STAT512 Division 1 HW 4

Yi Yang

10/13/2018

1. In a small scale experimental study of the relation between degree of brand liking (Y) and moisture content (X_1) and sweetness (X_2) of the product, data is in brand.csv. Sample size is 16. Use R to

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

b) fit regression model to the data without interaction, $\hat{Y} = \beta_1 X_1 + \beta_2 X_2$

c) Perform a test to see if the residuals are normal.

d) Perform BF test for accuracy of the residuals. You can make two groups ($Y \leq 81.75$, or > 81.75, where 81.75 is the average).

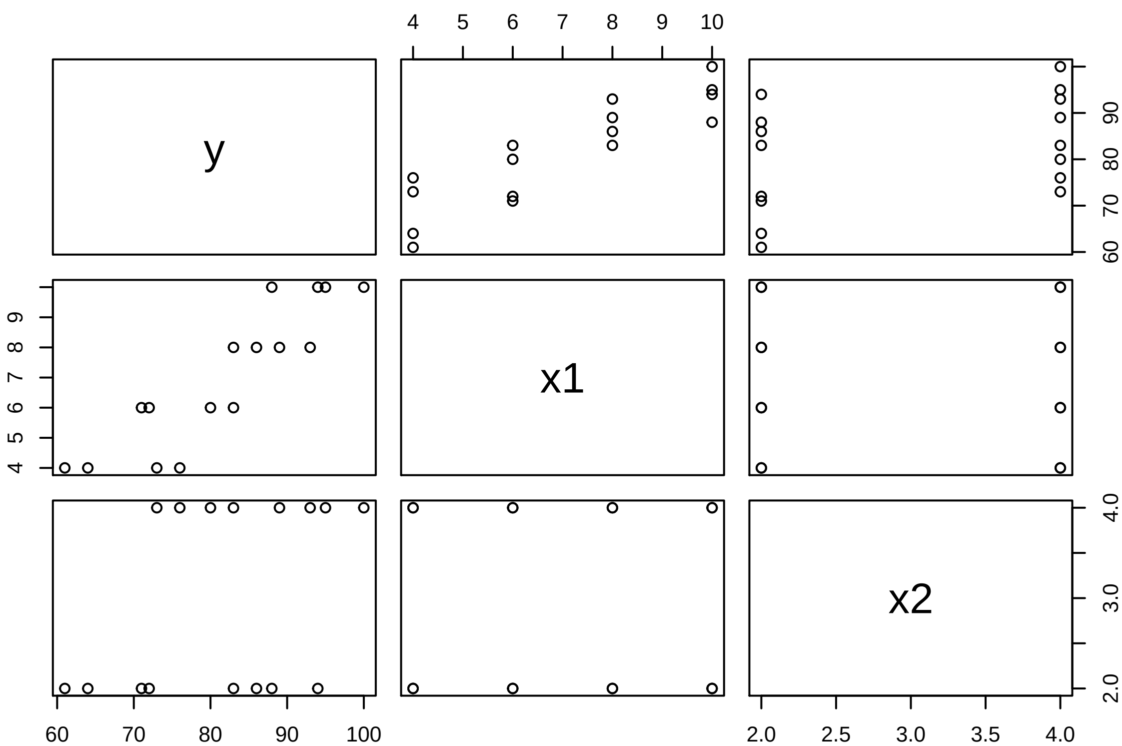
e) Perform a lack of fit test of the model use a significant level of 0.01. State H_0 and H_a, test statistics, critical value, p value and conclusion.

f) 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 $X_1 = 5, X_2 = 4$.

