Vijay Yedidi
Financial Econometrics HW 1
Due October 1, 2020
Professor Francis

Question 1:

Code For reading the CSV File created from the data into a dataframe

```
CsvRead <- function(data_location) {
   ##Function to define the dataframes,
   # converts the date column into r date format
   # for easy reading

   df <- read.csv(data_location, header = TRUE)
   df$date <- strptime(df$date, '%Y%m%d')

   ###Questions 1 and 2 Part c: Transform the simplereturns to log returns
   df$LOGRET <- log(1 + df$RET)
   return(df)
}
cat <- CsvRead('cat.csv')
proc_gamble <- CsvRead('Proc_Gam.csv')
```

Part A: *Compute the sample mean, standard deviation, skewness, excess kurtosis, minimum, and maximum of each simple return series.*

Code:

```
Moments <- function (ret) {
   ###Questions 1 and 2 part a: Find all the moments of the data sets.
   ## This function creates a dataframe that contains the moments of the returns
   #Moment 1: mean
   #Moment 2: standard deviation
   #Moment 3: skewness
   #Moment 4: kurtosis
   #Min and Max are outputted
   ret <- ret[-c(1, 2) ]
   mean <- apply(ret, 2, mean)
   sd <- apply(ret, 2, sd)
   skewness <- apply(ret, 2, skewness)
   kurtosis <- apply(ret, 2, kurtosis)
   min <- apply(ret, 2, min)
   max <- apply(ret, 2, max)
```

```
    return (data.frame(rbind(mean,
                             sd,
                             skewness,
                             kurtosis,
                             min,
                             max)))
}
```

Question 1A: Calls the function to write the descriptive statistics for the dataframes ##into a new dataframe.

```
cat_moments <- Moments(cat)
```

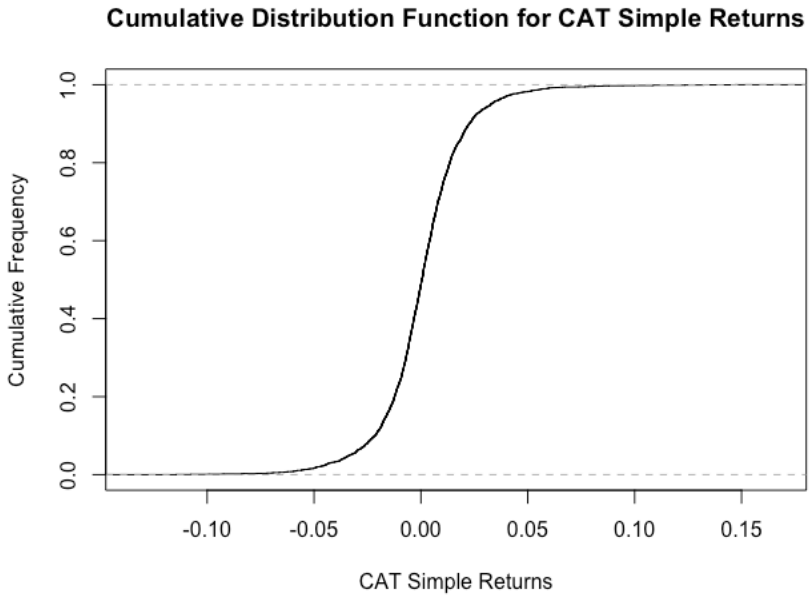
Output: *For Question 1A: The descriptive stats for each simple return series.*

