). (2) Obtain an interval estimate of a single predictor $\hat{Y}_h\{new\}$ when ($X1 = 5, X2 = 4$), with 99% confidence level.

c). (2) Obtain an interval estimate of the average of the next two predictors, when ($X1 = 5, X2 = 4$), with 90% confidence level.

d). (2) Obtain a simultaneous estimate of the two (single) predictor $\hat{Y}_h\{new\}$ when ($X1 = 5, X2 = 4$), and ($X1 = 6, X2 = 5$), with a 90% confidence level.

e). (2) Obtain a simultaneous confidence interval for all three estimators β_0, β_1 and β_2, with a 90% confidence level.

Q2

(a) From Q1 (f). $MSE = 7.253846$, $\Sigma\{\hat{Y}_h\} = s^2\{\hat{Y}_h\} = (1.269423)$

Now, $X_1 = 5$, $X_2 = 4$, so $X_h = \begin{pmatrix} 1 \\ 5 \\ 4 \end{pmatrix}$, From Q1(b), $b = \begin{pmatrix} 37.65 \\ 4.425 \\ 4.375 \end{pmatrix}$

$\hat{Y}_h = X_h^T b = (1 \quad 5 \quad 4) \begin{pmatrix} 37.65 \\ 4.425 \\ 4.375 \end{pmatrix} = 77.275$

$n=16$, $p=3$

$t(1-\frac{\alpha}{2}; n-p) = t(0.995; 13) = 3.012$

$s\{Y_h\} = \sqrt{s^2\{Y_h\}} = \sqrt{1.269423} = 1.1266867$

The 99% confidence limits for $E\{Y_h\}$ are

$\hat{Y}_h \pm t(1-\frac{\alpha}{2}; n-p) s\{\hat{Y}_h\}$

$= 77.275 \pm 3.012 \times 1.1266867$

$= (73.88142, 80.66858)$

(b) $s^2\{pred\} = MSE + s^2\{\hat{Y}_h\} = 7.253846 + 1.269423 = 8.523269$

$s\{pred\} = \sqrt{s^2\{pred\}} = 2.9194638$

$\hat{Y}_h = 77.275$, $t(1-\frac{\alpha}{2}; n-p) = 3.012$

The 99% confidence limit for one new observation $Y_{h(new)}$ according to X_h are

$= \hat{Y}_h \pm t(1-\frac{\alpha}{2}; n-p) s\{pred\} = 77.275 \pm 3.012 \times 2.9194638$

$= (68.48158, 86.068425)$

(c) $s^2\{predmean\} = \frac{MSE}{m} + s^2\{Y_h\} = \frac{7.253846}{2} + 1.269423 = 4.896346$

$s\{predmean\} = \sqrt{s^2\{predmean\}} = \sqrt{4.896346} = 2.2127689$

$\hat{Y}_h = 77.275$, $t(1-\frac{\alpha}{2}; n-p) = t(0.95; 13) = 1.771$

The 90% confidence limit for mean of 2 new observations at X_h are

$= \hat{Y}_h \pm t(1-\frac{\alpha}{2}; n-p) s\{predmean\}$

$= 77.275 \pm 1.771 \times 2.2127689$

$= (73.356186, 81.193814)$

(d) From Q1 (f), $S^2\{b\} = \begin{pmatrix} 8.9766346 & -0.6347115 & -1.3600962 \\ -0.6347115 & 0.09067308 & 0 \\ -1.3600962 & 1.887513 \times 10^{-6} & 0.4533654 \end{pmatrix}$

① $S^2\{pred\} = 8.523269$ when $(X_1, X_2) = (5, 4)$ from part (b).

② $S^2\{pred\} = MSE + S^2\{\hat{Y}_h\}$ when $(X_1, X_2) = (6, 5)$

$= MSE + X_h' S^2\{b\} X_h$

$= 7.253846 + (1\ 6\ 5)\ S^2\{b\} \begin{pmatrix} 1 \\ 6 \\ 5 \end{pmatrix}$

$= 7.253846 + 2.375$

$= 9.628846$

$B = t(1 - \frac{\alpha/2}{g};\ n-2) = t(1 - \frac{0.05}{2};\ 14) = t(0.975, 14) = 2.145$

① $s\{pred\} = \sqrt{8.523269} = 2.9194638$ when $(X_1, X_2) = (5, 4)$

② $s\{pred\} = \sqrt{9.628846} = 3.1030382$ when $(X_1, X_2) = (6, 5)$

$\hat{Y}_h = 77.275$ when $(X_1, X_2) = (5, 4)$.

$\hat{Y}_h = X_h^T b = (1\ \ 6\ \ 5) \begin{pmatrix} 37.65 \\ 4.425 \\ 4.375 \end{pmatrix} = 86.075$

The simultaneous CI:

$\hat{Y}_h \pm B\, s\{pred\} = 77.275 \pm 2.145 \times 2.9194638$ when $(X_1, X_2) = (5, 4)$

$= (71.012750,\ 83.537250)$

$\hat{Y}_h \pm B\, s\{pred\} = 86.075 \pm 2.145 \times 3.1030382$ when $(X_1, X_2) = (6, 5)$

$= (79.418983,\ 92.731017)$

(e) From $S^2\{b\}$, we can see that

$S^2\{b_0\} = 8.9766346$, $S^2\{b_1\} = 0.09067308$; $S^2\{b_2\} = 0.4533654$

$g = 3$, $B = t(1 - \frac{\alpha/2}{3};\ n-2) = t(0.983;\ 14) = 2.36$ by calculator.