Solution:

a)

```r
brand <- read.csv('data/brand.csv',header = TRUE)
plot(brand)
```




```r
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
```

```r
rcorr(as.matrix(brand))
```

```
##       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 matrix is given by

$$r = \begin{bmatrix} 1.00 & 0.89 & 0.39 \\ 0.89 & 1.00 & 0.00 \\ 0.39 & 0.00 & 1.00 \end{bmatrix}$$

b)

```r
brand.mod <- lm(y~x1 + x2, brand)
summary(brand.mod)
```

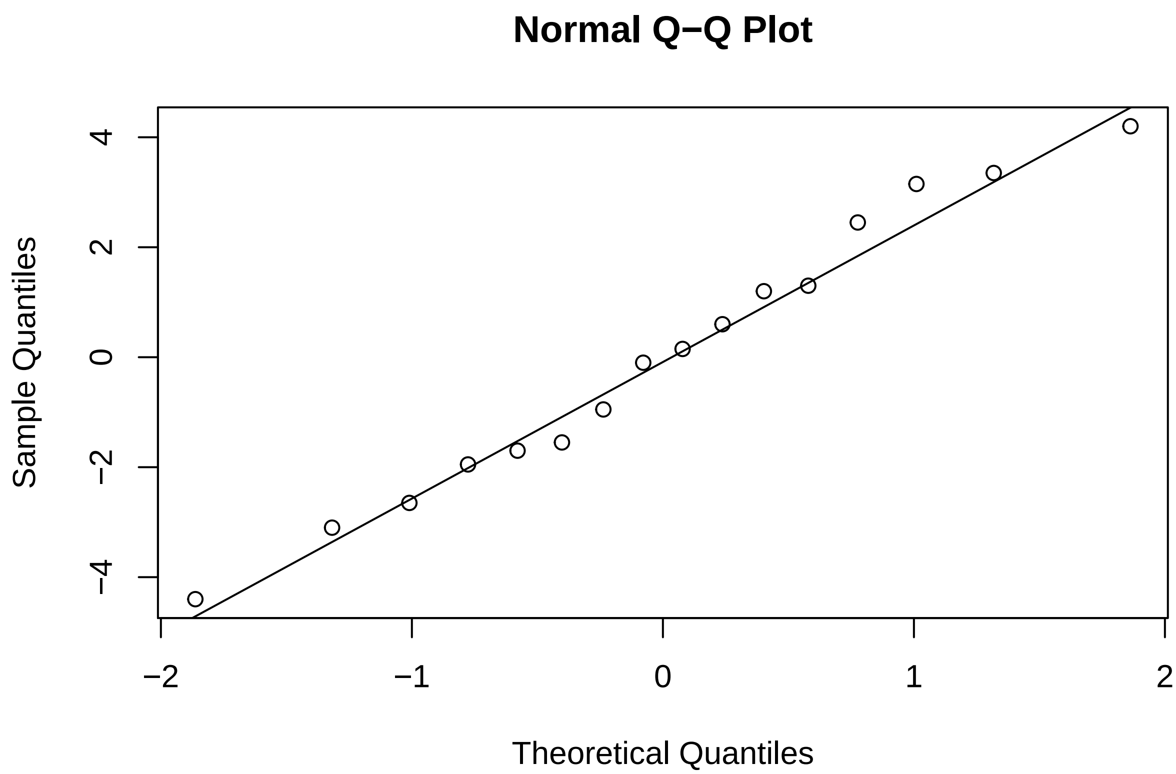
```
##
## Call:
## lm(formula = y ~ x1 + x2, data = brand)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.400 -1.762  0.025  1.587  4.200
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.6500     2.9961  12.566 1.20e-08 ***
## x1            4.4250     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
```

The linear regression fit model without interaction term is given by

$$Y = 37.650 + 4.425X_1 + 4.375X_2$$

c) Let us use a probability plot and a Shapiro-Wilk test to test the normality of the error terms.

```
brand.res <- residuals(brand.mod)
qqnorm(brand.res)
qqline(brand.res)
```




```
shapiro.test(brand.res)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  brand.res
## W = 0.97585, p-value = 0.9222
```

From the test results, we confirm that the residuals are normal.

d) Brown-Forsythe Test to test the constancy of residual variance is given by

```
library(ALSM)
```

```
## Loading required package: leaps

## Loading required package: SuppDists

## Loading required package: car

## Loading required package: carData

##
## Attaching package: 'ALSM'
```

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

```r
g <- rep(1,16)
g[brand$y <= 81.75] = 0
bftest(brand.mod,g)
```

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

The difference is not significant, which means the variance of residuals is constant.

e) The lack of fit test is stated as follows

$$H_0 : \mathbb{E}\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

$$H_a : \mathbb{E}\{Y\} \neq \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

