Background:

Choose 3 lists of 36 stocks on the TSX from
http://clouddc.chass.utoronto.ca.ezproxy.library.yorku.ca/ds/cfmrc/

Pick 36 stocks based on My name- YA JUN BAI

- 36 stocks whose symbols begin with Y or A (call this the 1st group from 2013)
- 36 stocks whose symbols begin with J or U (call this the 2nd group from 2014)
- 36 stocks whose symbols begin with B or A (call this the 3rd group from 2015)

and get an end of day price for ALL of them on the same date (April 9).

Analyse (1)

Introduction:

1. We are using SAS to estimate a multiple regression model, $y = f(x_1, x_2)$,

$$Y= B_0+B_1X_1+B_2X_2$$

y- Stock price of 2015

x_1- Stock price of 2013

x_2- Stock price of 2014

and justify the scientific significance of the variable(s) and the estimated model with t test, F test, R-squared, adjusted R squared. For each test, we identify the null and alternative hypotheses, the statistic value, and the p-value and conclusion.

2. Data collection:

We collected the data of stock price from 2013 to 2015,
Group1- Stock price of 2013
Group2- Stock price of 2014
Group3- Stock price of 2015

3. Report

Obs	Price2015	Price2013	Price2014
1	26.94	26.55	26.74
2	0.28	1.48	0.57
3	25.32	26.04	21.30
4	25.56	26.28	25.16
5	26.56	26.27	23.41
6	24.60	26.51	25.33
7	22.82	25.73	25.00
8	23.25	24.80	21.03
9	23.74	24.95	24.70
10	15.20	23.06	20.35

Obs	Price2015	Price2013	Price2014
11	19.75	17.86	17.13
12	19.80	24.54	20.82
13	22.72	24.87	20.90
14	25.70	26.65	25.63
15	19.84	26.10	24.25
16	17.09	26.01	21.43
17	24.15	27.00	25.82
18	12.01	15.10	8.66
19	2.71	4.11	4.08
20	2.60	4.08	4.05
21	9.00	15.35	13.76
22	19.25	25.31	22.17
23	9.17	15.25	13.18
24	9.56	10.45	11.00
25	10.31	10.57	10.32
26	17.00	24.00	20.72
27	17.03	23.83	21.40
28	17.10	24.25	20.71
29	17.00	23.80	21.82
30	17.02	23.55	21.64
31	16.60	24.68	21.52
32	16.72	24.85	21.79
33	16.98	23.35	21.60
34	16.75	24.60	21.50
35	17.27	23.54	21.85
36	17.05	22.25	25.88

“CORRELATION MATRIX”

The CORR Procedure

3 Variables:	Price2015 Price2013 Price2014

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Price2015	36	17.34583	6.84032	624.45000	0.28000	26.94000
Price2013	36	21.32278	7.04951	767.62000	1.48000	27.00000
Price2014	36	19.36722	6.71708	697.22000	0.57000	26.74000

Pearson Correlation Coefficients, N = 36
Prob > |r| under H0: Rho=0

	Price2015	Price2013	Price2014
Price2015	1.00000	0.90172 <.0001	0.90158 <.0001
Price2013	0.90172 <.0001	1.00000	0.96529 <.0001
Price2014	0.90158 <.0001	0.96529 <.0001	1.00000

“CORRELATION MATRIX”

The REG Procedure
Model: MODEL1
Dependent Variable: Price2015

Number of Observations Read	36
Number of Observations Used	36

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1354.87858	677.43929	79.06	<.0001
Error	33	282.77029	8.56880		
Corrected Total	35	1637.64888			

Root MSE	2.92725	R-Square	0.8273
Dependent Mean	17.34583	Adj R-Sq	0.8169
Coeff Var	16.87581		

