Assignment #3 (G1)

Regression

Group Assignment

Class section 23 team members:

Yutong Shen yshen50@simon.rochester.edu

Ruohong Li rli66@ur.rochester.edu

Yuxiao Yao yyao41@ur.rochester.edu

Yijia Liu yliu304@ur.rochester.edu

Saivarshini Ravichandran sravich5@simon.rochester.edu

1. Estimating Firm β's

As a write-up, fill in the tables below and use about one page of text to discuss:
a) Rationale for your choice of indexes and data frequency

Variables	Data source and/or index used	Data frequency used
R: Return for each stock HD AAPL VZ CSCO	R = p(t)/p(t-1) -1, p is the close price of the company in each month from 1990 to 1999;	monthly data, 12 per year
Rf: Return on a safe, risk-free investment	we are using the data from the monthly Treasury Yield between 1990 and 1999, it is equal to the monthly close price divided by 1200.	monthly data, 12 per year
Rm: Return on the "market portfolio"	is calculated based on Monthly Wilshire 5000 price(W5000) between 1990 and 1999, which equal to Rm = W5000(t)/W5000(t-1) -1.	monthly data, 12 per year

Data Preprocessing:

1) Download data from Yahoo finance. Link of data source:

    1990-1999 Apple's monthly close price:
    https://finance.yahoo.com/quote/AAPL/history?period1=633744000&period2=946598400&interval=1mo&filter=history&frequency=1mo&includeAdjustedClose=true

    1990-1999 Home Depot's monthly close price:
    https://finance.yahoo.com/quote/HD/history?period1=631152000&period2=946598400&interval=1mo&filter=history&frequency=1mo&includeAdjustedClose=true

    1990-1999 Verizon's monthly close price:
    https://finance.yahoo.com/quote/VZ/history?period1=631152000&period2=946598400&interval=1mo&filter=history&frequency=1mo&includeAdjustedClose=true

    1990-1999 Cisco's monthly close price:
    https://finance.yahoo.com/quote/CSCO/history?period1=635212800&period2=946598400&interval=1mo&filter=history&frequency=1mo&includeAdjustedClose=true

    1990-1999 Treasury Yield 10 Years monthly data:
    https://finance.yahoo.com/quote/%5ETNX/history?period1=631152000&period2=946598400&interval=1mo&filter=history&frequency=1mo&includeAdjustedClose=true

1990-1999 Wilshire 5000 monthly price data:

https://finance.yahoo.com/quote/%5EW5000/history?period1=633744000&period2=946598400&interval=1mo&filter=history&frequency=1mo&includeAdjustedClose=true

2) Use Excel to preprocess the data.
Use companies' close data and the equation R = p(t)/p(t-1) -1 to calculate the return value for four companies.

```
=B5/B4 -1
```

Use the monthly Treasury Yield 10 Years value divide 1200 to calculate the R_f

```
=V2/1200
```

Use the Wilshire 5000 monthly price data and the formula R = p(t)/p(t-1) -1 to calculate the R_m

```
=Z3/Z2-1
```

Calculate the value $(R_m - R_f)$ and (R - R_f)

```
=AA3-W3
```

```
=C5-W5
```

3) Clean data

```
rm(list=ls()); #clean the data
```

4) Download the "readxl" package and Load the data "CAPM.xlsx" to Rstudio

```
library(readxl)  # load package
CAPM <- read_excel("Desktop/CAPM.xlsx") #Load data(CAPM.xlsx)
```