$s\{b_0\} = 2.996103$, $s\{b_1\} = 0.3011197$, $s\{b_2\} = 0.673324$

For β_0: $b_0 \pm B\, S\{b_0\} = 37.65 \pm 2.36 \times 2.996103$

$= (30.579197,\ 44.720803)$

For β_1: $b_1 \pm B\, S\{b_1\} = 4.425 \pm 2.36 \times 0.3011197$

$= (3.714358,\ 5.135642)$

For β_2: $b_2 \pm B\, S\{b_2\} = 4.375 \pm 2.36 \times 0.673324$

$= (2.785955,\ 5.964045)$

3. (7) Refer to question 1,

R code is shown in Appendix.

a). (1) What is the ANOVA table that decomposes the regression sum of squares into extra sums of squares associated with **X2, then with X1, given X2. (**You may use R for this question)

Figure 1. "the regression sum of squares into extra sums of squares associated with **X2"**

```
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value Pr(>F)
x2         1  306.25  306.25  2.5817 0.1304
Residuals 14 1660.75  118.62
```

Figure 2. "the regression sum of squares into extra sums of squares associated with **X1, given X2"**

```
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
x2         1  306.25  306.25  42.219 2.011e-05 ***
x1         1 1566.45 1566.45 215.947 1.778e-09 ***
Residuals 13   94.30    7.25
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

b). (3) Test whether X1 can be dropped from the regression model given X2 is retained. Use the partial F test with a significant level of 0.01. Define Ho/Ha, test statistic, critical value, and state conclusion.

$H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0$

From Figure 1 & Figure 2, SSE(R) = 1660.75, $df_R = n - (p - 1) = 16 - (3 - 1) = 14$

SSE(F) = 94.3, $df_F = n - p = 16 - 3 = 13$

$$F^* = \frac{\frac{SSE(R) - SSE(F)}{df_R - df_F}}{\frac{SSE(F)}{df_F}} = \frac{\frac{1660.75 - 94.3}{14 - 13}}{\frac{94.3}{13}} = 215.947508$$

Critical value: $F(1 - \alpha; df_R - df_F, df_F) = F(0.99; 1,13) = 9.33$

Since $F^* = 215.947508 > F(1 - \alpha; df_R - df_F, df_F) = F(0.99; 1,13) = 9.33$, reject H_0

To estimate P-value,

$F^* > F(0.999,1,13) = 18.64$, so P-value < 0.001< $\alpha = 0.01$, so reject H_0.

Also, from Figure 2, we can see that $F^* = 215.947$ and P-value < 0.0001, so reject H_0, so $\beta_1 \neq 0$, X1 cannot be dropped from the regression Model.

c). (3) Compute R^2_{Y1},$R^2_{Y\,1\,|2}$, $R^2_{Y\,2|1}$ *and* R^2. Explain what each coefficient measures and interpret your result.

Figure 3.

```
> anova(lm(y~x1,data))
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)    
x1         1 1566.45 1566.45  54.751 3.356e-06 ***
Residuals 14  400.55   28.61                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

$$R^2_{Y\,1} = \frac{SSR(X1)}{SSR(X1) + SSE} = \frac{1566.45}{1566.45 + 400.55} = 0.796365$$

Figure 4.

```
> anova(lm(y~x2+x1,data))
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)    
x2         1  306.25  306.25  42.219 2.011e-05 ***
x1         1 1566.45 1566.45 215.947 1.778e-09 ***
Residuals 13   94.30    7.25                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

$$R^2_{Y\,1|2} = \frac{SSR(X1|X2)}{SSR(X1|X2) + SSE} = \frac{1566.45}{1566.45 + 94.3} = 0.943218$$

Figure 5.

```
> anova(lm(y~x1+x2,data))
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)    
x1         1 1566.45 1566.45 215.947 1.778e-09 ***
x2         1  306.25  306.25  42.219 2.011e-05 ***
Residuals 13   94.30    7.25                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

mn

$$R^2_{Y\,2|1} = \frac{SSR(X2|X1)}{SSR(X2|X1) + SSE} = \frac{306.25}{306.25 + 94.3} = 0.764574$$

From Figure 4,

Type I SS: $R^2 = \frac{SS1}{SS1+SSE} = \frac{306.25+1566.45}{306.25+1566.45+94.3} = 0.9521$

Type II SS: $R^2 = \frac{SS2}{SS2+SSE} = \frac{1566.45+306.25}{1566.45+306.25+94.3} = 0.9521$

4. (6) A commercial real estate company evaluate vacancy rates, square footage, rental rates, and operating expenses for commercial properties in a large metropolitan area in order to provide clients with quantitative information upon which to make rental decisions. N=81 suburban commercial properties are evaluated.

Y: rental sales
X1: age
X2: operating expense
X3: vacancy rates
X4: total square footage

According to the following ANOVA table, perform the following test, use a significant level of 0.01. Define Ho/Ha, test statistic, critical value, and state conclusion.

```
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq
x4         1 67.775  67.775
x1         1 42.275  42.275
x2         1 27.857  27.857
x3         1  0.420   0.420
Residuals 76 98.231   1.293
```

a). (2) whether X3 can be dropped from the regression model given that X1, X2 and X4 are retained.

b). (3) whether X2 and X3 can be dropped from the regression model given that X1 and X4 are retained.

c). (1) compute $R^2_{Y\ 3|1,2,4}$

Q4 $N=81,\ \alpha=0.01$

(a) $H_0: \beta_3 = 0$ (Reduced Model) $\quad H_a: \beta_3 \neq 0$ (Full Model)

$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_4 X_{i,4} + \varepsilon_i \quad Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \beta_4 X_{i,4} + \varepsilon_i$

$df_R = n-(p-1) = 81-(5-1) = 77 \quad df_F = n-p = 81-5 = 76$

$$F_S = \frac{\frac{SSE(R)-SSE(F)}{df_R - df_F}}{\frac{SSE(F)}{df_F}} = \frac{\frac{SSR(X_3 \mid X_1, X_2, X_4)}{(n-p+1)-(n-p)}}{\frac{SSE(X_1,X_2,X_3,X_4)}{n-p}} = \frac{\frac{0.42}{1}}{\frac{98.231}{76}} = 0.324948$$

$F(1-\alpha;\ df_R - df_F,\ df_F) = F(0.99;\ 1, 76) = 7.17$

$F_S = 0.324948 < 7.17$

so Do not reject $H_0: \beta_3 = 0$.

We're 99% confident that X_3 can be dropped from the regression Model.

(b) $H_0: \beta_2 = \beta_3 = 0$ (Reduced Model) $\quad H_a$: Not both β_2 and β_3 equal 0. (Full Model)

$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_4 X_{i,4} + \varepsilon_i \quad Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \beta_4 X_{i,4} + \varepsilon_i$

$df_R = n-(p-2) = 81-(5-2) = 78 \quad df_F = n-p = 81-5 = 76$

$$F_S = \frac{\frac{SSE(R)-SSE(F)}{df_R - df_F}}{\frac{SSE(F)}{df_F}} = \frac{\frac{SSR(X_2, X_3 \mid X_1, X_4)}{(n-p+2)-(n-p)}}{\frac{SSE(X_1,X_2,X_3,X_4)}{n-p}} = \frac{\frac{27.857+0.42}{2}}{\frac{98.231}{76}} = 10.9387668$$

$F(1-\alpha;\ df_R - df_F,\ df_F) = F(0.99;\ 2, 76) = 5.06$

$F_S = 10.9387668 > 5.06$

so reject $H_0: \beta_2 = \beta_3 = 0$.