Parameter Estimates

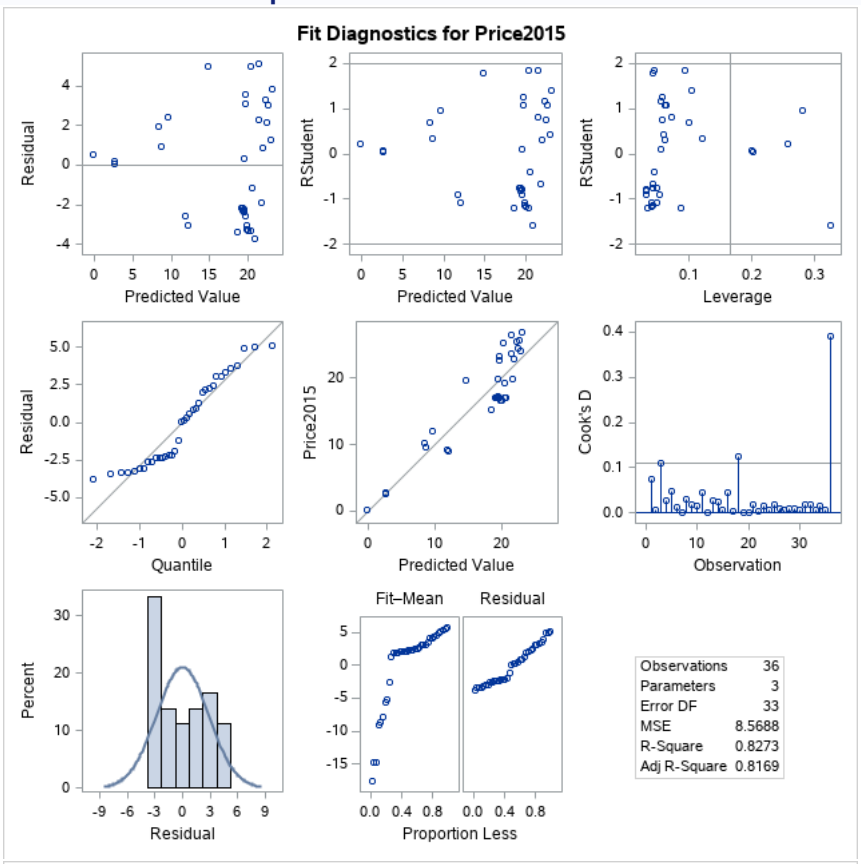
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-1.19682	1.57564	-0.76	0.4529
Price2013	1	0.44703	0.26872	1.66	0.1057
Price2014	1	0.46525	0.28202	1.65	0.1085

“CORRELATION MATRIX”

The REG Procedure

Model: MODEL1

Dependent Variable: Price2015



F test:

H_0: $B_1=B_2=0$ H_a: At least one of B_1 and B_2 is not 0

From the **Analysis of Variance**, we can see
F_{stat}= MSE/MSR = 677.43929/8.56880 = 79.05
Global F test (P-value < .0001) indicates that model is significant for predicting Price2015 based on a group of independent variables in the model
P-value < 0.05, we reject H_0.

T test: each individual variable

From the **Parameter Estimates**, we can see
H_0: $B_1 = 0$ H_a: $B_1 \neq 0$
t_{stat} = 1.66
P-value of Price2013 = 0.1057
P-value>0.05, we failed to reject H_0, therefore, Price2013 is not very useful to predict Price2015

H_0: $B_2 = 0$ H_a: $B_2 \neq 0$
t_{stat} = 1.65
P-value of Price2014 = 0.1085
P-value >0.05, we failed to reject H_0, therefore, Price2014 is not very useful to predict Price2015

R-squared:

From **Sum of Squares**, we can see

R^2=SSR/SST

=1354.87858 / 1637.64888

=0.82733.
which means approximately 83% of the variation of Price2015 is explained by the independent variables.
R-squared *is a statistical measure of how close the data are to the fitted regression line.0% indicates that the model explains none of the variability of the response data around its* ***mean****. 100% indicates that the model explains all the variability of the response data around its* ***mean****.*

Adjusted R squared

From **Parameter Estimate**, we can see

t_{state}= β_1/ s β_1
= 0.44703 / 0.26872
= 1.663
From parameter estimates, we can get the regression model
$Y= -1.19682 + 0.44703X_1+0.46525X_2$