5) Present and summary the data

```
CAPM #present data
summary(CAPM) #present information about data
```

```
> CAPM #present data
# A tibble: 120 × 28
   Date...1  Close...2 `R(CSCO)` R-Rf(CS…¹ ...5  Date…² Close…³ R(APP…⁴ R-Rf(A…⁵ ...10 Date.…⁶ Close…⁷ `R(VZ)`
   <chr>         <dbl>     <dbl>     <dbl> <lgl> <chr>    <dbl>   <dbl>    <dbl> <lgl> <chr>     <dbl>   <dbl>
 1 NA           NA        NA        NA     NA    1990/1…  0.304 NA       NA       NA    1990/1…    22.0 NA
 2 NA           NA        NA        NA     NA    1990/2…  0.304  0       -0.00709 NA    1990/2…    20.2 -0.0806
 3 1990/3/1     0.0816    NA        NA     NA    1990/3…  0.359  0.184    0.177   NA    1990/3…    21.2  0.0501
 4 1990/4/1     0.0825    0.0106    0.00310 NA   1990/4…  0.352 -0.0217  -0.0293  NA    1990/4…    21.0 -0.00927
 5 1990/5/1     0.0859    0.0421    0.0349 NA    1990/5…  0.368  0.0476   0.0405  NA    1990/5…    23.4  0.115
 6 1990/6/1     0.0920    0.0707    0.0637 NA    1990/6…  0.400  0.0848   0.0778  NA    1990/6…    22.3 -0.0480
 7 1990/7/1     0.0885   -0.0377   -0.0447 NA    1990/7…  0.375 -0.0615  -0.0684  NA    1990/7…    22.2 -0.00504
 8 1990/8/1     0.0859   -0.0294   -0.0368 NA    1990/8…  0.330 -0.119   -0.126   NA    1990/8…    19.2 -0.134
 9 1990/9/1     0.0781   -0.0909   -0.0983 NA    1990/9…  0.259 -0.216   -0.224   NA    1990/9…    21.2  0.105
10 1990/10/1    0.102     0.311     0.304  NA    1990/1…  0.275  0.0603   0.0531  NA    1990/1…    23.5  0.106
# … with 110 more rows, 15 more variables: `R-Rf(VZ)` <dbl>, ...15 <lgl>, Date...16 <chr>, Close...17 <dbl>,
#   `R(HD)` <dbl>, `R-Rf(HD)` <dbl>, ...20 <lgl>, Date...21 <chr>, Close...22 <dbl>, Rf <dbl>, ...24 <lgl>,
#   Date...25 <chr>, Close...26 <dbl>, Rm <dbl>, `Rm - Rf` <dbl>, and abbreviated variable names ¹`R-Rf(CSCO)`,
#   ²Date...6, ³Close...7, ⁴`R(APPLE)`, ⁵`R-Rf(APPLE)`, ⁶Date...11, ⁷Close...12
> summary(CAPM) #present information about data
   Date...1           Close...2          R(CSCO)           R-Rf(CSCO)          ...5            Date...6
 Length:120         Min.   : 0.07812   Min.   :-0.20251   Min.   :-0.20795   Mode:logical   Length:120
 Class :character   1st Qu.: 0.73611   1st Qu.:-0.03043   1st Qu.:-0.03577   NA's:120       Class :character
 Mode  :character   Median : 2.03125   Median : 0.05542   Median : 0.04972                  Mode  :character
                    Mean   : 7.04140   Mean   : 0.06354   Mean   : 0.05804
                    3rd Qu.: 8.02604   3rd Qu.: 0.13396   3rd Qu.: 0.12877
                    Max.   :53.56250   Max.   : 0.33898   Max.   : 0.33279
                    NA's   :2          NA's   :3          NA's   :3
   Close...7         R(APPLE)           R-Rf(APPLE)          ...10           Date...11            Close...12
 Min.   :0.1172   Min.   :-0.302427   Min.   :-0.307260   Mode:logical   Length:120         Min.   :18.66
 1st Qu.:0.2455   1st Qu.:-0.078239   1st Qu.:-0.083440   NA's:120       Class :character   1st Qu.:22.47
 Median :0.3320   Median :-0.002873   Median :-0.007963                  Mode  :character   Median :25.63
 Mean   :0.3489   Mean   : 0.019079   Mean   : 0.013554                                     Mean   :30.08
 3rd Qu.:0.4180   3rd Qu.: 0.107857   3rd Qu.: 0.102495                                     3rd Qu.:32.57
 Max.   :0.9180   Max.   : 0.395228   Max.   : 0.390635                                     Max.   :60.54
                  NA's   :1           NA's   :1
```

```
     R(VZ)               R-Rf(VZ)             ...15            Date...16            Close...17          R(HD)
 Min.   :-0.134177   Min.   :-0.141561   Mode:logical   Length:120          Min.   : 1.735   Min.   :-0.11039
 1st Qu.:-0.033844   1st Qu.:-0.039616   NA's:120       Class :character    1st Qu.: 8.170   1st Qu.:-0.03042
 Median : 0.002652   Median :-0.002753                  Mode  :character    Median :10.000   Median : 0.03067
 Mean   : 0.009519   Mean   : 0.003993                                      Mean   :14.365   Mean   : 0.03442
 3rd Qu.: 0.044699   3rd Qu.: 0.040148                                      3rd Qu.:15.432   3rd Qu.: 0.08372
 Max.   : 0.194064   Max.   : 0.189222                                      Max.   :68.750   Max.   : 0.30229
 NA's   :1           NA's   :1                                                               NA's   :1
    R-Rf(HD)           ...20            Date...21            Close...22          Rf                ...24
 Min.   :-0.11736   Mode:logical   Length:120          Min.   :4.410   Min.   :0.003675   Mode:logical
 1st Qu.:-0.03564   NA's:120       Class :character    1st Qu.:5.827   1st Qu.:0.004856   NA's:120
 Median : 0.02578                  Mode  :character    Median :6.503   Median :0.005419
 Mean   : 0.02890                                      Mean   :6.646   Mean   :0.005538
 3rd Qu.: 0.07820                                      3rd Qu.:7.343   3rd Qu.:0.006119
 Max.   : 0.29693                                      Max.   :9.040   Max.   :0.007533
 NA's   :1
  Date...25            Close...26           Rm                 Rm - Rf
 Length:120          Min.   : 2834   Min.   :-0.15686   Min.   :-0.16106
```

```
 Length:120          Min.   : 2834   Min.   :-0.15686   Min.   :-0.16106
 Class :character    1st Qu.: 4038   1st Qu.:-0.01190   1st Qu.:-0.01765
 Mode  :character    Median : 4752   Median : 0.01541   Median : 0.01027
                     Mean   : 6271   Mean   : 0.01322   Mean   : 0.00769
                     3rd Qu.: 8468   3rd Qu.: 0.03730   3rd Qu.: 0.03229
                     Max.   :13813   Max.   : 0.10715   Max.   : 0.10156
                                     NA's   :1          NA's   :1
>
```