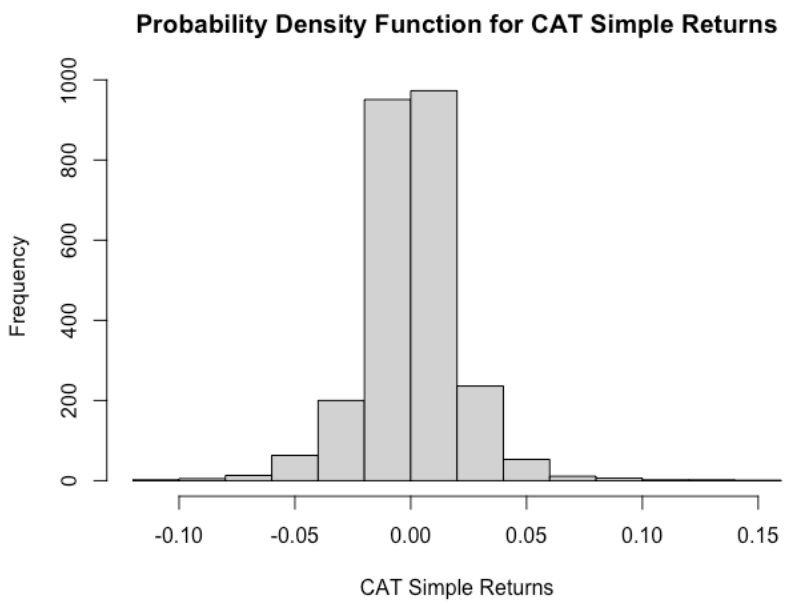
	RET	vwretd	ewretd	sprtrn	LOGRET
mean	0.0004945544	0.0003372415	0.0004484198	0.0002682593	0.0002760543
sd	0.0209288110	0.0132067743	0.0122023530	0.0131790153	0.0208998419
skewness	0.2305660851	-0.1880250583	-0.2247949821	-0.0895182926	0.0164783265
kurtosis	8.0354031658	12.1506090755	10.8562375229	13.1545094211	7.7452472400
min	-0.1143400000	-0.0897710000	-0.0782400000	-0.0903500000	-0.1214221492
max	0.1472290000	0.1148870000	0.1074220000	0.1158000000	0.1373494695

Part b: Empirical Density Function of the simple returns of Caterpillar Stock. Are the daily simple returns normally distributed?

Code:

```
plot(cat_cum_dist, xlab = 'Simple Returns', ylab = 'Cumulative Frequency', main = 'Cumulative Distribution Function for Cat Simple Returns')
hist(cat$RET, xlab = 'Simple Returns', ylab = 'Frequency', main = 'Probability Distribution Function for Cat Simple Returns')
```





At first glance, if you look at the pots, the cumulative distribution function follows a relatively sigmoid curve, and the probability density histogram is relatively normal. But when it is run through the Jarque-Bera test of normality, the test proves that this distribution is not normal. When testing further, as explained in the R Guide handbook, the D'Agostino test proves that the data is in fact skewed, and the Anscombe-Glynn test proves the kurtosis is not equal to 3. So the null hypothesis of this being a normal distribution can be rejected.

```
> jarque.test(cat$RET)

        Jarque-Bera Normality Test

data:  cat$RET
JB = 2682.5, p-value < 2.2e-16
alternative hypothesis: greater

> agostino.test(cat$RET)

        D'Agostino skewness test

data:  cat$RET
skew = 0.23057, z = 4.67701, p-value = 2.911e-06
alternative hypothesis: data have a skewness

> anscombe.test(cat$RET)

        Anscombe-Glynn kurtosis test

data:  cat$RET
kurt = 8.0354, z = 16.8852, p-value < 2.2e-16
alternative hypothesis: kurtosis is not equal to 3

>
```

Part C: *Calculate the log returns and find the same descriptive stats from part A for the log returns.*

In order to calculate the log returns, I simply used the equation found in the class notes.

Log Returns = log(1 + Returns)

$$\text{Log Returns} = \log(1 + \text{ Returns})$$

Code:

```
df$LOGRET <- log(1 + df$RET)
```

I implemented this line of code within the CsvRead Function (seen on first page) so it can easily be called for both Caterpillar and Proctor & Gamble data

	RET	vwretd	ewretd	sprtrn	LOGRET
mean	0.0004945544	0.0003372415	0.0004484198	0.0002682593	0.0002760543
sd	0.0209288110	0.0132067743	0.0122023530	0.0131790153	0.0208998419
skewness	0.2305660851	-0.1880250583	-0.2247949821	-0.0895182926	0.0164783265
kurtosis	8.0354031658	12.1506090755	10.8562375229	13.1545094211	7.7452472400
min	-0.1143400000	-0.0897710000	-0.0782400000	-0.0903500000	-0.1214221492
max	0.1472290000	0.1148870000	0.1074220000	0.1158000000	0.1373494695

The Descriptive Stats for the Log Returns can be seen in the last column of this table (Produced in Part A)