We're 99% confident that X_2 and X_3 have significant relationship with Y, and they cannot be dropped.

(c) $$R^2_{Y3 \mid 1,2,4} = \frac{SSR(X_3 \mid X_1, X_2, X_4)}{SSR(X_3 \mid X_1, X_2, X_4) + SSE} = \frac{0.42}{0.42 + 98.231} = 0.00425743$$

5. (12) In a study of insurance industry, an economist wished to related the speed with which a particular insurance innovation is adopted (Y) to the size of the insurance firm (X1) and the type of firm (X2, stock company and mutual company). Data is in insurance.csv

$$X_2 = \begin{cases} 0 & if\ mutual\ company \\ 1 & if\ stock\ company \end{cases}$$

Preform hypothesis test for the following question. Use a significant level of 0.1. Define Ho/Ha, test statistic, critical value, and state conclusion.

R code is shown in Appendix.

a). (3) The mutual firm and the stock firm have the same average adopt time for any firm size.

$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon_i$, where $\beta_3 \neq 0$

$H_0: \beta_2 = 0, \beta_3 = 0$

$H_a: \beta_2 \neq 0, \beta_3 \neq 0\ or \beta_2 \neq 0\ and\ \beta_3 \neq 0$

Full Model: $\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$

Reduced Model: $\hat{Y} = \beta_0 + \beta_1 X_1$

Figure 1. Full Model

```
> anova(fit1)
Analysis of Variance Table

Response: months
          Df  Sum Sq Mean Sq  F value    Pr(>F)
size       1 1188.17 1188.17 107.7819 1.627e-08 ***
type       1  316.25  316.25  28.6875 6.430e-05 ***
size:type  1    0.01    0.01   0.0005    0.9821
Residuals 16  176.38   11.02
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 2. Reduced Model

```
> anova(fit2)
Analysis of Variance Table

Response: months
          Df  Sum Sq Mean Sq F value    Pr(>F)
size       1 1188.17 1188.17  43.414 3.452e-06 ***
Residuals 18  492.63   27.37
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

From Figure 1 & Figure 2, SSE(R) = 492.63, $df_R = n - (p-2) = 20 - (4-2) = 18$

SSE(F) = 176.38, $df_F = n - p = 20 - 4 = 16$

$$F^* = \frac{\frac{SSE(R) - SSE(F)}{df_R - df_F}}{\frac{SSE(F)}{df_F}} = \frac{\frac{SSR(X2, X3|X1)}{18-16}}{\frac{SSE(X1, X2, X3)}{20-4}} = \frac{\frac{316.25 + 0.01}{18-16}}{\frac{176.38}{16}} = 14.3445$$

Critical value: $F(1-\alpha; df_R - df_F, df_F) = F(0.9; 2{,}16) = 2.7$

Since $F^* = 14.34403 > F(1-\alpha; df_R - df_F, df_F) = F(0.9; 2{,}16) = 2.7$, reject H_0

we reject H_0: $\beta_2 = 0, \beta_3 = 0$, so at least of the β_2 or β_3 is not equal to zero, which means at least one of them has impact on Y, and they cannot be dropped. Hence, the mutual firm and the stock firm have different average adopt time for any firm size.

b). (3) The firm size (X1) has the same impact on the adopt time in mutual firm and stock firm.

$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon_i$, where $\beta_3 \neq 0$

H_0: $\beta_3 = 0$

H_a: $\beta_3 \neq 0$

Full Model:Ŷ=$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$

Reduced Model: Ŷ=$\beta_0 + \beta_1 X_1 + \beta_2 X_2$

Figure 3. Full Model

```
> summary(fit1)

Call:
lm(formula = months ~ size + type + size * type, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.7144 -1.7064 -0.4557  1.9311  6.3259 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 33.8383695  2.4406498  13.864 2.47e-10 ***
size        -0.1015306  0.0130525  -7.779 7.97e-07 ***
type         8.1312501  3.6540517   2.225   0.0408 *  
size:type   -0.0004171  0.0183312  -0.023   0.9821    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.32 on 16 degrees of freedom
Multiple R-squared:  0.8951,    Adjusted R-squared:  0.8754 
F-statistic: 45.49 on 3 and 16 DF,  p-value: 4.675e-08
```

From Figure 3,

the t_s for β_3 is -0.023

n=20, p=4

$t_c = t\left(1-\frac{\alpha}{2}; n-4\right) = t(0.95; 16) = 1.746$

the t_s for β_3 is -0.023 < 1.746,

since the t_s for β_3 is -0.023 < 1.746, we do not reject H_0: $\beta_3 = 0$, so β_3 may be equal to zero, which means β_3 has no impact on Y. Hence, the firm size (X1) has the same impact on the adopt time in mutual firm and stock firm.

c). (3) The firm size (X1) has the no impact on the adopt time in mutual firm and stock firm.

$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon_i$, where $\beta_3 \neq 0$

H_0: $\beta_1 = \beta_3 = 0$

H_a: $\beta_1 \neq 0$ or $\beta_3 \neq 0$

Full Model: $\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$

Reduced Model: $\hat{Y} = \beta_0 + \beta_2 X_2$

Figure 4. Reduced Model

```
> anova(fit3)
Analysis of Variance Table

Response: months
          Df Sum Sq Mean Sq F value Pr(>F)
type       1  145.8 145.800  1.7097 0.2075
Residuals 18 1535.0  85.278
```

From Figure 1 & Figure 3, SSE(R) = 1535.0, $df_R = n-(p-2) = 20-(4-2) = 18$

SSE(F) = 176.38, $df_F = n - p = 20 - 4 = 16$

$$F^* = \frac{\frac{SSE(R) - SSE(F)}{df_R - df_F}}{\frac{SSE(F)}{df_F}} = \frac{\frac{SSR(X1, X3|X2)}{18-16}}{\frac{SSE(X1, X2, X3)}{20-4}} = \frac{\frac{1188.17 + 0.01}{18-16}}{\frac{176.38}{16}} = 53.8918$$

Critical value: $F(1-\alpha; df_R - df_F, df_F) = F(0.9; 2{,}16) = 2.7$

Since $F^* = 61.6224 > F(1-\alpha; df_R - df_F, df_F) = F(0.9; 2{,}16) = 2.7$, reject H_0

we reject H_0: $\beta_1 = \beta_3 = 0$, so at least of the β_1 or β_3 is not equal to zero, which means at least one of them has impact on Y, and they cannot be dropped Hence, the firm size (X1) has impact on the adopt time in mutual firm and stock firm.

d). (3) If the firm size (X1) has the same impact on the two insurance company (i.e.$\beta_3 = 0$), the average adoption time for the stock firm, at any given firm size, is also the same as the mutual firm.