b) **What your findings / interpretations are on E(R), the estimates**

Construct	HD	AAPL	VZ	CSCO
E(R): average return	0.03442089	0.01907947	0.009518852	0.06354
$\hat{\alpha}$	0.020720	0.00473	-0.0002613	0.046562
P value for $Ho: \alpha = 0$	0.00171	0.707954	0.96	5.00e-06
$\hat{\beta}$	1.063014	1.14730	0.5532043	1.500948
P value for $H1: \beta = 0$	1.97e-09	0.000472	5.47e-05	1.16e-08
R^2	0.2658	0.09961	0.1304	0.2474

Home Depot –

HD: Average Return = 0.03442089

```
> average_return_HD = mean(na.omit(CAPM$`R(HD)`)) #calculate the average return of Home Depot
> average_return_HD
[1] 0.03442089
```

R-R_f = 0.020720 + 1.063014($R_m - R_f$)

Interpretation: the original value of (R-R_f) is 0.020720, and each additional ($R_m - R_f$) is expected to increase (R - R_f) 1.063014

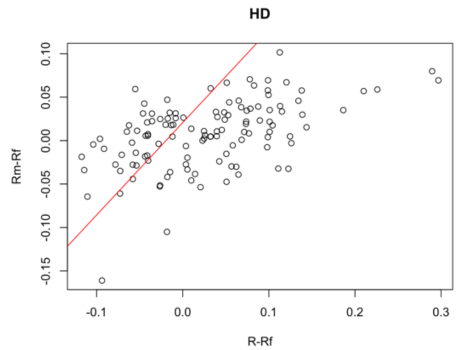
R^2: we can explain almost 26.58% of the variation in (R-R_f) with ($R_m - R_f$)

```
Call:
lm(formula = CAPM$`R-Rf(HD)` ~ CAPM$`Rm - Rf`, data = CAPM)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.138914 -0.054650  0.002416  0.046462  0.202442 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)    0.020720   0.006452   3.211  0.00171 ** 
CAPM$`Rm - Rf` 1.063014   0.163336   6.508 1.97e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06904 on 117 degrees of freedom
  (因为不存在，1个观察量被删除了)
Multiple R-squared:  0.2658,	Adjusted R-squared:  0.2595 
F-statistic: 42.36 on 1 and 117 DF,  p-value: 1.974e-09
```




```
fit_HD = lm(CAPM$`R-Rf(HD)`~CAPM$`Rm - Rf`,data=CAPM) # fit the HD data to a linear model
summary(fit_HD) # present the information about the linear model(fit_HD)'s result
plot(CAPM$`R-Rf(HD)`,CAPM$`Rm - Rf`,xlab='R-Rf',ylab='Rm-Rf',main='HD')# create scatter plot for linear model
abline(lm(CAPM$`R-Rf(HD)`~CAPM$`Rm - Rf`),col="red") #add a red-color regression line to the scatter plot
```

Apple-

AAPL: Average Return = 0.01907947

```
> average_return_APPLE = mean(na.omit(CAPM$`R(APPLE)`))#calculate the average return of Apple
> average_return_APPLE
[1] 0.01907947
```