Part D: *Test that the mean of the log returns of Caterpillar stock is 0. Do the same with the S&P composite index.*

```
> t.test(cat$LOGRET, mu = 0 )

        One Sample t-test

data:  cat$LOGRET
t = 0.6628, df = 2517, p-value = 0.5075
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.0005406637  0.0010927723
sample estimates:
   mean of x
0.0002760543

> t.test(cat$sprtrn, mu = 0)

        One Sample t-test

data:  cat$sprtrn
t = 1.0214, df = 2517, p-value = 0.3072
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.0002467464  0.0007832651
sample estimates:
   mean of x
0.0002682593
```

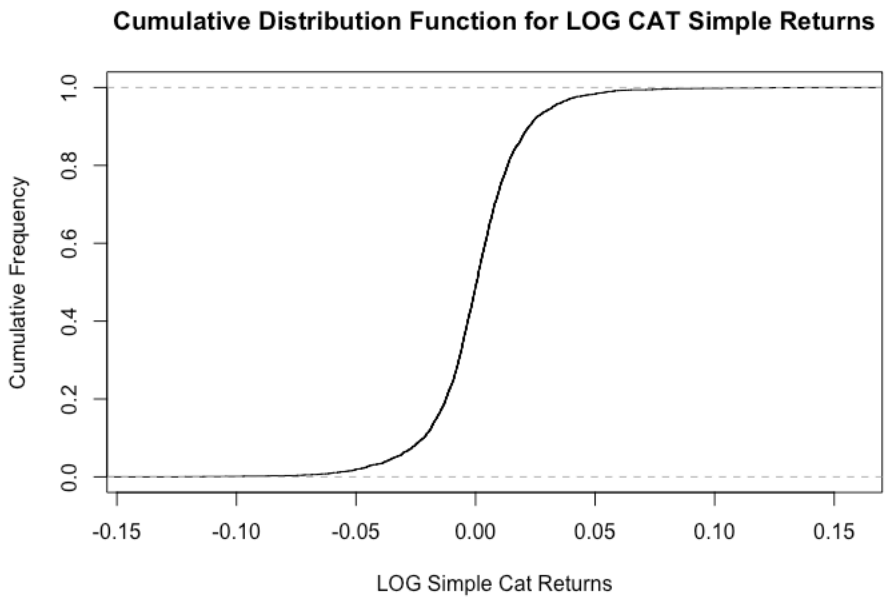
The t test was performed with a 95% confidence interval, and the range of values within the confidence interval contained 0. So we can accept the null hypothesis that the mean of the log returns of the Caterpillar Stock is 0, and we can accept the null hypothesis that the mean of the S&P composite index is 0 as well.

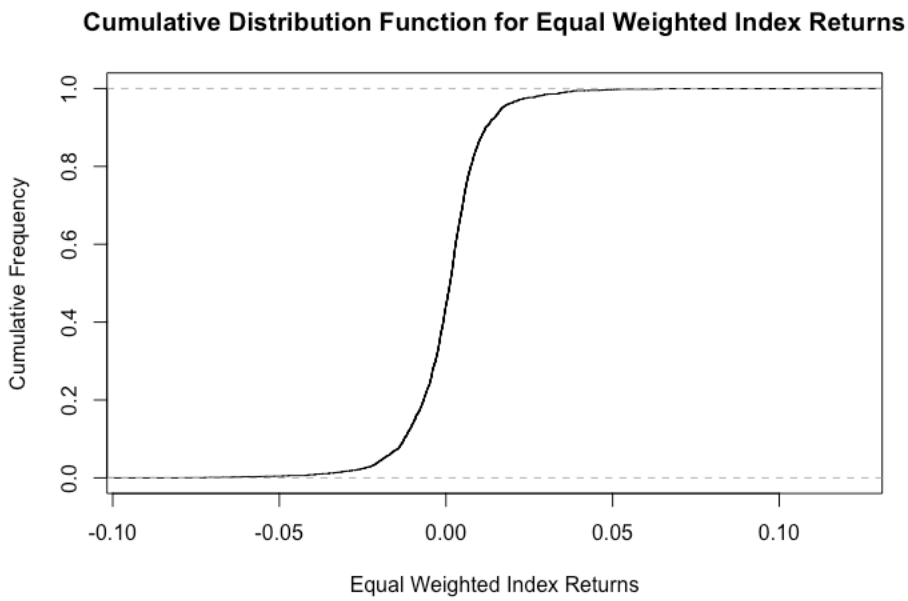
Part E: Obtain the empirical density plot of the daily log returns of Caterpillar stock and the equal weighted index