$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon_i$

H_0: $\beta_2 = 0$

H_a: $\beta_2 \neq 0$

Full Model: $\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

Reduced Model: $\hat{Y} = \beta_0 + \beta_1 X_1$

Figure 5. Full model

```
> summary(fit4)

Call:
lm(formula = months ~ size + type, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.6915 -1.7036 -0.4385  1.9210  6.3406 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 33.874069   1.813858  18.675 9.15e-13 ***
size        -0.101742   0.008891 -11.443 2.07e-09 ***
type         8.055469   1.459106   5.521 3.74e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.221 on 17 degrees of freedom
Multiple R-squared:  0.8951,	Adjusted R-squared:  0.8827 
F-statistic:  72.5 on 2 and 17 DF,  p-value: 4.765e-09

> anova(fit4)
Analysis of Variance Table

Response: months
          Df  Sum Sq Mean Sq F value    Pr(>F)    
size       1 1188.17 1188.17  114.51 5.683e-09 ***
type       1  316.25  316.25   30.48 3.742e-05 ***
Residuals 17  176.39   10.38                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

From Figure 5,

the $t_s\ for\ \beta_2$ is 5.521

n=20, p=3

$t_c = t\left(1 - \frac{\alpha}{2}; n - 3\right) = t(0.95; 17) = 1.74$

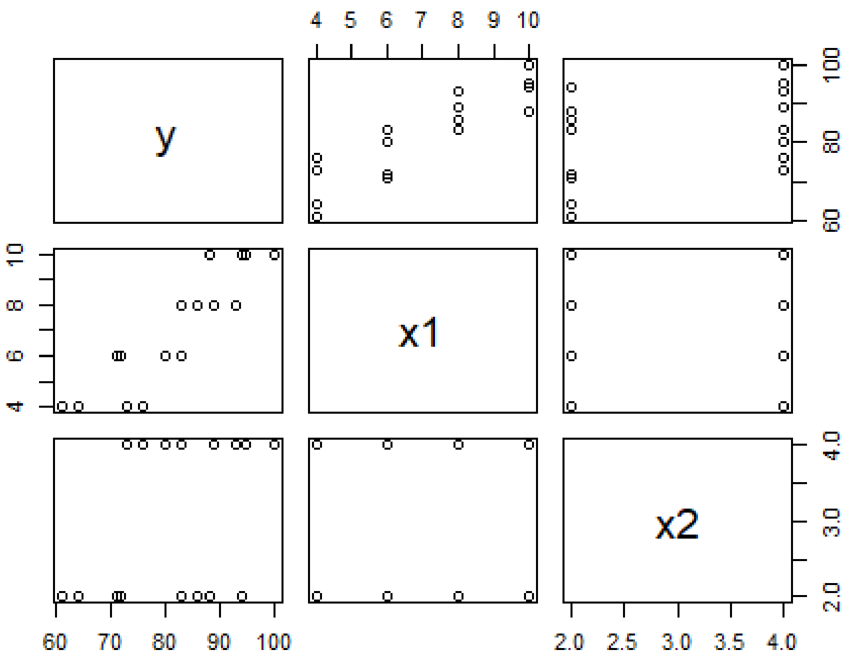
the $t_s\ for\ \beta_2$ is 5.521> 1.74,

since the $t_s\ for\ \beta_2$ is 5.521 > 1.74, we reject H_0: $\beta_2 = 0$, so β_2 would not be equal to zero, which means β_2 has impact on Y. Hence, the average adoption time for the stock firm, at any given firm size, is also different from the mutual firm.

Appendix

HW4 Q1&Q3

```
data<-read.csv("C:/Users/candi/Desktop/STAT 512/brand.csv",header=TRUE,sep = ",")
data

##      y x1 x2
## 1   64  4  2
## 2   73  4  4
## 3   61  4  2
## 4   76  4  4
## 5   72  6  2
## 6   80  6  4
## 7   71  6  2
## 8   83  6  4
## 9   83  8  2
## 10  89  8  4
## 11  86  8  2
## 12  93  8  4
## 13  88 10  2
## 14  95 10  4
## 15  94 10  2
## 16 100 10  4

colnames(data)<-c("y","x1","x2")
x1<-data$x1
x2<-data$x2
y<-data$y

##Q1 (a)
plot(data)
##install.packages("Hmisc")
library("Hmisc")

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units
```



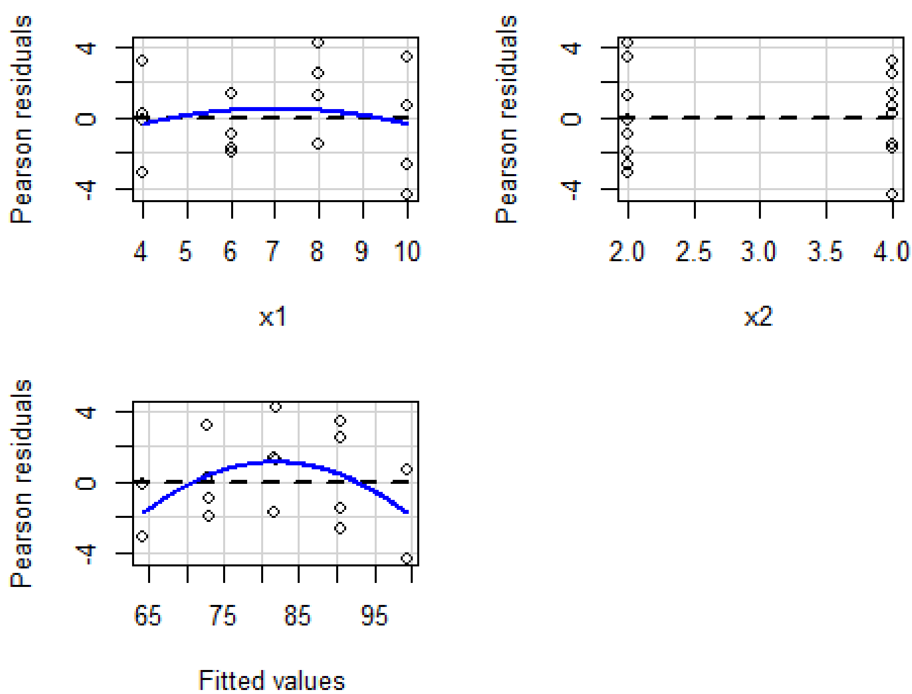