```r
brand.mod.full <- lm(y~factor(x1)*factor(x2), brand)
anova(brand.mod, brand.mod.full)
```

```
## 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
```

```r
qf(1-0.05,5,8)
```

```
## [1] 3.687499
```

Based on the ANOVA table, $MSPE = 57.0/8 = 7.125$, $MSLF = 37.3/5 = 7.46$, the test statistic is $Ts = \frac{MSLF}{MSPE} = 1.047$, the critical value is 3.687499 at significance level $\alpha = 0.05$, p-value is 0.453, which is greater than α. Therefore, we fail to reject the null hypothesis.

f) Based on design matrix, we are able to obtain the variance-covariance matrix of estimators, variance-covariance matrixx of predictors.

```r
MSE <- 94.3/13
X.design <- matrix(c(rep(1,16),brand$x1,brand$x2),nrow = 16,ncol = 3)
Normal.inv <- solve(t(X.design)%*%X.design)
Sigmab <- MSE*Normal.inv
Xh <- matrix(c(1,5,4),nrow = 3,ncol = 1)
SigmaYh <- t(Xh)%*%Sigmab%*%Xh
MSE
```

```
## [1] 7.253846
```

```r
print(Sigmab)
```

```
##            [,1]          [,2]       [,3]
## [1,]  8.9766346 -6.347115e-01 -1.3600962
## [2,] -0.6347115  9.067308e-02  0.0000000
## [3,] -1.3600962  1.887513e-16  0.4533654
```

```r
print(SigmaYh)
```

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

Based on the calculation, $MSE = 7.253846$.

$$\Sigma_{\{b\}} = \begin{bmatrix} 8.9766346 & -0.6347115 & -1.36009621 \\ -0.6347115 & 0.09067308 & 0.0000000 \\ -1.3600962 & 1.887513e-16 & 0.4533654 \end{bmatrix}, \quad \Sigma_{\{\hat{Y}_h\}} = 1.269423$$

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

a) Obtain an interval estimate of $\mathbb{E}\{Y_h\}$ (i.e., $\hat{Y}_h$) when $(X_1 = 5, X_2 = 4)$, with 99% confidence level.

b) Obtain an interval estimate of a single predictor $\hat{Y}_h\{new\}$ when $(X_1 = 5, X_2 = 4)$, with 99% confidence level.

c) Obtain an interval estimate of the average of the next two predictors, when $(X_1 = 5, X_2 = 4)$, with 90% confidence level.

d) Obtain a simultaneous estimate of the two (single) predictor $\hat{Y}_h\{new\}$ when $(X_1 = 5, X_2 = 4)$, and $(X_1 = 6, X_2 = 5)$, with a 90% confidence level.

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

Solution:

a) The interval estimate is given by $\hat{Y}_h \pm t(1-\alpha/2; n-p)s\{\hat{Y}_h\}$. $t(1-0.01/2; 13) = 3.012276$, Therefore, the interval is given by $[73.88111, 80.66889]$.

```
Xh2 <- matrix(c(1,6,5),nrow = 3,ncol = 1)
SigmaYh2 <- t(Xh2)%*%Sigmab%*%Xh2
qt(1-0.1/2,13)
```