R-R_f = 0.00473 +1.14730$(R_m - R_f)$

Interpretation: the original value of (R-R_f) is 0.00473, and each additional $(R_m - R_f)$ is expected to increase (R - R_f) 1.14730

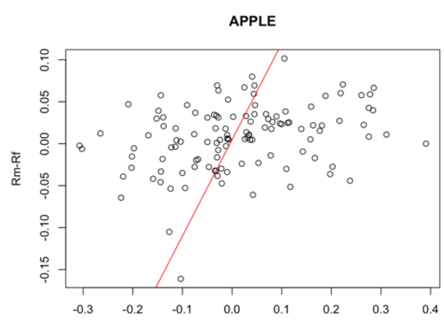
R^2: we can explain almost 9.961% of the variation in (R-R_f) with $(R_m - R_f)$

```
lm(formula = CAPM$`R-Rf(APPLE)` ~ CAPM$`Rm - Rf`, data = CAPM)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.30929 -0.08529 -0.01373  0.07772  0.38592 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)     0.00473    0.01260   0.376 0.707954    
CAPM$`Rm - Rf`  1.14730    0.31889   3.598 0.000472 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1348 on 117 degrees of freedom
  (因为不存在，1个观察量被删除了)
Multiple R-squared:  0.09961,	Adjusted R-squared:  0.09192 
F-statistic: 12.94 on 1 and 117 DF,  p-value: 0.0004715
```




```
fit_APPLE = lm(CAPM$`R-Rf(APPLE)`~CAPM$`Rm - Rf`,data=CAPM) #fit the Apple data to a linear model
summary(fit_APPLE) #present the information about the linear model(fit_APPLE)'s result
plot(CAPM$`R-Rf(APPLE)`,CAPM$`Rm - Rf`,xlab='R-Rf',ylab='Rm-Rf',main='APPLE')#create scatter plot for model
abline(lm(CAPM$`R-Rf(APPLE)`~CAPM$`Rm - Rf`),col="red")#add a red-color regression line to the scatter plot
```

Verizon-

VZ: Average Return = 0.009518852

```
> average_return_VZ = mean(na.omit(CAPM$`R(VZ)`)) # calculate the average return of Verizon
> average_return_VZ
[1] 0.009518852
```

R-R_f = -0.0002613 +0.5532043$(R_m - R_f)$

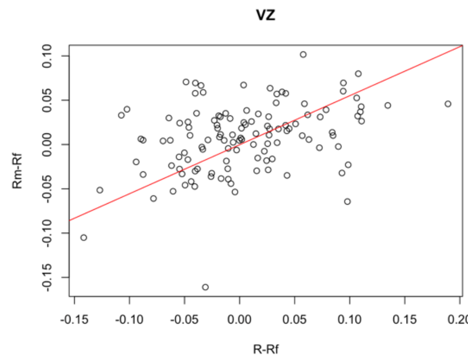
Interpretation: the original value of (R-R_f) is -0.0002613, and each additional $(R_m - R_f)$ is expected to increase (R - R_f) 0.5532043

R^2: we can explain almost 13.04% of the variation in (R-R_f) with $(R_m - R_f)$

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.125389 -0.034083 -0.002845  0.034160  0.164132

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.0002613  0.0052174  -0.050     0.96
CAPM$`Rm - Rf`  0.5532043  0.1320741   4.189 5.47e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05583 on 117 degrees of freedom
  (因为不存在，1个观察量被删除了)
Multiple R-squared:  0.1304,	Adjusted R-squared:  0.123
F-statistic: 17.54 on 1 and 117 DF,  p-value: 5.469e-05
```




```
fit_VZ = lm(CAPM$`R-Rf(VZ)`~CAPM$`Rm - Rf`,data=CAPM) #fit the VZ data to a linear model
summary(fit_VZ)#present the information about the linear model(fit_VZ)'s result
plot(CAPM$`R-Rf(VZ)`,CAPM$`Rm - Rf`,xlab='R-Rf',ylab='Rm-Rf',main='CSCO')#create scatter plot for linear model
abline(lm(CAPM$`R-Rf(VZ)`~CAPM$`Rm - Rf`),col="red")#add a red-color regression line to the scatter plot
```

Cisco –

CSCO : Average Return = 0.06354