```
rcorr(as.matrix(data))

##       y    x1   x2
## y  1.00 0.89 0.39
## x1 0.89 1.00 0.00
## x2 0.39 0.00 1.00
##
## n= 16
##
##
## P
##    y      x1     x2
## y         0.0000 0.1304
## x1 0.0000        1.0000
## x2 0.1304 1.0000

##Q1 (b)
model1<-lm(y~x1+x2,data)
summary(model1)

##
## Call:
## lm(formula = y ~ x1 + x2, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.400 -1.762  0.025  1.587  4.200
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.6500     2.9961  12.566 1.20e-08 ***
## x1            4.4250     0.3011  14.695 1.78e-09 ***
## x2            4.3750     0.6733   6.498 2.01e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.693 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
```

```
##Q1 (c)
library(alr4)
```

```
## Loading required package: car

## Loading required package: carData

## Loading required package: effects

## Use the command
##     lattice::trellis.par.set(effectsTheme())
##   to customize lattice options for effects plots.
## See ?effectsTheme for details.
```

```
residualPlots(model1, tests=TRUE, quadratic=TRUE, smooth=FALSE)
```



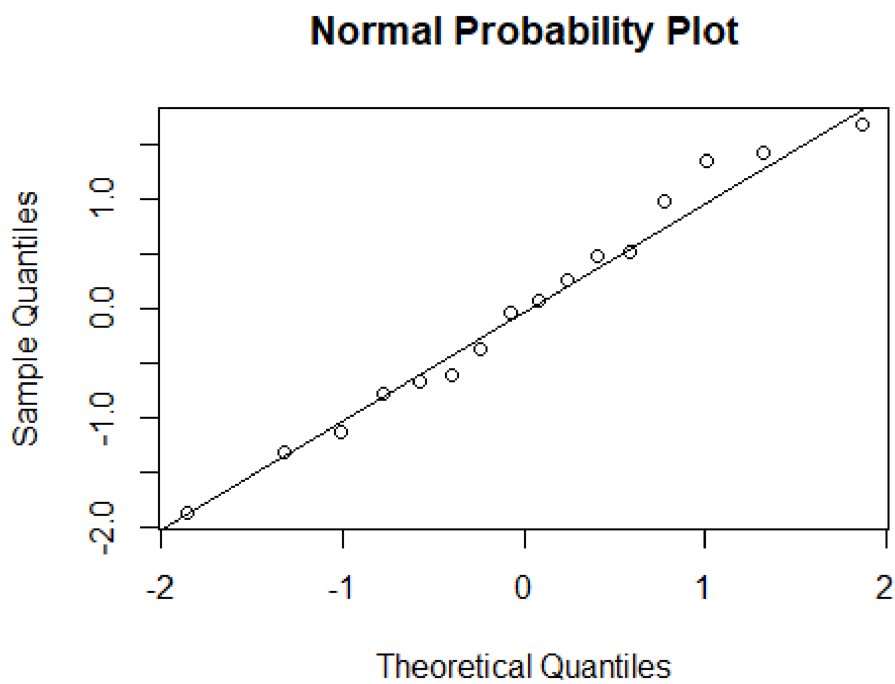

```
##            Test stat Pr(>|Test stat|)
## x1           -0.5416          0.59801
```

```
## x2             0.1994          0.84527
## Tukey test    -1.7221          0.08506 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
concentration.stdres=rstandard(model1)
shapiro.test(concentration.stdres)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  concentration.stdres
## W = 0.97496, p-value = 0.9111
```

```
qqnorm(concentration.stdres, main="Normal Probability Plot")
qqline(concentration.stdres)
```




```
##Q1 (d)
library(ALSM)
```

```
## Loading required package: leaps

## Loading required package: SuppDists

##
## Attaching package: 'ALSM'

## The following object is masked from 'package:lattice':
##
##     oneway
```

```
g<-rep(1,16)
g[y<=81.75]=0
bftest(model1,g)

##        t.value   P.Value alpha df
## [1,] 0.8629512 0.4027057  0.05 14

##Q1 (e)
model0<-lm(y~factor(x1)*factor(x2))
anova(model1,model0)

## Analysis of Variance Table
##
## Model 1: y ~ x1 + x2
## Model 2: y ~ factor(x1) * factor(x2)
##   Res.Df  RSS Df Sum of Sq     F Pr(>F)
## 1     13 94.3
## 2      8 57.0  5      37.3 1.047  0.453

model2<-lm(y~x2,data)
model3<-lm(y~x2+x1,data)
model4<-lm(y~x1,data)
summary(model3)

##
## Call:
## lm(formula = y ~ x2 + x1, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.400 -1.762  0.025  1.587  4.200
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.6500     2.9961  12.566 1.20e-08 ***
## x2            4.3750     0.6733   6.498 2.01e-05 ***
## x1            4.4250     0.3011  14.695 1.78e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.693 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

##Q1 (f)
#MSE
week<-c(64,73,61,76,72,80,71,83,83,89,86,93,88,95,94,100)
y0<-as.matrix(week)
colnames(y0)<-c("week")
temp1<-c(4,4,4,4,6,6,6,6,8,8,8,8,10,10,10,10)
temp2<-c(2,4,2,4,2,4,2,4,2,4,2,4,2,4,2,4)
Intercept<-rep(1,16)
x<-cbind(Intercept,temp1,temp2)
```

```
xty<-t(x)%*%y0
xty

##           week
## Intercept 1308
## temp1     9510
## temp2     3994

xtx<-t(x)%*%x
xtxinv<-solve(xtx)
xtxinv

##           Intercept         temp1   temp2
## Intercept    1.2375 -8.750000e-02 -0.1875
## temp1       -0.0875  1.250000e-02  0.0000
## temp2       -0.1875  2.602085e-17  0.0625

class(xty)

## [1] "matrix"

class(xtxinv)