```
## [1] 1.770933
```

```
Y <-37.650 + 4.425*5 + 4.375*4
Y2 <- 37.650 + 4.425*6 + 4.375*5
Y
```

```
## [1] 77.275
```

```
sqrt(SigmaYh + MSE)
```

```
##          [,1]
## [1,] 2.919464
```

```
Y + sqrt(SigmaYh + MSE/2)*1.770933
```

```
##          [,1]
## [1,] 81.19367
```

```
qt(1-0.1/4,13)
```

```
## [1] 2.160369
```

```
Y - sqrt(SigmaYh + MSE)*qt(1-0.1/4,13)
```

```
##          [,1]
## [1,] 70.96788
```

```
Y + sqrt(SigmaYh + MSE)*qt(1-0.1/4,13)
```

```
##          [,1]
## [1,] 83.58212
```

```r
Y2 - sqrt(SigmaYh2 + MSE)*qt(1-0.1/4,13)
```

```
##          [,1]
## [1,] 79.37739
```

```r
Y2 + sqrt(SigmaYh2 + MSE)*qt(1-0.1/4,13)
```

```
##          [,1]
## [1,] 92.77261
```

```r
37.650 - qt(1-0.1/6,13)*sqrt(8.9766346)
```

```
## [1] 30.5205
```

```r
37.650 + qt(1-0.1/6,13)*sqrt(8.9766346)
```

```
## [1] 44.7795
```

```r
4.425 - qt(1-0.1/6,13)*sqrt(9.067308e-02)
```

```
## [1] 3.708458
```

```r
4.425 + qt(1-0.1/6,13)*sqrt(9.067308e-02)
```

```
## [1] 5.141542
```

```r
4.375 - qt(1-0.1/6,13)*sqrt(0.4533654)
```

```
## [1] 2.772763
```

```r
4.375 + qt(1-0.1/6,13)*sqrt(0.4533654)
```

```
## [1] 5.977237
```

b) The confidence interval estimate for the new observation is given by $\hat{Y}_h \pm t(1-\alpha/2; n-p)s\{\hat{Y}_h\{new\}\}$ with $s\{\hat{Y}_h\{new\}\} = \sqrt{MSE + s^2\{\hat{Y}_h\}} = 2.916953$. Therefore, the numerical interval is estimated by $[68.48077, 86.06923]$.

c) Similarly, the standard deviation of prediction mean is given by $s\{predmean\} = \sqrt{MSE/m + s^2\{\hat{Y}_h\}} = 2.209455$. Therefore, the estimated confidence interval is given by $[73.35633, 81.19367]$.

d) Let us estimate with the Bonferoni simultaneous prediction, which is $\hat{Y}_h \pm Bs\{pred\}$ with $B = t(1-\alpha/2g; n-p) = t(1-0.1/4; 13) = 2.160369$. Therefore, the estimated interval is $[70.96788, 83.58212]$ and $[79.37739, 92.77261]$.

e) The Bonferroni estimated interval for coefficients can be computed, given $s\{b_0\} = \sqrt{8.9766346}, s\{b_1\} = \sqrt{9.067308e-02}, s\{b_2\} = \sqrt{0.4533654}$. Then interval is given by $b_k \pm Bs\{b_k\}$ with $B = t(1-\alpha/6, 13)$. Therefore, the interval for β_0 is $[30.5205, 44.7795]$, the interval for β_1 is $[3.708458, 5.141542]$ and the interval for β_3 is $[2.772763, 5.977237]$.

3. Refer to question 1,

a) What is the ANOVA table that decomposes the regression sum of squares into extra sums of squares associated with X_2, then with X_1, given X_2. (You may use R for this question).

b) Test whether X_1 can be dropped from the regression model given X_2 is retained. Use the partial F test with a significant level of 0.01. Define H_0 and H_a, test statistic, critical value, and state conclusion.

c) Compute R^2_{Y1}, $R^2_{Y1|2}$, $R^2_{Y2|1}$ and R^2. Explain what each coefficient measures and interpret your result.

Solution:

a) Type I ANOVA table can do that

```r
brand.mod.reverse <- lm(y~x2 + x1, brand)
brand.mod
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, data = brand)
##
## Coefficients:
## (Intercept)           x1           x2
##      37.650        4.425        4.375
```

```r
brand.mod.reverse
```

```
##
## Call:
## lm(formula = y ~ x2 + x1, data = brand)
##
## Coefficients:
## (Intercept)           x2           x1
##      37.650        4.375        4.425
```

```r
anova(lm(y~x2,brand))
```

```
## 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
```

```r
anova(lm(y~x1,brand))
```

```
## 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
anova(brand.mod.reverse)
```

```
## 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
```

The extra sum of squares are given by $SSR(X_2) = 306.25$ and $SSR(X_1|X_2) = 1566.45$.

b) The Hypothesis test is drafted as follows

$$H_0 : \beta_1 = 0, \quad H_a : \beta_1 \neq 0$$

The test statistic is given by

$$F^* = \frac{MSR(X_1|X_2)}{MSE(X_1,X_2)} = \frac{1566.45/1}{94.3/13} = 215.95$$

The critical value is given by $qf(0.99;1,13) = 9.073806$, which is less than the test statistic. It is concluded that the Null hypothesis can be rejected, so that X_1 can be dropped.