Code:

```
data Stockprice;
input Price2015 Price2013 Price2014;
datalines;
26.94 26.55 26.74
 0.28 1.48 0.57
 25.32 26.04 21.30
 25.56 26.28 25.16
 26.56 26.27 23.41
 24.60 26.51 25.33
 22.82 25.73 25.00
 23.25 24.8 21.03
 23.74 24.95 24.70
 15.20 23.06 20.35
 19.75 17.86 17.13
 19.80 24.54 20.82
 22.72 24.87 20.90
 25.70 26.65 25.63
 19.84 26.1 24.25
 17.09 26.01 21.43
 24.15 27 25.82
 12.01 15.1 8.66
 2.71 4.11 4.08
 2.60 4.08 4.05
 9.00 15.35 13.76
 19.25 25.31 22.17
 9.17 15.25 13.18
 9.56 10.45 11.00
 10.31 10.57 10.32
 17.00 24 20.72
 17.03 23.83 21.40
 17.10 24.25 20.71
 17.00 23.8 21.82
 17.02 23.55 21.64
 16.60 24.68 21.52
 16.72 24.85 21.79
 16.98 23.35 21.60
 16.75 24.6 21.50
 17.27 23.54 21.85
 17.05 22.25 25.88

proc print;
run;

PROC CORR DATA=Stockprice;
TITLE “CORRELATION MATRIX”;
VAR Price2015 Price2013 Price2014;
RUN;
proc reg data=stockprice alpha=0.05;
model Price2015 = Price2013 Price2014;
run;
```

Analyse (2)

Introduction:

1. We are using SAS to estimate a multiple regression model, $\log y = f(x_1, x_2)$,

$\log Y = B_0 + B_1X_1 + B_2X_2$

y- Stock price of 2015

x_1- Stock price of 2013

x_2- Stock price of 2014

and justify the scientific significance of the variable(s) and the estimated model with t test, F test, R-squared, adjusted R squared. For each test, we identify the null and alternative hypotheses, the statistic value, and the p-value and conclusion.

2. Data collection:

We collected the data of stock price from 2013 to 2015,
Group1- Stock price of 2013
Group2- Stock price of 2014
Group3- ln value of Stock price of 2015

3. Report

Obs	LnofPrice2015	Price2013	Price2014
1	3.29	26.55	26.74
2	-1.27	1.48	0.57
3	3.23	26.04	21.30
4	3.24	26.28	25.16
5	3.24	26.27	23.41
6	3.20	26.51	25.33
7	3.13	25.73	25.00
8	3.15	24.80	21.03
9	3.17	24.95	24.70
10	2.72	23.06	20.35

Obs	LnofPrice2015	Price2013	Price2014
11	2.98	17.86	17.13
12	2.99	24.54	20.82
13	3.12	24.87	20.90
14	3.25	26.65	25.63
15	2.99	26.10	24.25
16	2.84	26.01	21.43
17	3.18	27.00	25.82
18	2.49	15.10	8.66
19	0.99	4.11	4.08
20	0.96	4.08	4.05
21	2.20	15.35	13.76
22	2.96	25.31	22.17
23	2.22	15.25	13.18
24	2.26	10.45	11.00
25	2.33	10.57	10.32
26	2.83	24.00	20.72
27	2.83	23.83	21.40
28	2.84	24.25	20.71
29	2.83	23.80	21.82
30	2.83	23.55	21.64
31	2.81	24.68	21.52
32	2.82	24.85	21.79
33	2.83	23.35	21.60
34	2.82	24.60	21.50
35	2.85	23.54	21.85
36	2.84	22.25	25.88

“CORRELATION MATRIX”

The CORR Procedure

3 Variables:	LnofPrice2015 Price2013 Price2014

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
LnofPrice2015	36	2.66639	0.85904	95.99000	-1.27000	3.29000
Price2013	36	21.32278	7.04951	767.62000	1.48000	27.00000
Price2014	36	19.36722	6.71708	697.22000	0.57000	26.74000