```
> average_return_CSCO = mean(na.omit(CAPM$`R(CSCO)`)) #calculate the average return of Cisco
> average_return_CSCO
[1] 0.06354
```

R-R_f=0.046562 -1.500948 $(R_m - R_f)$

Interpretation: the original value of (R-R_f) is 0.046562, and each additional $(R_m - R_f)$ is expected to increase (R - R_f) 1.500948

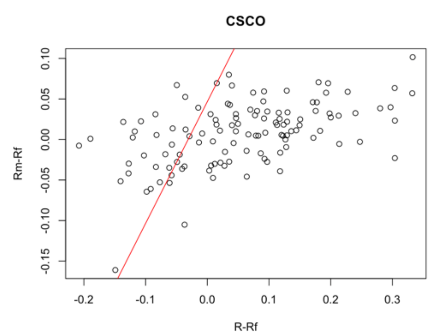
R^2:we can explain almost 24.74% of the variation in (R-R_f) with $(R_m - R_f)$

```
lm(formula = CAPM$`R-Rf(CSCO)` ~ CAPM$`Rm - Rf`, data = CAPM)

Residuals:
      Min        1Q    Median        3Q       Max
-0.243099 -0.063268  0.006474  0.069272  0.291806

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.046562   0.009719   4.791 5.00e-06 ***
CAPM$`Rm - Rf` 1.500948   0.244120   6.148 1.16e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1032 on 115 degrees of freedom
  (因为不存在，3个观察量被删除了)
Multiple R-squared:  0.2474,	Adjusted R-squared:  0.2409
F-statistic:  37.8 on 1 and 115 DF,  p-value: 1.164e-08
```




```
fit_CSCO = lm(CAPM$`R-Rf(CSCO)`~CAPM$`Rm - Rf`,data=CAPM) #fit the Cisco data to a linear model
summary(fit_CSCO) #present the information about the linear model(fit_CSCO)'s result
plot(CAPM$`R-Rf(CSCO)`,CAPM$`Rm - Rf`,xlab='R-Rf',ylab='Rm-Rf',main='CSCO') #create scatter plot for linear mode
abline(lm(CAPM$`R-Rf(CSCO)`~CAPM$`Rm - Rf`),col="red")#add a red-color regression line to the scatter plot
```

c) Whether or not you reject the null hypotheses H_0: α = 0 and $H_{0:}\beta$ and at which significance level

Assume the significant level is 0.05-

HD : $H_0: \alpha = 0$ $\qquad$ $H_1: \alpha \neq 0$

P Value equal to 0.00171 < 0.05, we reject the null at 0.05 level, we conclude α is statistically significant.

$H_{0:}\beta = 0$ $H_1: \beta \neq 0$

P Value equal to 1.97e-09 < 0.05, we reject the null at 0.05 level, we conclude β is statistically significant.

AAPL: $H_0: \alpha = 0$ $H_1: \alpha \neq 0$

P Value equal to 0.707954 >0.05, we cannot reject the null at 0.05 level, we conclude α is not statistically significant.

$H_{0:}\beta = 0$ $H_1: \beta \neq 0$

P Value equal to 0.000472 < 0.05, we reject the null at 0.05 level, we conclude β is statistically significant.

VZ: $H_0: \alpha = 0$ $H_1: \alpha \neq 0$

P Value equal to 0.96>0.05, we cannot reject the null at 0.05 level, we conclude α is not statistically significant.

$H_{0:}\beta = 0$ $H_1: \beta \neq 0$

P Value equal to 5.47e-05< 0.05, we reject the null at 0.05 level, we conclude β is statistically significant.

CSCO: $H_0: \alpha = 0$ $H_1: \alpha \neq 0$

P Value equal to 5.00e-06 < 0.05, we reject the null at 0.05 level, we conclude α is statistically significant.

$H_{0:}\beta = 0$ $H_1: \beta \neq 0$

P Value equal to 1.16e-08 < 0.05, we reject the null at 0.05 level, we conclude β is statistically significant.

2 Computing Price Elasticities (and optimal prices)

Using the data posted on the class website (rfj_data.xlsx) run three simple regressions of the following type

$$\boldsymbol{Q_t = \alpha + \beta p_t + u_t}$$

for each of the three brands in the data (Tropicana, Minute Maid and Private label). In this regression, Q_t (ounces sold) is the dependent variable, p_t (price) is the independent variable, and u_t represents the residual (or prediction error) term. Also, α and β are the intercept and the slope parameters, respectively, in the regression model.

Data-Preprocessing:

1) Cleaning the existing data to avoid data/variable obfuscation.
2) Load the package for future use, 'readxl' contains a method for loading excel data.

```
#cleaning data
rm(list=ls());
gc();
```

```
#Load package
library(readxl)
```

3) Load data to Rstudio and present the data.