```
qf(0.99,1,13)
```

```
## [1] 9.073806
```

c) The coefficients of partial determination are given by

```
anova(brand.mod)
```

```
## 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
```

```
library(car)
Anova(lm(y~x1+x2,brand), 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
```

$$R^2_{Y1} = \frac{SSR(X_1)}{SSR(X_1)+SSR(X_2|X_1)+SSE(X_1,X_2)} = \frac{1566.45}{1566.45+306.25+94.30} = 0.7964$$

$$R^2_{Y1|2} = \frac{SSR(X_1|X_2)}{SSE(X_2)} = \frac{1566.45}{94.30+1566.45} = 0.9432$$

$$R^2_{Y2|1} = \frac{SSR(X_2|X_1)}{SSE(X_1)} = \frac{306.25}{306.25+94.30} = 0.7646$$

$$R^2 = \frac{SSR(X_1,X_2)}{SST(X_1,X_2)} = \frac{1566.45+306.25}{1566.45+306.25+94.30} = 0.9521$$

R^2_{Y1} measures the portion of total variation of Y due to introducing predictor X_1. $R^2_{Y1|2}$ measures the relative marginal reduction in the variation in Y assocaited with X_1 when X_2 already in the model. $R^2_{Y2|1}$ measures the relative marginal reduction in the variation in Y assocaited with X_2 when X_1 already in the model. R^2 measures the proportion of variation in Y explained by the model.

4. 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

X_1: age

X_2: operating expense

X_3: vacancy rates

X_4: total square footage

According to the following ANOVA table, perform the following test, use a significant level of 0.01. Define H_0 and H_a, 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) Whether X_3 can be dropped from the regression model given that X_1, X_2 and X_4 are retained.

b) Whether X_2 and X_3 can be dropped from the regression model given that X_1 and X_4 are retained.

c) Compute $R^2_{Y3|1,2,4}$.

Solution:

a) The test hypothesis is given by

$$H_0 : \beta_3 = 0, \quad H_a : \beta_3 \neq 0$$

The test statistic is given by

$$F^* = \frac{MSR(X_3|X_1, X_2, X_4)}{MSE(X_1, X_2, X_3, X_4)} = \frac{0.420}{1.293} = 0.325$$

The critical value is $qf(0.99, 1, 76) = 6.980578$, which is greater than F^*. Therefore, the null hypothesis can not be rejected. X_3 can be dropped from the regression model.

```
qf(0.99,1,76)
```

```
## [1] 6.980578
```

b) The test hypothesis is given by

$$H_0 : \beta_2 = \beta_3 = 0, \quad H_a : \text{not all } \beta_2, \beta_3 \text{ equal to zero}$$

The test statistic is given by

$$F^* = \frac{MSR(X_2, X_3|X_1, X_4)}{MSE(X_1, X_2, X_3, X_4)} = \frac{(0.420 + 27.857)/2}{1.293} = 10.935$$

The critical value is given by $qf(0.99, 2, 76) = 4.89584$, which is less than the test statistic. The null hypothesis can be rejected, which means not all β_2 and β_3 equal to zero.