Code:

```
## Cumulative Distribution Functions
logret_cdf <- ecdf(cat$LOGRET)
ew_cdf <- ecdf(cat$ewretd)
##Plots for Cumulative Distribution Functions
plot(ew_cdf, xlab = 'LOG Simple Cat Returns',
    ylab = 'Cumulative Frequency',
    main = 'Cumulative Distribution Function for LOG CAT Simple Returns')
plot(logret_cdf, xlab = 'Equal Weighted Index Returns',
    ylab = 'Cumulative Frequency',
    main = 'Cumulative Distribution Function for Equal Weighted Index Returns')
```

Output


Cumulative Distribution Function for LOG CAT Simple Returns
Cumulative Frequency
0.0
0.2
0.4
0.6
0.8
1.0
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
LOG Simple Cat Returns


Cumulative Distribution Function for Equal Weighted Index Returns
Cumulative Frequency
0.0
0.2
0.4
0.6
0.8
1.0
-0.10
-0.05
0.00
0.05
0.10
Equal Weighted Index Returns

Question 2:

Part A: *Calculate the mean, standard deviation, skewness, excess kurtosis, minimum, and maximum.*

Code: Same function used as in question 1. Moments was called
proc_gamble_moments <- Moments(proc_gamble)

Output: *A data frame for all the descriptive statistics being asked for.*

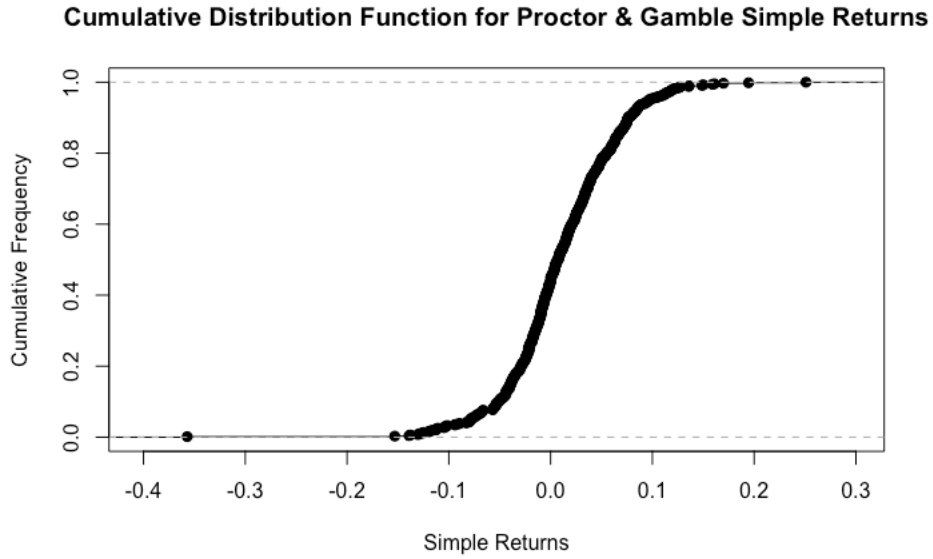
	RET	vwretd	ewretd	sprtrn
mean	0.01034235	0.00887971	0.01136855	0.006359545
sd	0.05538326	0.04403532	0.05554414	0.042515192
skewness	-0.29759770	-0.52360232	-0.18107535	-0.423835943
kurtosis	6.23883030	4.99943376	5.93700886	4.828034399
min	-0.35704100	-0.22536300	-0.27224800	-0.217630000
max	0.25093100	0.16558500	0.29926000	0.163047000