```
#Load data (rfj_data.xlsx) to csv and present the data
rfj <- read_excel('rfj_data.xlsx',sheet = 'rfj')
head(rfj)#present first 6 rows of data
summary(rfj[c(2,3,4,5,6,7)])#present information of data
```

```
> head(rfj)#present data
# A tibble: 6 × 7
   WEEK      q1      q2      q3     p1     p2     p3
  <dbl>   <dbl>   <dbl>   <dbl>  <dbl>  <dbl>  <dbl>
1     1 3592320 1363666 1316800 0.0391 0.0328 0.0249
2     2 2310768 3513756 1060416 0.0412 0.0307 0.0263
3     3 5076400 1335114 1013696 0.0344 0.0325 0.0260
```

```
> summary(rfj[c(2,3,4,5,6,7)])#present information of data
       q1                 q2                 q3                 p1                p2                p3         
 Min.   :  294464   Min.   :  858466   Min.   :  753600   Min.   :0.02459   Min.   :0.01646   Min.   :0.01394  
 1st Qu.: 2538064   1st Qu.: 1296548   1st Qu.: 1447744   1st Qu.:0.03598   1st Qu.:0.03102   1st Qu.:0.02243  
 Median : 3309412   Median : 1597826   Median : 1994592   Median :0.03973   Median :0.03261   Median :0.02413  
 Mean   : 3935819   Mean   : 2477531   Mean   : 2691890   Mean   :0.03913   Mean   :0.03251   Mean   :0.02424  
 3rd Qu.: 4772334   3rd Qu.: 2808464   3rd Qu.: 3034720   3rd Qu.:0.04217   3rd Qu.:0.03537   3rd Qu.:0.02608  
 Max.   :17900608   Max.   :16958464   Max.   :13188800   Max.   :0.04842   Max.   :0.04025   Max.   :0.03557  
```

4) Use the Linear Regression Model find the α(Intercept), β(Slope) and $\boldsymbol{u_t}$(Residuals) for each three companies. Those variables will be used in calculating elasticity(Q2a) and Optimization(Q2b).

Tropicana:

```
fit_1 = lm(q1~p1,data=rfj)#fit the Q and p to linear regression model
summary(fit_1)#present the information of fitting result
intercept_1 = coefficients(summary(fit_1))[1,1] #store the intercept
slope_1=coefficients(summary(fit_1))[2,1]#store the slope
# scatter plot: quantity versus prices
plot(rfj$q1,rfj$p1,xlab='Quantity',ylab='Price',main='Tropicana')
#add a blue-color regression line to the existing plot
abline(lm(rfj$p1~rfj$q1), col="blue")
```

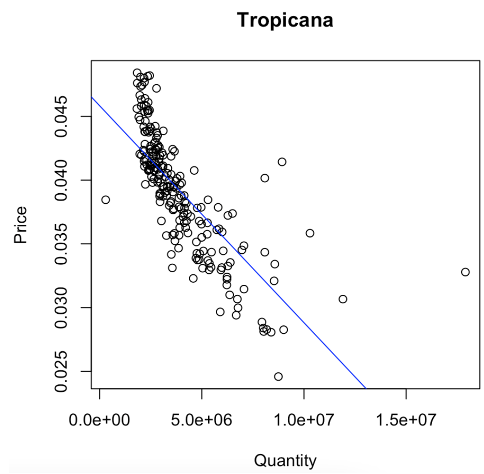
```
> summary(fit_1)#present the information of fitting result

Call:
lm(formula = q1 ~ p1, data = rfj)

Residuals:
     Min       1Q   Median       3Q      Max 
-3865440  -720429  -170082   402385 11860966 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)   16908578     781684   21.63   <2e-16 ***
p1          -331498553   19833155  -16.71   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1372000 on 216 degrees of freedom
Multiple R-squared:  0.564,	Adjusted R-squared:  0.5619 
F-statistic: 279.4 on 1 and 216 DF,  p-value: < 2.2e-16
```



For brand Tropicana, intercept is 16908578, slope is -331498553, residual is -170082(Median). The slope (beta1) of the regression line is -331498553 which means that the expected value of ounces sold will decrease by 331498553 if the price of Tropicana increases by one dollar. The absolute value of beta1(slope) is large, meaning that there will be a big influence on the dependent variable (ounces sold), so it is economically significant. The p-value is small, and the t-value is big, so we can reject the null hypothesis. The residual is -170082 which means the predicted value is too high (a negative residual). In other words, it represents the difference between the observed value and the fitted response value which is 170082 deviates from the model. R square is 0.564 which means that 56.4% of the variation in quantity is explained by the price in the regression model.

Minute Maid:

```
fit_2 = lm(q2~p2,data=rfj)#fit the Q and p to linear regression model
summary(fit_2)#present the information of fitting result
intercept_2 = coefficients(summary(fit_2))[1,1] #store the intercept
slope_2=coefficients(summary(fit_2))[2,1]#store the slope
# scatter plot: quantity versus prices
plot(rfj$q2,rfj$p2,xlab='Quantity',ylab='Price',main='Minute Maid')
#add a blue-color regression line to the existing plot
abline(lm(rfj$p2~rfj$q2), col="blue")
```

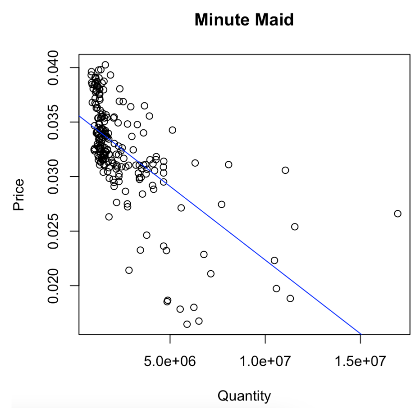
```
> summary(fit_2)#present the information of fitting result

Call:
lm(formula = q2 ~ p2, data = rfj)

Residuals:
     Min       1Q   Median       3Q      Max 
-2943719  -948760  -373986   496154 12718311 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)   12171921     811113   15.01   <2e-16 ***
p2          -298175734   24703396  -12.07   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1672000 on 216 degrees of freedom
Multiple R-squared:  0.4028,	Adjusted R-squared:    0.4 
F-statistic: 145.7 on 1 and 216 DF,  p-value: < 2.2e-16
```



For brand Minute Maid, intercept is 12171921, slope is -298175734, residual is -373986(Median). The slope (beta1) of the regression line is -298175734 which means that the expected value of ounces sold will decrease by 298175734 if the price of Minute Maid increases by one dollar. Comparing the absolute value of beta1(slope) to Tropicana, the value is smaller, meaning Minute Maid is less economically significant than Tropicana. The p-value is small, and the t-value is big, so we can reject the null hypothesis. The residual is -373986 which means the predicted value is too high (a negative residual). In other words, it represents the difference between the observed value and the fitted response value which is 373986 deviates from the model. R square is 0.4028 which means that 40.28% of the variation in quantity is explained by the price in the regression model.

Private Label:

```
fit_3 = lm(q3~p3,data=rfj)#fit the Q and p to linear regression model
summary(fit_3)#present the information of fitting result
intercept_3 = coefficients(summary(fit_3))[1,1] #store the intercept
slope_3=coefficients(summary(fit_3))[2,1]#store the slope
# scatter plot: quantity versus prices
plot(rfj$q3,rfj$p3,xlab='Quantity',ylab='Price',main='Private Label')
#add a blue-color regression line to the existing plot
abline(lm(rfj$p3~rfj$q3), col="blue")
```

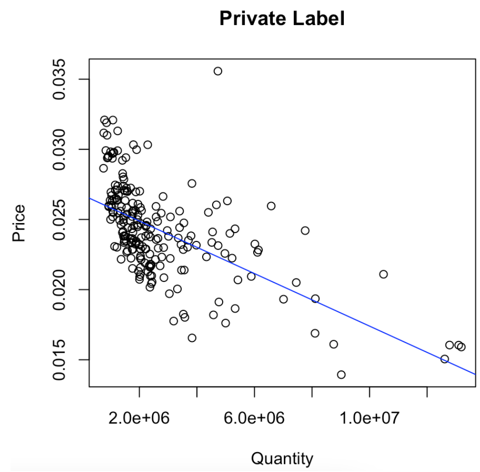
```
> summary(fit_3)#present the information of fitting result

Call:
lm(formula = q3 ~ p3, data = rfj)

Residuals:
     Min       1Q   Median       3Q      Max
-2140390 -1143199  -415911   531089  7234901

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   12182980     857012   14.22   <2e-16 ***
p3          -391515424   35014574  -11.18   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1745000 on 216 degrees of freedom
Multiple R-squared:  0.3666,	Adjusted R-squared:  0.3637
F-statistic:   125 on 1 and 216 DF,  p-value: < 2.2e-16
```



For brand Private Label intercept is 12182980, slope is -391515424, residual is -415911(Median). The slope (beta1) of the regression line is -391515424 which means that the expected value of ounces sold will decrease by 391515424 if the price of Private Label increases by one dollar. The absolute value of beta1(slope) is the largest, meaning that the price of Private Label is both statically and economically significant. The residual is -415911 which means the predicted value is too high (a negative residual). In other words, it represents the difference between the observed value and the fitted response value which is 415911 deviates from the model. R square is 0.3666 which means that 36.66% of the variation in quantity is explained by the price in the regression model.

(a) Using the regression results, compute own price elasticity at (average P, average Q) for each brand. Price elasticity, ε, can be expressed as:

$$\varepsilon = \frac{\partial Q}{\partial p}\frac{p}{Q}$$

where $\frac{\partial Q}{\partial p}$ is the slope coefficient from the regression result.