```
qf(0.99,2,76)
```

```
## [1] 4.89584
```

c) The coefficient of partial determination is given by

$$R^2_{Y3|1,2,4} = \frac{SSR(X_3|X_1, X_2, X_4)}{SSE(X_1, X_2, X_4)} = \frac{0.420}{98.231 + 0.420} = 0.004257$$

5. 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 (X_1) and the type of firm (X_2, stock company and mutual company). Data is in insurance.csv

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

Perform hypothesis test for the following question. Use a significant level of 0.1. Define H_0 and H_a, test statistic, critical value, and state conclusion.

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

b) The firm size (X_1) has the same impact on the adopt time in mutual firm and stock firm.

c) The firm size (X_1) has no impact on the adopt time in mutual firm and stock firm.

d) If the firm size (X_1) 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.

Solution:

a) For a mutual company, the average adopt time is given by $Y = \beta_0 + \beta_1 X_1$. For a stock company, the average adopt time is given by $Y = \beta_0 + \beta_2 + (\beta_1 + \beta_3) X_1$. The hypothesis test for this subproblem is given by

$$H_0 : \beta_2 = \beta_3 = 0, \quad H_a : \text{not all } \beta_2, \beta_3 = 0$$

```r
insurance <- read.csv('data/insurance.csv',header = FALSE)
Y <- insurance$V1
X1 <- insurance$V2
X2 <- insurance$V3
insurance.mod <- lm(Y~X1+X2+X1*X2,insurance)
summary(insurance.mod)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X1 * X2, data = insurance)
##
## 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 ***
## X1          -0.1015306  0.0130525  -7.779 7.97e-07 ***
## X2           8.1312501  3.6540517   2.225   0.0408 *
## X1:X2       -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
```

```r
anova(insurance.mod)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df  Sum Sq Mean Sq  F value    Pr(>F)
## X1         1 1188.17 1188.17 107.7819 1.627e-08 ***
## X2         1  316.25  316.25  28.6875 6.430e-05 ***
## X1:X2      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
```

The test statistic is given by $F^* = MSR(X_2, X_3|X_1)/MSE = 316.26/2/11.02 = 14.35$, the critical value is $qf(0.9, 2, 16) = 2.668171$, which is less than the test statistic. Therefore, the null hypothesis can be rejected, the mutual firm and the stock firm do not have the same average adopt time for any firm size.

```r
qf(0.9,2,16)
```

```
## [1] 2.668171
```

b) The test hypothesis is given by

$$H_0 : \beta_3 = 0, \quad H_a : \beta_3 \neq 0$$

The test statistic is given by $F^* = MSR(X_3|X_1, X_2)/MSE = 0.001$, the critical value is given by $qf(0.9, 1, 16) = 3.04811$, which is greater than test statistic. The null hypothesis can not be rejected, the firm size has the same impact on the adoption time in both type firms.

```r
qf(0.9,1,16)
```

```
## [1] 3.04811
```

c) The test hypothesis is given by

$$H_0 : \beta_1 = \beta_3 = 0, \quad H_a : \text{not all } \beta_1, \beta_3 = 0$$

Test statistic is given by $F^* = MSR(X_1, X_3|X_2)/MSE = 1358.62/2/11.02 = 61.643$, the critical value is 2.668171, which is less than the test statistic. Therefore, the null hypothesis can be rejected. The firm size can not have no impact on the adopt time in mutual firm and stock firm.

```r
anova(lm(Y~X2+X1+X1*X2,insurance))
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df  Sum Sq Mean Sq  F value    Pr(>F)
## X2         1  145.80  145.80  13.2259   0.00222 **
## X1         1 1358.61 1358.61 123.2435 6.299e-09 ***
## X2:X1      1    0.01    0.01   0.0005   0.98213
## Residuals 16  176.38   11.02
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

d) Since $\beta_3 = 0$, the new ANOVA table is given by

```r
anova(lm(Y~X1+X2, insurance))
```

```
## Analysis of Variance Table
##
## Response: Y
```

```
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## X1         1 1188.17 1188.17  114.51 5.683e-09 ***
## X2         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
```

```r
qf(0.9,1,17)
```

```
## [1] 3.026232
```

The test hypothesis is given by

$$H_0 : \beta_2 = 0, \quad H_a : \beta_2 \neq 0$$

The test statistic is given by $F^* = MSR(X_2|X_1)/MSE = 316.25/10.38 = 30.467$, the critical value is given by $qf(0.9, 1, 17) = 3.026232$, which is less than the test statistic. Therefore, the null hypothesis can be rejected, the average adoption time for the stock firm can not be the same as the mutual firm at any given firm size.