THE CAPITAL ASSET PRICING MODEL

$$R_i = R_f + \beta_i (R_m - R_f)$$

EXAMPLE 2: FIND THE BETA OF GOOGLE (IN R)

EXAMPLE 2: FIND THE BETA OF GOOGLE (IN R)

LET'S GO THROUGH THE STEPS WE'LL NEED TO DO

STEP 1: READ THE HISTORICAL PRICES FOR GOOGLE AND NASDAQ INTO DATAFRAMES

EXAMPLE 2: FIND THE BETA OF GOOGLE (IN R)

THESE ARE THE FILE PATHS FOR THE DATA DOWNLOADED FROM YAHOO FINANCE

```
googFile <- '/Users/swethakolalapudi/Desktop/Regression/goog.csv'
nasdaqFile <-'/Users/swethakolalapudi/Desktop/Regression/nasdaq.csv'
```

READ.TABLE() CAN READ DATA FROM A FILE INTO A DATA FRAME

```
goog <- read.table(googFile,header = TRUE, sep =",")[,c("Date","Adj.Close")]
```

THE FILE PATH

EXAMPLE 2: FIND THE BETA OF GOOGLE (IN R)

READ.TABLE() CAN READ DATA FROM A FILE INTO A DATA FRAME

```
goog <- read.table(googFile,header = TRUE, sep =",")[,c("Date","Adj.Close")]
```

SPECIFYING THAT THE FILE HAS A HEADER AND ITS COMMA SEPARATED

EXAMPLE 2: FIND THE BETA OF GOOGLE (IN R)

WE ARE ONLY KEEPING THE COLUMNS DATE AND ADJ. CLOSE FROM THIS DATA FRAME

```
goog <- read.table(googFile,header = TRUE, sep =",")[,c("Date","Adj.Close")]
```

EXAMPLE 2: FIND THE BETA OF GOOGLE (IN R)

```
goog <- read.table(googFile,header = TRUE, sep =",")[,c("Date","Adj.Close")]
```

```
names(goog)[2]<-"goog.price"
```

RENAME THE SECOND COLUMN TO PRICE

EXAMPLE 2: FIND THE BETA OF GOOGLE (IN R)

```
goog <- read.table(googFile,header = TRUE, sep =",")[,c("Date","Adj.Close")]names(goog)
[2]<-"goog.price"
```

```
nasdaq <- read.table(nasdaqFile,header = TRUE, sep =“,")[,c("Date","Adj.Close")]
names(nasdaq)[2]<-"nasdaq.price"
```

WE READ THE NASDAQ PRICES AS WELL

```
goog <- merge(goog, nasdaq, by = "Date")
```

AND MERGE BOTH INTO 1 DATA FRAME

EXAMPLE 1: FIND THE BETA OF GOOGLE (IN EXCEL)

STEP 2: CONVERT THE PRICES TO RETURNS

STEP 2: CONVERT THE PRICES TO RETURNS

	Date	goog.price	nasdaq.price
1	2010-01-04	264.7058	2147.35
2	2010-02-01	263.1373	2238.26
3	2010-03-01	283.2772	2397.96
4	2010-04-01	262.5879	2461.19
5	2010-05-03	242.5728	2257.04
6	2010-06-01	222.2531	2109.24
7	2010-07-01	242.1832	2254.70
8	2010-08-02	224.7856	2114.03
9	2010-09-01	262.6328	2368.62
10	2010-10-01	306.5440	2507.41
11	2010-11-01	277.5779	2498.23
12	2010-12-01	296.6888	2652.87
13	2011-01-03	299.8806	2700.08
14	2011-02-01	306.3941	2782.27
15	2011-03-01	293.0874	2781.07
16	2011-04-01	271.7787	2873.54
17	2011-05-02	264.2462	2835.30
18	2011-06-01	252.9375	2773.52