```
elasticity_1 = slope_1 * (mean(rfj$p1)/mean(rfj$q1))
elasticity_2 = slope_2 * (mean(rfj$p2)/mean(rfj$q2))
elasticity_3 = slope_3 * (mean(rfj$p3)/mean(rfj$q3))
```

```
> print(c(elasticity_1,elasticity_2,elasticity_3))#present 3 elasticity
[1] -3.296077 -3.912924 -3.525809
```

#The elasticity of Tropicana is -3.296, which means that the relationship between price and ounces sold is sensitive. There is a 3.296 unit decrease in ounces sold to respond to a unit increase in price.

#The elasticity of Minute Maid is -3.913, which means that the relationship between price and ounces sold is sensitive. There is a 3.913 unit decrease in ounces sold to respond to a unit increase in price.

#The elasticity of Private Label is -3.526, which means that the relationship between price and ounces sold is sensitive. There is a 3.526 unit decrease in ounces sold to respond to a unit increase in price. However, the elasticity of Private Label is in the middle of the three brands, meaning that the consumers of Private Label are more care about the price compared to Tropicana rather than Minute Maid.

(b) Assume that the per unit cost of producing an additional unit is the same across brands and is 1 cent per ounce. Based on the data provided and the estimated demand equation from above, compute the optimal price for each brand (that is, find the price point at which the profit is maximized). For this problem, you can ignore the error term u_t when writing down the profit function. The attached "intro to solver.pdf" or "intro to optimization in R.pdf" should prove useful for setting up this problem.

Since we can ignore error term $\boldsymbol{u_t}$ and the marginal cost c = 1 cent = 0.01 $
Tropicana:
Demand function is $Q = 16908578 + -331498553 \times p$
Profit function is $\pi = (p-c)Q = (p-c)(16908578 + -331498553 \times p)$

```
#calculate optimal profit for Tropicana
profit_function_1 = function(price,intercept_1,slope_1){
    #create function for profit, take price,intercept,slope as parameters
    profit = (price-0.01)*(intercept_1 + slope_1 *price)
    return(profit)}#return profit as output
#optimize the function for maximum the profit, set interval to (0,1000)
optimize(profit_function_1,interval =c(0,1000),maximum=TRUE,intercept_1,slope_1)
```

```
$maximum
[1] 0.03050325

$objective
[1] 139356.4
```

Price of maximized profit for Tropicana is 0.0305, and the optimized profit is 139356.4 dollars.

Minute Maid:
Demand function is $Q = 12171921 + -298175734 \times p$
Profit function is $\pi = (p-c)Q = (p-c)(12171921 + -298175734 \times p)$

```
#calculate optimal profit for Minute Maid
profit_function_2 = function(price,intercept_2,slope_2){
    #create function for profit, take price,intercept,slope as parameters
    profit = (price-0.01)*(intercept_2 + slope_2 *price)
    return(profit)}#return profit as output
#optimize the function for maximum the profit, set interval to (0,1000)
optimize(profit_function_2,interval =c(0,1000),maximum=TRUE,intercept_2,slope_2)
```

```
$maximum
[1] 0.02541065

$objective
[1] 70813.2
```

Price of maximized profit for Minute Maid is 0.025, and the optimized profit is 70813.2 dollars.

Private Label:
Demand function is $Q = 12182980 + -391515424 \times p$
Profit function is $\pi = (p-c)Q = (p-c)(12182980 + -391515424 \times p)$

```
#calculate optimal profit for Private Label
profit_function_3 = function(price,intercept_3,slope_3){
    #create function for profit, take price,intercept,slope as parameters
    profit = (price-0.01)*(intercept_3 + slope_3 *price)
    return(profit)}#return profit as output
#optimize the function for maximum the profit, set interval to (0,1000)
optimize(profit_function_3,interval =c(0,1000),maximum=TRUE,intercept_3,slope_3)
```

```
$maximum
[1] 0.02055875

$objective
[1] 43648.95
```

Price of maximized profit for Private Label is 0.021, and the optimized profit is 43648.95 dollars.

In general, the price of maximized profit of each three brands is all lower than average price respectively. Because of the fluctuating value of price, it causes inaccuracy.