## [1] "matrix"

betahat<-xtxinv %*% xty
betahat

##             week
## Intercept 37.650
## temp1      4.425
## temp2      4.375

x

##       Intercept temp1 temp2
##  [1,]         1     4     2
##  [2,]         1     4     4
##  [3,]         1     4     2
##  [4,]         1     4     4
##  [5,]         1     6     2
##  [6,]         1     6     4
##  [7,]         1     6     2
##  [8,]         1     6     4
##  [9,]         1     8     2
## [10,]         1     8     4
## [11,]         1     8     2
## [12,]         1     8     4
## [13,]         1    10     2
## [14,]         1    10     4
## [15,]         1    10     2
## [16,]         1    10     4

xtxinv
```

```
##             Intercept         temp1   temp2
## Intercept      1.2375 -8.750000e-02 -0.1875
## temp1         -0.0875  1.250000e-02  0.0000
## temp2         -0.1875  2.602085e-17  0.0625
```

```
xty
```

```
##             week
## Intercept 1308
## temp1     9510
## temp2     3994
```

```
hat<-x%*%xtxinv%*%t(x)
hat
```

```
##          [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]    [,8]
##  [1,]  0.2375  0.1125  0.2375  0.1125  0.1625  0.0375  0.1625  0.0375
##  [2,]  0.1125  0.2375  0.1125  0.2375  0.0375  0.1625  0.0375  0.1625
##  [3,]  0.2375  0.1125  0.2375  0.1125  0.1625  0.0375  0.1625  0.0375
##  [4,]  0.1125  0.2375  0.1125  0.2375  0.0375  0.1625  0.0375  0.1625
##  [5,]  0.1625  0.0375  0.1625  0.0375  0.1375  0.0125  0.1375  0.0125
##  [6,]  0.0375  0.1625  0.0375  0.1625  0.0125  0.1375  0.0125  0.1375
##  [7,]  0.1625  0.0375  0.1625  0.0375  0.1375  0.0125  0.1375  0.0125
##  [8,]  0.0375  0.1625  0.0375  0.1625  0.0125  0.1375  0.0125  0.1375
##  [9,]  0.0875 -0.0375  0.0875 -0.0375  0.1125 -0.0125  0.1125 -0.0125
## [10,] -0.0375  0.0875 -0.0375  0.0875 -0.0125  0.1125 -0.0125  0.1125
## [11,]  0.0875 -0.0375  0.0875 -0.0375  0.1125 -0.0125  0.1125 -0.0125
## [12,] -0.0375  0.0875 -0.0375  0.0875 -0.0125  0.1125 -0.0125  0.1125
## [13,]  0.0125 -0.1125  0.0125 -0.1125  0.0875 -0.0375  0.0875 -0.0375
## [14,] -0.1125  0.0125 -0.1125  0.0125 -0.0375  0.0875 -0.0375  0.0875
## [15,]  0.0125 -0.1125  0.0125 -0.1125  0.0875 -0.0375  0.0875 -0.0375
## [16,] -0.1125  0.0125 -0.1125  0.0125 -0.0375  0.0875 -0.0375  0.0875
##          [,9]   [,10]   [,11]   [,12]   [,13]   [,14]   [,15]   [,16]
##  [1,]  0.0875 -0.0375  0.0875 -0.0375  0.0125 -0.1125  0.0125 -0.1125
##  [2,] -0.0375  0.0875 -0.0375  0.0875 -0.1125  0.0125 -0.1125  0.0125
##  [3,]  0.0875 -0.0375  0.0875 -0.0375  0.0125 -0.1125  0.0125 -0.1125
##  [4,] -0.0375  0.0875 -0.0375  0.0875 -0.1125  0.0125 -0.1125  0.0125
##  [5,]  0.1125 -0.0125  0.1125 -0.0125  0.0875 -0.0375  0.0875 -0.0375
##  [6,] -0.0125  0.1125 -0.0125  0.1125 -0.0375  0.0875 -0.0375  0.0875
##  [7,]  0.1125 -0.0125  0.1125 -0.0125  0.0875 -0.0375  0.0875 -0.0375
##  [8,] -0.0125  0.1125 -0.0125  0.1125 -0.0375  0.0875 -0.0375  0.0875
##  [9,]  0.1375  0.0125  0.1375  0.0125  0.1625  0.0375  0.1625  0.0375
## [10,]  0.0125  0.1375  0.0125  0.1375  0.0375  0.1625  0.0375  0.1625
## [11,]  0.1375  0.0125  0.1375  0.0125  0.1625  0.0375  0.1625  0.0375
## [12,]  0.0125  0.1375  0.0125  0.1375  0.0375  0.1625  0.0375  0.1625
## [13,]  0.1625  0.0375  0.1625  0.0375  0.2375  0.1125  0.2375  0.1125
## [14,]  0.0375  0.1625  0.0375  0.1625  0.1125  0.2375  0.1125  0.2375
## [15,]  0.1625  0.0375  0.1625  0.0375  0.2375  0.1125  0.2375  0.1125
## [16,]  0.0375  0.1625  0.0375  0.1625  0.1125  0.2375  0.1125  0.2375
```