WE NOW HAVE A DATA FRAME THAT LOOKS LIKE THIS

	Date	goog.price	nasdaq.price
74	2016-02-01	752.0000	4620.37
73	2016-01-04	742.9500	4613.95
72	2015-12-01	758.8800	5007.41
71	2015-11-02	742.6000	5108.67
70	2015-10-01	710.8100	5053.75
69	2015-09-01	608.4200	4620.16
68	2015-08-03	618.2500	4776.51
67	2015-07-01	625.6100	5128.28
66	2015-06-01	520.5100	4986.87
65	2015-05-01	532.1100	5070.03
64	2015-04-01	537.3400	4941.42
63	2015-03-02	548.0024	4900.88
62	2015-02-02	558.4025	4963.53
61	2015-01-02	534.5225	4635.24
60	2014-12-01	526.4024	4736.05
59	2014-11-03	541.8325	4791.63
58	2014-10-01	559.0825	4630.74
57	2014-09-02	577.3626	4493.39
56	2014-08-01	571.6026	4580.27

THE DATE COLUMN IS READ AS STRINGS

CONVERT TO DATE

```
goog[,c("Date")] <- as.Date(goog[,c("Date")])
```

SORT IN DESCENDING ORDER OF DATE

```
goog <- goog[order(goog$Date, decreasing = TRUE),]
```

	Date	goog.price	nasdaq.price
74	2016-02-01	752.0000	4620.37
73	2016-01-04	742.9500	4613.95
72	2015-12-01	758.8800	5007.41
71	2015-11-02	742.6000	5108.67
70	2015-10-01	710.8100	5053.75
69	2015-09-01	608.4200	4620.16
68	2015-08-03	618.2500	4776.51
67	2015-07-01	625.6100	5128.28
66	2015-06-01	520.5100	4986.87
65	2015-05-01	532.1100	5070.03
64	2015-04-01	537.3400	4941.42
63	2015-03-02	548.0024	4900.88
62	2015-02-02	558.4025	4963.53
61	2015-01-02	534.5225	4635.24
60	2014-12-01	526.4024	4736.05
59	2014-11-03	541.8325	4791.63
58	2014-10-01	559.0825	4630.74
57	2014-09-02	577.3626	4493.39
56	2014-08-01	571.6026	4580.27

RETURN = (NEW PRICE) / OLD PRICE - 1

```
goog[-nrow(goog),-1] <-
                    goog[-nrow(goog),-1]/goog[-1,-1]-1
```

ALL BUT THE LAST ROW OF THE DATA

	Date	goog.price	nasdaq.price
74	2016-02-01	752.0000	4620.37
73	2016-01-04	742.9500	4613.95
72	2015-12-01	758.8800	5007.41
71	2015-11-02	742.6000	5108.67
70	2015-10-01	710.8100	5053.75
69	2015-09-01	608.4200	4620.16
68	2015-08-03	618.2500	4776.51
67	2015-07-01	625.6100	5128.28
66	2015-06-01	520.5100	4986.87
65	2015-05-01	532.1100	5070.03
64	2015-04-01	537.3400	4941.42
63	2015-03-02	548.0024	4900.88
62	2015-02-02	558.4025	4963.53
61	2015-01-02	534.5225	4635.24
60	2014-12-01	526.4024	4736.05
59	2014-11-03	541.8325	4791.63
58	2014-10-01	559.0825	4630.74
57	2014-09-02	577.3626	4493.39
56	2014-08-01	571.6026	4580.27

RETURN = (NEW PRICE) / OLD PRICE − 1

```
goog[-nrow(goog),-1] <-
                    goog[-nrow(goog),-1]/goog[-1,-1]-1
```