Pearson Correlation Coefficients, N = 36
Prob > |r| under H0: Rho=0

	LnofPrice2015	Price2013	Price2014
LnofPrice2015	1.00000	0.88087 <.0001	0.86576 <.0001
Price2013	0.88087 <.0001	1.00000	0.96529 <.0001
Price2014	0.86576 <.0001	0.96529 <.0001	1.00000

"CORRELATION MATRIX"

The REG Procedure
Model: MODEL1
Dependent Variable: LnofPrice2015

Number of Observations Read	36
Number of Observations Used	36

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	20.13128	10.06564	58.31	<.0001
Error	33	5.69675	0.17263		
Corrected Total	35	25.82803			

Root MSE	0.41549	R-Square	0.7794
Dependent Mean	2.66639	Adj R-Sq	0.7661
Coeff Var	15.58236		

Parameter Estimates

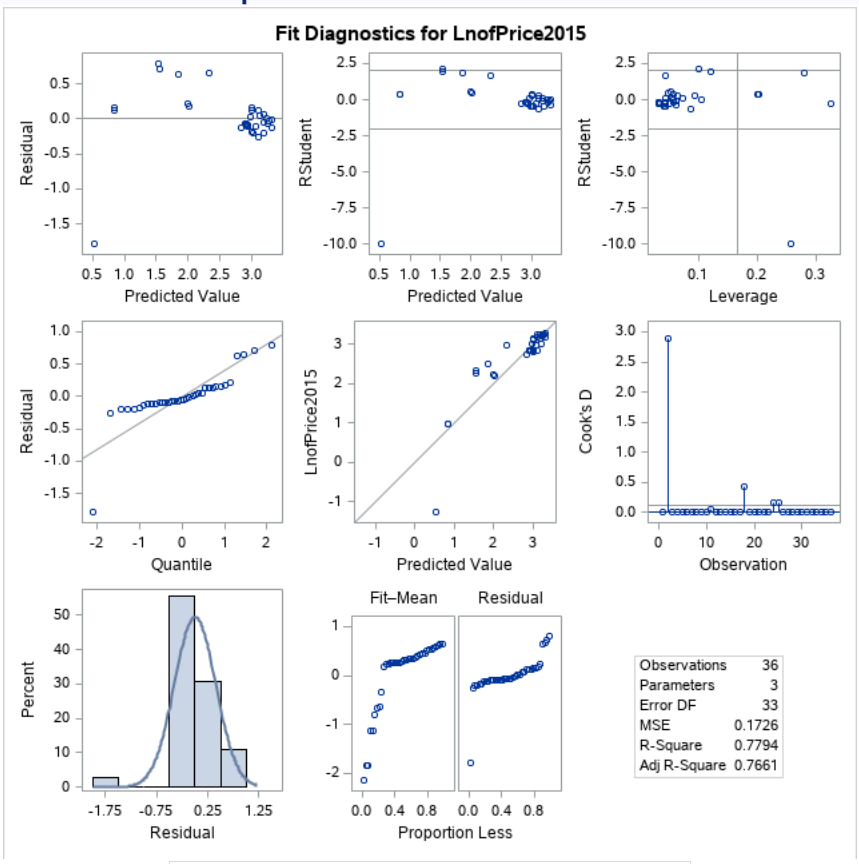
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.38469	0.22364	1.72	0.0948
Price2013	1	0.08066	0.03814	2.11	0.0421
Price2014	1	0.02900	0.04003	0.72	0.4738

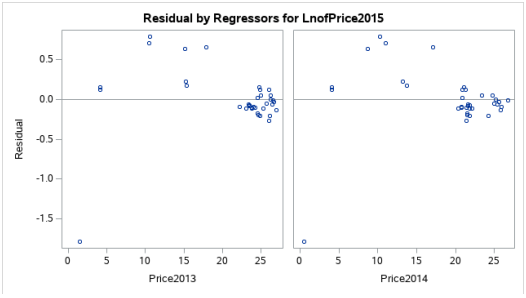
“CORRELATION MATRIX”

The REG Procedure

Model: MODEL1

Dependent Variable: LnofPrice2015





F test:

H_0: $B_1=B_2=0$ H_a: At least one of B_1 and B_2 is not 0

From the **Analysis of Variance**, we can see
F_{stat}= MSE/MSR = 10.06564/0.17263= 58.308
Global F test (P-value < .0001) indicates that model is significant for predicting Price2015 based on a group of independent variables in the model
P-value < 0.05, we reject H_0.