```
ide<-diag(16)
ide-hat
```

```
##          [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]    [,8]
##  [1,]  0.7625 -0.1125 -0.2375 -0.1125 -0.1625 -0.0375 -0.1625 -0.0375
##  [2,] -0.1125  0.7625 -0.1125 -0.2375 -0.0375 -0.1625 -0.0375 -0.1625
##  [3,] -0.2375 -0.1125  0.7625 -0.1125 -0.1625 -0.0375 -0.1625 -0.0375
##  [4,] -0.1125 -0.2375 -0.1125  0.7625 -0.0375 -0.1625 -0.0375 -0.1625
##  [5,] -0.1625 -0.0375 -0.1625 -0.0375  0.8625 -0.0125 -0.1375 -0.0125
##  [6,] -0.0375 -0.1625 -0.0375 -0.1625 -0.0125  0.8625 -0.0125 -0.1375
##  [7,] -0.1625 -0.0375 -0.1625 -0.0375 -0.1375 -0.0125  0.8625 -0.0125
##  [8,] -0.0375 -0.1625 -0.0375 -0.1625 -0.0125 -0.1375 -0.0125  0.8625
##  [9,] -0.0875  0.0375 -0.0875  0.0375 -0.1125  0.0125 -0.1125  0.0125
## [10,]  0.0375 -0.0875  0.0375 -0.0875  0.0125 -0.1125  0.0125 -0.1125
## [11,] -0.0875  0.0375 -0.0875  0.0375 -0.1125  0.0125 -0.1125  0.0125
## [12,]  0.0375 -0.0875  0.0375 -0.0875  0.0125 -0.1125  0.0125 -0.1125
## [13,] -0.0125  0.1125 -0.0125  0.1125 -0.0875  0.0375 -0.0875  0.0375
## [14,]  0.1125 -0.0125  0.1125 -0.0125  0.0375 -0.0875  0.0375 -0.0875
## [15,] -0.0125  0.1125 -0.0125  0.1125 -0.0875  0.0375 -0.0875  0.0375
## [16,]  0.1125 -0.0125  0.1125 -0.0125  0.0375 -0.0875  0.0375 -0.0875
##          [,9]   [,10]   [,11]   [,12]   [,13]   [,14]   [,15]   [,16]
##  [1,] -0.0875  0.0375 -0.0875  0.0375 -0.0125  0.1125 -0.0125  0.1125
##  [2,]  0.0375 -0.0875  0.0375 -0.0875  0.1125 -0.0125  0.1125 -0.0125
##  [3,] -0.0875  0.0375 -0.0875  0.0375 -0.0125  0.1125 -0.0125  0.1125
##  [4,]  0.0375 -0.0875  0.0375 -0.0875  0.1125 -0.0125  0.1125 -0.0125
##  [5,] -0.1125  0.0125 -0.1125  0.0125 -0.0875  0.0375 -0.0875  0.0375
##  [6,]  0.0125 -0.1125  0.0125 -0.1125  0.0375 -0.0875  0.0375 -0.0875
##  [7,] -0.1125  0.0125 -0.1125  0.0125 -0.0875  0.0375 -0.0875  0.0375
##  [8,]  0.0125 -0.1125  0.0125 -0.1125  0.0375 -0.0875  0.0375 -0.0875
##  [9,]  0.8625 -0.0125 -0.1375 -0.0125 -0.1625 -0.0375 -0.1625 -0.0375
## [10,] -0.0125  0.8625 -0.0125 -0.1375 -0.0375 -0.1625 -0.0375 -0.1625
## [11,] -0.1375 -0.0125  0.8625 -0.0125 -0.1625 -0.0375 -0.1625 -0.0375
## [12,] -0.0125 -0.1375 -0.0125  0.8625 -0.0375 -0.1625 -0.0375 -0.1625
## [13,] -0.1625 -0.0375 -0.1625 -0.0375  0.7625 -0.1125 -0.2375 -0.1125
## [14,] -0.0375 -0.1625 -0.0375 -0.1625 -0.1125  0.7625 -0.1125 -0.2375
## [15,] -0.1625 -0.0375 -0.1625 -0.0375 -0.2375 -0.1125  0.7625 -0.1125
## [16,] -0.0375 -0.1625 -0.0375 -0.1625 -0.1125 -0.2375 -0.1125  0.7625
```

```
resid<-y0-x%*%betahat
resid
```

```
##        week
##  [1,] -0.10
##  [2,]  0.15
##  [3,] -3.10
##  [4,]  3.15
##  [5,] -0.95
##  [6,] -1.70
##  [7,] -1.95
##  [8,]  1.30
##  [9,]  1.20
## [10,] -1.55
## [11,]  4.20
## [12,]  2.45
## [13,] -2.65
```

```
## [14,] -4.40
## [15,]  3.35
## [16,]  0.60

t(resid)%*%resid

##      week
## week 94.3

model<-lm(week~temp1+temp2)  #double check
#model$residuals
summary(model)

##
## Call:
## lm(formula = week ~ temp1 + temp2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.400 -1.762  0.025  1.587  4.200
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.6500     2.9961  12.566 1.20e-08 ***
## temp1         4.4250     0.3011  14.695 1.78e-09 ***
## temp2         4.3750     0.6733   6.498 2.01e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.693 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

anova(model)

## Analysis of Variance Table
##
## Response: week
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## temp1      1 1566.45 1566.45 215.947 1.778e-09 ***
## temp2      1  306.25  306.25  42.219 2.011e-05 ***
## Residuals 13   94.30    7.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

J<-matrix(1,nrow=16,ncol=16)
J

##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
##  [1,]    1    1    1    1    1    1    1    1    1     1     1     1     1
##  [2,]    1    1    1    1    1    1    1    1    1     1     1     1     1
##  [3,]    1    1    1    1    1    1    1    1    1     1     1     1     1
##  [4,]    1    1    1    1    1    1    1    1    1     1     1     1     1
```

```
##  [5,]    1    1    1    1    1    1    1    1    1     1     1     1     1
##  [6,]    1    1    1    1    1    1    1    1    1     1     1     1     1
##  [7,]    1    1    1    1    1    1    1    1    1     1     1     1     1
##  [8,]    1    1    1    1    1    1    1    1    1     1     1     1     1
##  [9,]    1    1    1    1    1    1    1    1    1     1     1     1     1
## [10,]    1    1    1    1    1    1    1    1    1     1     1     1     1
## [11,]    1    1    1    1    1    1    1    1    1     1     1     1     1
## [12,]    1    1    1    1    1    1    1    1    1     1     1     1     1
## [13,]    1    1    1    1    1    1    1    1    1     1     1     1     1
## [14,]    1    1    1    1    1    1    1    1    1     1     1     1     1
## [15,]    1    1    1    1    1    1    1    1    1     1     1     1     1
## [16,]    1    1    1    1    1    1    1    1    1     1     1     1     1
##       [,14] [,15] [,16]
##  [1,]     1     1     1
##  [2,]     1     1     1
##  [3,]     1     1     1
##  [4,]     1     1     1
##  [5,]     1     1     1
##  [6,]     1     1     1
##  [7,]     1     1     1
##  [8,]     1     1     1
##  [9,]     1     1     1
## [10,]     1     1     1
## [11,]     1     1     1
## [12,]     1     1     1
## [13,]     1     1     1
## [14,]     1     1     1
## [15,]     1     1     1
## [16,]     1     1     1
```

```
SST<-t(y0)%*%y0-(1/16)*t(y0)%*%J%*%y0
SST
```

```
##      week
## week 1967
```

```
SSE<-t(resid)%*%resid
SSE
```

```
##      week
## week 94.3
```

```
SSR<-SST-SSE
SSR
```

```
##        week
## week 1872.7
```

```
MSE<-SSE/13
MSE
```

```
##          week
## week 7.253846
```

```
##Vairance-Covariance Matrix of Estimator
sigmaBeta<-MSE[1,1]*solve(t(x)%*%x)
sigmaBeta