ALL BUT THE FIRST ROW OF THE DATA

	Date	goog.returns	nasdaq.returns
74	2016-02-01	0.012181153	1.391416e-03
73	2016-01-04	-0.020991452	-7.857554e-02
72	2015-12-01	0.021923013	-1.982116e-02
71	2015-11-02	0.044723594	1.086716e-02
70	2015-10-01	0.168288383	9.384736e-02
69	2015-09-01	-0.015899744	-3.273302e-02
68	2015-08-03	-0.011764494	-6.859416e-02
67	2015-07-01	0.201917298	2.835640e-02
66	2015-06-01	-0.021799957	-1.640221e-02
65	2015-05-01	-0.009733208	2.602690e-02
64	2015-04-01	-0.019456873	8.271992e-03
63	2015-03-02	-0.018624712	-1.262205e-02
62	2015-02-02	0.044675515	7.082471e-02
61	2015-01-02	0.015425568	-2.128558e-02
60	2014-12-01	-0.028477535	-1.159941e-02
59	2014-11-03	-0.030854232	3.474383e-02
58	2014-10-01	-0.031661368	3.056714e-02
57	2014-09-02	0.010076949	-1.896829e-02
56	2014-08-01	0.000000000	4.817187e-02

```
goog[-nrow(goog),-1] <-
                    goog[-nrow(goog),-1]/goog[-1,-1]-
```

RENAME THE COLUMNS

```
names(goog)[2:3]<-c("goog.returns",
                    "nasdaq.returns"
```

DROP THE LAST ROW

```
goog <- goog[-nrow(goog),]
```

STEP 3: SUBTRACT THE HISTORICAL YIELDS FOR A 5 YR TREASURY BOND (RISK FREE RATE) FROM THE RETURNS

STEP 3: SUBTRACT THE HISTORICAL YIELDS FOR A 5 YR TREASURY BOND (RISK FREE RATE) FROM THE RETURNS

READ THE HISTORICAL YIELDS

```
tbonds <- read.table('/Users/swethakolalapudi/Desktop/
                     Regression/tbonds.csv',header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(tbonds)[2]<-"tbonds.return"
```

MERGE WITH THE RETURNS DATA FRAME

```
tbonds[,c("Date")] <- as.Date(tbonds[,c("Date")])
goog <- merge(goog, tbonds, by="Date")
```

DIVIDE TBOND RETURNS BY 100 TO GET THE %

```
goog$tbonds.return <- goog$tbonds.return/100
```

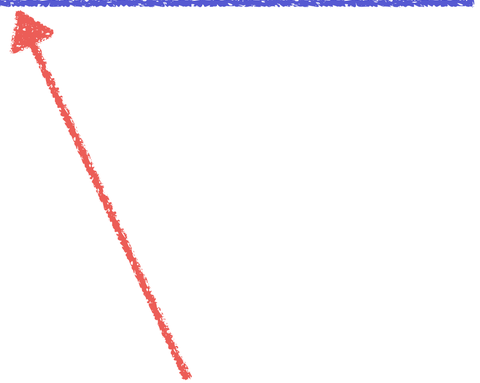
STEP 3: SUBTRACT THE HISTORICAL YIELDS FOR A 5 YR TREASURY BOND (RISK FREE RATE) FROM THE RETURNS

SUBTRACT TBOND RETURNS FROM THE GOOGLE AND NASDAQ RETURNS

```
goog[,c(“goog.returns","nasdaq.returns")] <-
     goog[,c(“goog.returns”,"nasdaq.returns")]-goog[,”tbonds.returns"]
```

STEP 4: REGRESS THE ADJUSTED GOOGLE RETURNS AGAINST NASDAQ RETURNS

```
googM <- lm(goog$goog.returns~goog$nasdaq.returns)
```



Y (THE DEPENDENT VARIABLE)

STEP 4: REGRESS THE ADJUSTED GOOGLE RETURNS AGAINST NASDAQ RETURNS

```
googM <- lm(goog$goog.returns~goog$nasdaq.returns)
```

IF YOU HAVE MULTIPLE INDEPENDENT VARIABLES SEPARATE THEM BY +

X (THE INDEPENDENT VARIABLE)

STEP 5: PRINT THE REGRESSION RESULTS

```
summary(googM)
```

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.184265 -0.040596 -0.002853  0.028320  0.192225