T test: each individual variable

From the **Parameter Estimates**, we can see
H_0: $B_1= 0$ H_a: $B_1 \neq 0$
t_{stat} = 2.11
P-value of Price2013 = 0.0421
P-value of Price2013 < 0.05, we reject H_0, therefore, Price2013 is significant to predict Price2015.

H_0: $B_2=0$ H_a: $B_2 \neq 0$
t_{stat} = 0.72
P-value of Price2014 = 0.4738
P-value of Price2014 >0.05, we failed to reject H_0, therefore, Price2014 is not very useful to predict Price2015.

R-squared:

From **Sum of Squares**, we can see

R^2=SSR/SST=20.13128/ 25.82803 = 0.779435.

which means approximately 78% of the variation of Price2015 is explained by the independent variables.

R-squared *is a statistical measure of how close the data are to the fitted regression line.0% indicates that the model explains none of the variability of the response data around its* ***mean****. 100% indicates that the model explains all the variability of the response data around its* ***mean****.*

Adjusted R squared

From **Parameter Estimate**, we can see

Price2013: $t_{state}= \beta_1 / s\beta_1 =$ 0.08066/ 0.03814 = 2.114

Price2014: $t_{state}= \beta_1 / s\beta_1 =$ 0.02900/ 0.04003 = 0.7196

From parameter estimates, we can get the regression model

$$\log Y= 0.38469+ 0.08066X_1+0.02900X_2$$

Code:

```
data Stockprice;
input LnofPrice2015 Price2013 Price2014;
datalines;

3.29 26.55 26.74
-1.27 1.48 0.57
3.23 26.04 21.30
3.24 26.28 25.16
3.24 26.27 23.41
3.20 26.51 25.33
3.13 25.73 25.00
3.15 24.8 21.03
3.17 24.95 24.70
2.72 23.06 20.35
2.98 17.86 17.13
2.99 24.54 20.82
3.12 24.87 20.90
3.25 26.65 25.63
2.99 26.1 24.25
2.84 26.01 21.43
3.18 27 25.82
2.49 15.1 8.66
0.99 4.11 4.08
0.96 4.08 4.05
2.20 15.35 13.76
2.96 25.31 22.17
2.22 15.25 13.18
2.26 10.45 11.00
2.33 10.57 10.32
2.83 24 20.72
2.83 23.83 21.40
2.84 24.25 20.71
2.83 23.8 21.82
2.83 23.55 21.64
2.81 24.68 21.52
2.82 24.85 21.79
2.83 23.35 21.60
2.82 24.6 21.50
2.85 23.54 21.85
2.84 22.25 25.88

proc print;
run;

PROC CORR DATA=Stockprice;
TITLE “CORRELATION MATRIX”;
VAR LnofPrice2015 Price2013 Price2014;
RUN;

proc reg data=stockprice alpha=0.05;
model LnofPrice2015 = Price2013 Price2014;
run;
```

Comparation of two regression model

P-value

In regression, low p-values indicate terms that are statistically significant.

Items	p-value in F Test	p-value in T Test
1st model	<0.001	P-value of Price2013 = 0.1057 P-value of Price2014 = 0.1085
2nd model	<0.001	P-value of Price2013 = 0.0421 B_1 is important to predict Price2015 P-value of Price2014 = 0.4738 B_2 is not important to predict Price2015

From P-value, we may conclude the 2nd model is better

Adjusted R squared and predicted R-squared values.

Generally, we choose the model that has higher adjusted and predicted R-squared values.

Items	R squared	Adjusted R squared
1st model	0.82733	1.66/1.65
2nd model	0.779435	2.11/0.72

From Adjusted R squared and predicted R-squared values, we may conclude that the 2nd model is better.

All in all, we should conclude that the 2nd model is much better to predict the stock price of 2015