##             Intercept         temp1      temp2
## Intercept  8.9766346 -6.347115e-01 -1.3600962
## temp1     -0.6347115  9.067308e-02  0.0000000
## temp2     -1.3600962  1.887513e-16  0.4533654

##Vairance-Covariance Matrix of predictors
newx<-rbind(1,5,4)
newx

##      [,1]
## [1,]    1
## [2,]    5
## [3,]    4

t(newx)

##      [,1] [,2] [,3]
## [1,]    1    5    4

yhat<-t(newx)%*%betahat
yhat

##        week
## [1,] 77.275

varYhat<-t(newx)%*%sigmaBeta%*%newx
varYhat

##          [,1]
## [1,] 1.269423

##Q3 (a)
anova(lm(y~x2,data))

## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq F value Pr(>F)
## x2         1  306.25  306.25  2.5817 0.1304
## Residuals 14 1660.75  118.62

anova(lm(y~x2+x1,data))

## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq F value    Pr(>F)    
## x2         1  306.25  306.25  42.219 2.011e-05 ***
## x1         1 1566.45 1566.45 215.947 1.778e-09 ***
## Residuals 13   94.30    7.25
```

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

##Q3 (c)
##install.packages("rsq")
anova(lm(y~x1,data))

## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## x1         1 1566.45 1566.45  54.751 3.356e-06 ***
## Residuals 14  400.55   28.61
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(lm(y~x1+x2,data))

## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## x1         1 1566.45 1566.45 215.947 1.778e-09 ***
## x2         1  306.25  306.25  42.219 2.011e-05 ***
## Residuals 13   94.30    7.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(lm(y~x2+x1,data))

## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## x2         1  306.25  306.25  42.219 2.011e-05 ***
## x1         1 1566.45 1566.45 215.947 1.778e-09 ***
## Residuals 13   94.30    7.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova(lm(y~x1+x2,data),type="II")

## Anova Table (Type II tests)
##
## Response: y
##            Sum Sq Df F value    Pr(>F)
## x1        1566.45  1 215.947 1.778e-09 ***
## x2         306.25  1  42.219 2.011e-05 ***
## Residuals   94.30 13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(lm(y~x1+x2,data))
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.400 -1.762  0.025  1.587  4.200
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.6500     2.9961  12.566 1.20e-08 ***
## x1            4.4250     0.3011  14.695 1.78e-09 ***
## x2            4.3750     0.6733   6.498 2.01e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.693 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
```

```r
summary(lm(y~x2+x1,data))
```

```
##
## Call:
## lm(formula = y ~ x2 + x1, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.400 -1.762  0.025  1.587  4.200
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.6500     2.9961  12.566 1.20e-08 ***
## x2            4.3750     0.6733   6.498 2.01e-05 ***
## x1            4.4250     0.3011  14.695 1.78e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.693 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
```

```r
##library("rsq")
##rsq.partial(model4)
##rsq.partial(model3,model2)
##rsq.partial(model1,model4)
```

HW4 Q5

```
data<-read.csv("C:/Users/candi/Desktop/STAT 512/insurance.csv",header=TRUE,sep = ",
")
data

##    months size type
## 1      17  151    0
## 2      26   92    0
## 3      21  175    0
## 4      30   31    0
## 5      22  104    0
## 6       0  277    0
## 7      12  210    0
## 8      19  120    0
## 9       4  290    0
## 10     16  238    0
## 11     28  164    1
## 12     15  272    1
## 13     11  295    1
## 14     38   68    1
## 15     31   85    1
## 16     21  224    1
## 17     20  166    1
## 18     13  305    1
## 19     30  124    1
## 20     14  246    1

colnames(data)<-c("months","size","type")
months<-data$months
size<-data$size
type<-data$type
fit1<-lm(months~size+type+size*type,data)
fit2<-lm(months~size,data)

##Q5 (a)
anova(fit1)

## Analysis of Variance Table
##
## Response: months
##           Df  Sum Sq Mean Sq  F value    Pr(>F)
## size       1 1188.17 1188.17 107.7819 1.627e-08 ***
## type       1  316.25  316.25  28.6875 6.430e-05 ***
## size:type  1    0.01    0.01   0.0005    0.9821
## Residuals 16  176.38   11.02
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(fit2)

## Analysis of Variance Table
##
```

```
## Response: months
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## size       1 1188.17 1188.17  43.414 3.452e-06 ***
## Residuals 18  492.63   27.37
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##Q5 (b)
summary(fit1)
```

```
##
## Call:
## lm(formula = months ~ size + type + size * type, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7144 -1.7064 -0.4557  1.9311  6.3259
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.8383695  2.4406498  13.864 2.47e-10 ***
## size        -0.1015306  0.0130525  -7.779 7.97e-07 ***
## type         8.1312501  3.6540517   2.225   0.0408 *
## size:type   -0.0004171  0.0183312  -0.023   0.9821
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.32 on 16 degrees of freedom
## Multiple R-squared:  0.8951, Adjusted R-squared:  0.8754
## F-statistic: 45.49 on 3 and 16 DF,  p-value: 4.675e-08
```

```
##Q5 (c)
fit3<-lm(months~type,data)
anova(fit3)
```

```
## Analysis of Variance Table
##
## Response: months
##           Df Sum Sq Mean Sq F value Pr(>F)
## type       1  145.8 145.800  1.7097 0.2075
## Residuals 18 1535.0  85.278
```

```
##Q4 (d)
fit4<-lm(months~size+type, data)
summary(fit4)
```

```
##
## Call:
## lm(formula = months ~ size + type, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.6915 -1.7036 -0.4385  1.9210  6.3406
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.874069   1.813858  18.675 9.15e-13 ***
## size        -0.101742   0.008891 -11.443 2.07e-09 ***
## type         8.055469   1.459106   5.521 3.74e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.221 on 17 degrees of freedom
## Multiple R-squared:  0.8951, Adjusted R-squared:  0.8827
## F-statistic:  72.5 on 2 and 17 DF,  p-value: 4.765e-09
```

```r
anova(fit4)
```

```
## Analysis of Variance Table
##
## Response: months
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## size       1 1188.17 1188.17  114.51 5.683e-09 ***
## type       1  316.25  316.25   30.48 3.742e-05 ***
## Residuals 17  176.39   10.38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```