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.005303   0.007086   0.748    0.457
goog$nasdaq.returns 0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```

LET'S PARSE THIS RESULT

```
Residuals:
      Min          1Q      Median          3Q         Max
-0.184265  -0.040596  -0.002853   0.028320   0.192225

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.005303   0.007086   0.748    0.457
goog$nasdaq.returns 0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```

MIN, MAX, MEDIAN, QUARTILES OF THE RESIDUALS

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.184265 -0.040596 -0.002853  0.028320  0.192225

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.005303   0.007086   0.748    0.457
goog$nasdaq.returns 0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```

THIS TELLS US WHAT THE RANGE AND OTHER CHARACTERISTICS OF THE ERRORS ARE

```
Residuals:
      Min          1Q      Median          3Q         Max
-0.184265 -0.040596 -0.002853  0.028320  0.192225

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.005303   0.007086   0.748    0.457
goog$nasdaq.returns  0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```

THIS SECTION DEALS WITH THE CO-EFFICIENTS / INTERCEPT

```
Residuals:
      Min          1Q      Median          3Q         Max
-0.184265 -0.040596 -0.002853  0.028320  0.192225

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.005303   0.007086   0.748    0.457
goog$nasdaq.returns 0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```

THE ESTIMATED VALUE OF THE CO-EFFICIENTS

```
Residuals:
      Min          1Q      Median          3Q         Max
-0.184265 -0.040596 -0.002853  0.028320  0.192225

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.005303   0.007086   0.748    0.457
goog$nasdaq.returns 0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```

BETA FOR GOOGLE (SLOPE OF THE REGRESSION LINE)

```
Residuals:
      Min          1Q      Median          3Q         Max
-0.184265 -0.040596 -0.002853  0.028320  0.192225

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.005303   0.007086   0.748    0.457
goog$nasdaq.returns 0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```

STANDARD ERROR OF THE CO-EFFICIENTS

(THIS TELLS US THE MAGNITUDE THE CO-EFFICIENT MIGHT BE OFF BY)

```
Residuals:
      Min          1Q      Median          3Q         Max
-0.184265 -0.040596 -0.002853  0.028320  0.192225

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.005303   0.007086   0.748    0.457
goog$nasdaq.returns  0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```

THE VALUE OF THE T-STATISTIC

(USED TO TEST THE STATICAL SIGNIFICANCE OF THE RELATIONSHIP BETWEEN THIS X AND Y)

```
Residuals:
      Min          1Q      Median          3Q         Max
-0.184265 -0.040596 -0.002853  0.028320  0.192225

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.005303   0.007086   0.748    0.457
goog$nasdaq.returns  0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```

THE P-VALUE (1-SIGNIFICANCE LEVEL)

(IF THIS VALUE IS <0.1, THE RELATIONSHIP IS STATISTICALLY SIGNIFICANT)

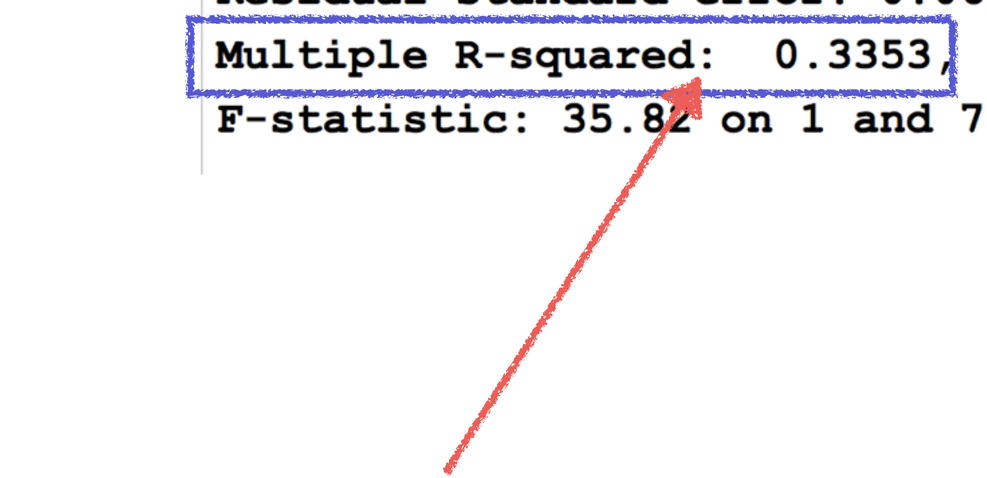
```
Residuals:
      Min          1Q      Median          3Q         Max
-0.184265 -0.040596 -0.002853  0.028320  0.192225

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.005303   0.007086   0.748    0.457
goog$nasdaq.returns  0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```

SINCE THIS VALUE IS <0.01, THE RELATIONSHIP BETWEEN GOOGLE AND NASDAQ RETURNS IS STATISTICALLY SIGNIFICANT

```
(Intercept)          0.005303   0.007086   0.748    0.457
goog$nasdaq.returns  0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```

THIS SECTION DEALS WITH THE STRENGTH OF THE OVERALL MODEL (THE PREVIOUS SECTION WAS CONCERNED WITH EACH VARIABLE'S ROLE IN THE MODEL)

```
(Intercept)          0.005303   0.007086   0.748    0.457
goog$nasdaq.returns 0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```



R-SQUARED

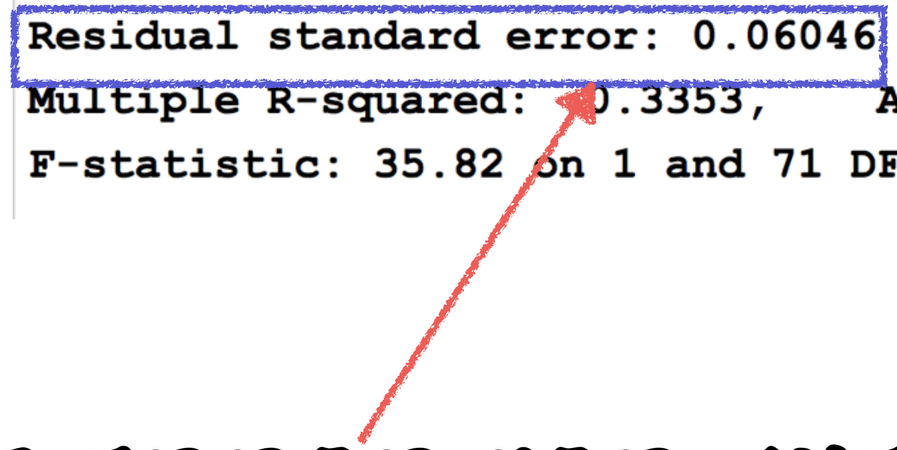
WHAT % OF THE VARIATION IN THE DEPENDENT VARIABLE HAS BEEN EXPLAINED

```
(Intercept)          0.005303   0.007086   0.748    0.457
goog$nasdaq.returns 0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```

ADJUSTED R-SQUARED

ADJUSTS R-SQUARED FOR THE NUMBER OF VARIABLES IN THE MODEL, THIS IS SO THAT WE AVOID "OVERFITTING"

```
(Intercept)          0.005303   0.007086   0.748    0.457
goog$nasdaq.returns  0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```



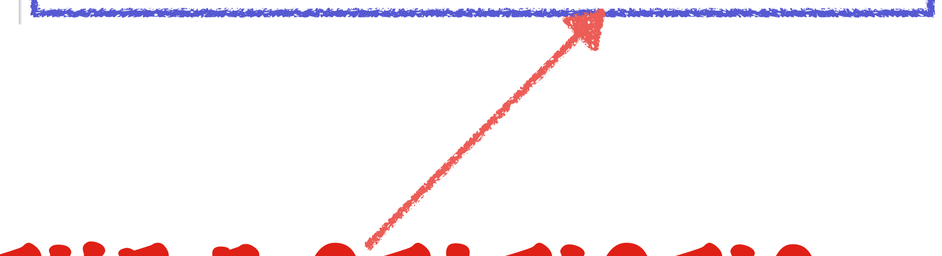
STANDARD ERROR FOR THE RESIDUALS

```
(Intercept)          0.005303   0.007086   0.748    0.457
goog$nasdaq.returns  0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```

DEGREES OF FREEDOM

NUMBER OF DATA POINTS - NUMBER OF VARIABLES - 1

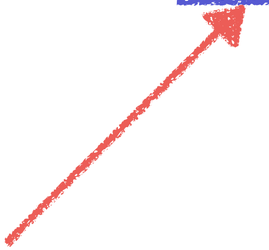
```
(Intercept)          0.005303   0.007086   0.748    0.457
goog$nasdaq.returns  0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```



THE F-STATISTIC

THIS IS A TEST STATISTIC THAT WE CAN USE TO TELL IF THE TRUE VALUE OF R-SQUARED IS 33% OR NOT

```
(Intercept)          0.005303   0.007086   0.748    0.457
goog$nasdaq.returns 0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```



THE P-VALUE (1-SIGNIFICANCE LEVEL)

(IF THIS VALUE IS <0.1, THE MODEL IS STATISTICALLY SIGNIFICANT)

R VS EXCEL

```
Residuals:
      Min         1Q     Median         3Q        Max
-0.184265 -0.040596 -0.002853  0.028320  0.192225

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.005303   0.007086   0.748    0.457
goog$nasdaq.returns 0.958860   0.160220   5.985 8.02e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06046 on 71 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3259
F-statistic: 35.82 on 1 and 71 DF,  p-value: 8.018e-08
```

AS YOU CAN SEE THE R VERSION GIVES MUCH RICHER INFORMATION ON THE REGRESSION STATISTICS THAN EXCEL

R VS EXCEL

THE R VERSION GIVES MUCH RICHER INFORMATION ON THE REGRESSION STATISTICS THAN EXCEL

ALL THE PREPROCESSING STEPS CAN BE PUT INTO A FUNCTION

```
preProcess <- function(googFile,nasdaqFile){
goog <-read.table(googFile,header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(goog)[2]<-"goog.price"
nasdaq <- read.table(nasdaqFile,header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(nasdaq)[2]<-"nasdaq.price"
goog <- merge(goog, nasdaq, by = "Date")
goog[,c("Date")] <- as.Date(goog[,c("Date")])
goog <- goog[order(goog$Date, decreasing = TRUE),]
goog[-nrow(goog),-1] <- goog[-nrow(goog),-1]/goog[-1,-1]-1
goog <- goog[-nrow(goog),]
tbonds <- read.table('/Users/swethakolalapudi/Desktop/Regression/tbond5yr.csv',header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(tbonds)[2]<-"tbonds.returns"
tbonds[,c("Date")] <- as.Date(tbonds[,c("Date")])
goog <- merge(goog, tbonds, by="Date")
goog$tbonds.returns <- goog$tbonds.returns/100
names(goog)[2:3] <- c("goog.returns","nasdaq.returns")
goog[,c("goog.returns","nasdaq.returns")] <- goog[,c("goog.returns","nasdaq.returns")]-goog[,"tbonds.returns"]
return(goog)}
```

R VS EXCEL

ALL THE PREPROCESSING STEPS CAN BE PUT INTO A FUNCTION

```
preProcess <- function(googFile,nasdaqFile){
goog <-read.table(googFile,header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(goog)[2]<-"goog.price"
nasdaq <- read.table(nasdaqFile,header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(nasdaq)[2]<-"nasdaq.price"
goog <- merge(goog, nasdaq, by = "Date")
goog[,c("Date")] <- as.Date(goog[,c("Date")])
goog <- goog[order(goog$Date, decreasing = TRUE),]
goog[-nrow(goog),-1] <- goog[-nrow(goog),-1]/goog[-1,-1]-1
goog <- goog[-nrow(goog),]
tbonds <- read.table('/Users/swethakolalapudi/Desktop/Regression/tbond5yr.csv',header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(tbonds)[2]<-"tbonds.returns"
tbonds[,c("Date")] <- as.Date(tbonds[,c("Date")])
goog <- merge(goog, tbonds, by="Date")
goog$tbonds.returns <- goog$tbonds.returns/100
names(goog)[2:3] <- c("goog.returns","nasdaq.returns")
goog[,c("goog.returns","nasdaq.returns")] <- goog[,c("goog.returns","nasdaq.returns")]-goog[,"tbonds.returns"]
return(goog)}
```

ALL THE CODE WE USED TO PREPROCESS THE DATA

R VS EXCEL

ALL THE PREPROCESSING STEPS CAN BE PUT INTO A FUNCTION

```
preProcess <- function(googFile,nasdaqFile){
goog <-read.table(googFile,header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(goog)[2]<-"goog.price"
nasdaq <- read.table(nasdaqFile,header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(nasdaq)[2]<-"nasdaq.price"
goog <- merge(goog, nasdaq, by = "Date")
goog[,c("Date")] <- as.Date(goog[,c("Date")])
goog <- goog[order(goog$Date, decreasing = TRUE),]
goog[-nrow(goog),-1] <- goog[-nrow(goog),-1]/goog[-1,-1]-1
goog <- goog[-nrow(goog),]
tbonds <- read.table('/Users/swethakolalapudi/Desktop/Regression/tbond5yr.csv',header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(tbonds)[2]<-"tbonds.returns"
tbonds[,c("Date")] <- as.Date(tbonds[,c("Date")])
goog <- merge(goog, tbonds, by="Date")
goog$tbonds.returns <- goog$tbonds.returns/100
names(goog)[2:3] <- c("goog.returns","nasdaq.returns")
goog[,c("goog.returns","nasdaq.returns")] <- goog[,c("goog.returns","nasdaq.returns")]-goog[,"tbonds.returns"]
return(goog)}
```

FUNCTION THAT TAKES IN 2 FILE NAMES

R VS EXCEL

ALL THE PREPROCESSING STEPS CAN BE PUT INTO A FUNCTION

```
preProcess <- function(googFile,nasdaqFile){
goog <-read.table(googFile,header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(goog)[2]<-"goog.price"
nasdaq <- read.table(nasdaqFile,header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(nasdaq)[2]<-"nasdaq.price"
goog <- merge(goog, nasdaq, by = "Date")
goog[,c("Date")] <- as.Date(goog[,c("Date")])
goog <- goog[order(goog$Date, decreasing = TRUE),]
goog[-nrow(goog),-1] <- goog[-nrow(goog),-1]/goog[-1,-1]-1
goog <- goog[-nrow(goog),]
tbonds <- read.table('/Users/swethakolalapudi/Desktop/Regression/tbond5yr.csv',header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(tbonds)[2]<-"tbonds.returns"
tbonds[,c("Date")] <- as.Date(tbonds[,c("Date")])
goog <- merge(goog, tbonds, by="Date")
goog$tbonds.returns <- goog$tbonds.returns/100
names(goog)[2:3] <- c("goog.returns","nasdaq.returns")
goog[,c("goog.returns","nasdaq.returns")] <- goog[,c("goog.returns","nasdaq.returns")]-goog[,"tbonds.returns"]
return(goog)}
```

RETURNS A DATA FRAME WITH RETURNS - RISK FREE RETURNS OF THE 2 SECURITIES

R VS EXCEL

ALL THE PREPROCESSING STEPS CAN BE PUT INTO A FUNCTION

```
preProcess <- function(googFile,nasdaqFile){
goog <-read.table(googFile,header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(goog)[2]<-"goog.price"
nasdaq <- read.table(nasdaqFile,header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(nasdaq)[2]<-"nasdaq.price"
goog <- merge(goog, nasdaq, by = "Date")
goog[,c("Date")] <- as.Date(goog[,c("Date")])
goog <- goog[order(goog$Date, decreasing = TRUE),]
goog[-nrow(goog),-1] <- goog[-nrow(goog),-1]/goog[-1,-1]-1
goog <- goog[-nrow(goog),]
tbonds <- read.table('/Users/swethakolalapudi/Desktop/Regression/tbond5yr.csv',header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(tbonds)[2]<-"tbonds.returns"
tbonds[,c("Date")] <- as.Date(tbonds[,c("Date")])
goog <- merge(goog, tbonds, by="Date")
goog$tbonds.returns <- goog$tbonds.returns/100
names(goog)[2:3] <- c("goog.returns","nasdaq.returns")

goog[,c("goog.returns","nasdaq.returns")] <- goog[,c("goog.returns","nasdaq.returns")]-goog[,"tbonds.returns"]
return(goog)}
```

NEXT TIME YOU CAN USE JUST 1 LINE OF CODE TO DO ALL THE PRE-PROCESSING

```
goog <- preProcess(googFile,nasdaqFile)
```

R VS EXCEL

THE R VERSION GIVES MUCH RICHER INFORMATION ON THE REGRESSION STATISTICS THAN EXCEL

ALL THE PREPROCESSING STEPS CAN BE PUT INTO A FUNCTION

IN R YOU CAN DEAL WITH MISSING VALUES IN THE LINEAR REGRESSION

```
googM <- lm(goog$goog.returns~goog$nasdaq.returns, na.action = na.omit)
```

R VS EXCEL

IN R YOU CAN DEAL WITH MISSING VALUES IN THE LINEAR REGRESSION

```
googM <- lm(goog$goog.returns~goog$nasdaq.returns, na.action = na.omit)
```

THIS IS THE
DEFAULT OPTION

HERE WE ARE SPECIFYING
THAT MISSING VALUES
(VALUES WHICH ARE NA)
SHOULD BE OMITTED

R VS EXCEL

IN R YOU CAN DEAL WITH MISSING VALUES IN THE LINEAR REGRESSION

```
googM <- lm(goog$goog.returns~goog$nasdaq.returns, na.action = na.omit)
```

HERE WE ARE SPECIFYING THAT MISSING VALUES (VALUES WHICH ARE NA) SHOULD BE OMITTED

OTHER OPTIONS YOU COULD SPECIFY:

NA.EXCLUDE OMITS THE MISSING VALUES , BUT PADS THE Y VALUES RETURNED WITH NAS

NA.FAIL THROWS AN ERROR IF THERE ARE MISSING VALUES

R VS EXCEL

IN R YOU CAN DEAL WITH MISSING VALUES IN THE LINEAR REGRESSION

```
googM <- lm(goog$goog.returns~goog$nasdaq.returns, na.action = na.omit)
```

ANOTHER OPTION IS TO DEAL WITH NA VALUES IN THE PREPROCESSING STEP

```
goog[,"goog.returns"][is.na(goog[,"goog.returns"])] <-mean(goog[,"goog.returns"])
```

IF ANY OF THE RETURNS IS NA, REPLACE IT WITH THE MEAN RETURNS

R VS EXCEL

THE R VERSION GIVES MUCH RICHER INFORMATION ON THE REGRESSION STATISTICS THAN EXCEL

ALL THE PREPROCESSING STEPS CAN BE PUT INTO A FUNCTION

IN R YOU CAN DEAL WITH MISSING VALUES IN THE LINEAR REGRESSION

THERE ARE BUNCH OF OTHER COOL THINGS YOU CAN DO WITH REGRESSION IN R

WORKING WITH CATEGORICAL VARIABLES

ROBUST MODELS (WHICH DEAL WITH OUTLIERS)

VALIDATING THE ASSUMPTIONS OF LINEAR REGRESSION

WE'LL SEE ALL OF THESE IN THE NEXT FEW EXAMPLES