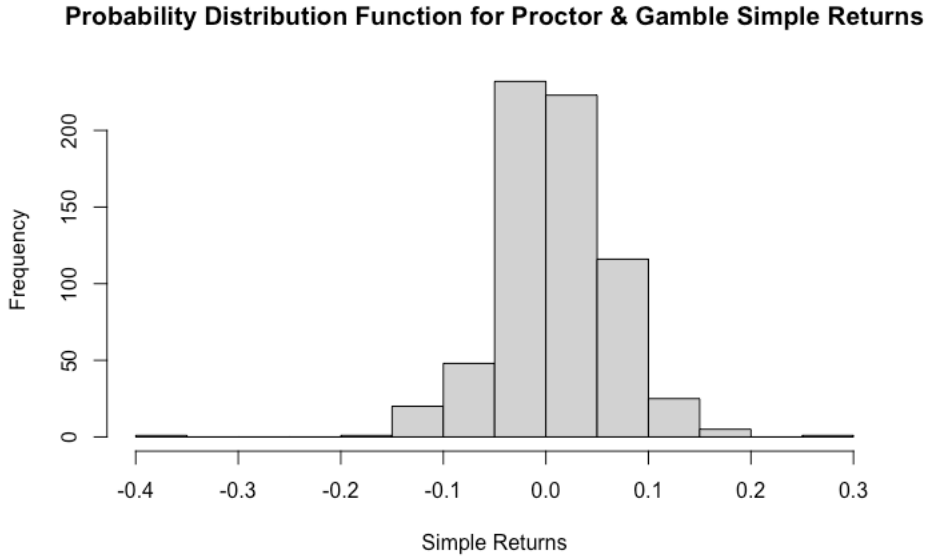
Part B: *Obtain the empirical density function of the simple returns for Caterpillar stock. Are they normally distributed?*

Code:

```
proc_cum_dist <- ecdf(proc_gamble$RET)
plot(proc_cum_dist,
    xlab = 'Simple Returns',
    ylab = 'Cumulative Frequency',
    main = 'Cumulative Distribution Function for Proctor & Gamble Simple Returns')
hist(proc_gamble$RET,
    xlab = 'Simple Returns',
    ylab = 'Frequency',
    main = 'Probability Distribution Function for Proctor & Gamble Simple Returns')
```

Output:





It can be seen from the plots (both in the cumulative distribution and the probability distribution) that there are outliers, the data is skewed, and the kurtosis is not correct. But to make sure, we run the Jarque Bera test, the D'Agostino skewness test, and the Anscombe-Glynn kurtosis test (as covered in the R Guide).

```
> jarque.test(proc_gamble$RET)

        Jarque-Bera Normality Test

data:  proc_gamble$RET
JB = 303.64, p-value < 2.2e-16
alternative hypothesis: greater

> agostino.test(proc_gamble$RET)

        D'Agostino skewness test

data:  proc_gamble$RET
skew = -0.2976, z = -3.1192, p-value = 0.001813
alternative hypothesis: data have a skewness

> anscombe.test(proc_gamble$RET)

        Anscombe-Glynn kurtosis test

data:  proc_gamble$RET
kurt = 6.2388, z = 7.5115, p-value = 5.844e-14
alternative hypothesis: kurtosis is not equal to 3
```

The hypotheses in all three tests can be rejected, so the data is clearly not normally distributed.

Part C: Just like in question 1, the log transform was added to the CsvRead function for re-usability purposes. Code for the function can be seen on page 1.

Log Returns = log(1 + Returns)

$$\text{Log Returns} = \log(1 + \text{Returns})$$

Code:

```
df$LOGRET <- log(1 + df$RET)
```

proc_gamble <- CsvRead('Proc_Gam.csv')

Output:

	RET	vwretd	ewretd	sprtrn	LOGRET
mean	0.01034235	0.00887971	0.01136855	0.006359545	0.008756106
sd	0.05538326	0.04403532	0.05554414	0.042515192	0.055806475
skewness	-0.29759770	-0.52360232	-0.18107535	-0.423835943	-0.834910553
kurtosis	6.23883030	4.99943376	5.93700886	4.828034399	9.379717213
min	-0.35704100	-0.22536300	-0.27224800	-0.217630000	-0.441674320
max	0.25093100	0.16558500	0.29926000	0.163047000	0.223888074

The log statistics for the log returns can be seen in the last column labeled 'LOGRET'.

Part D: *Test the null hypothesis that the log returns of Proctor & Gamble stock is 0. Do the same test for the S&P composite index.*

```
> t.test(proc_gamble$LOGRET, mu = 0)

        One Sample t-test

data:  proc_gamble$LOGRET
t = 4.0673, df = 671, p-value = 5.32e-05
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.004529107 0.012983105
sample estimates:
  mean of x
0.008756106

> t.test(proc_gamble$sprtrn, mu = 0 )

        One Sample t-test

data:  proc_gamble$sprtrn
t = 3.8776, df = 671, p-value = 0.0001159
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.003139279 0.009579810
sample estimates:
  mean of x
0.006359545
```

In both cases, we can reject the null hypothesis, and accept that the returns for both the log returns of Proctor and Gamble and the S&P composite index are not 0.

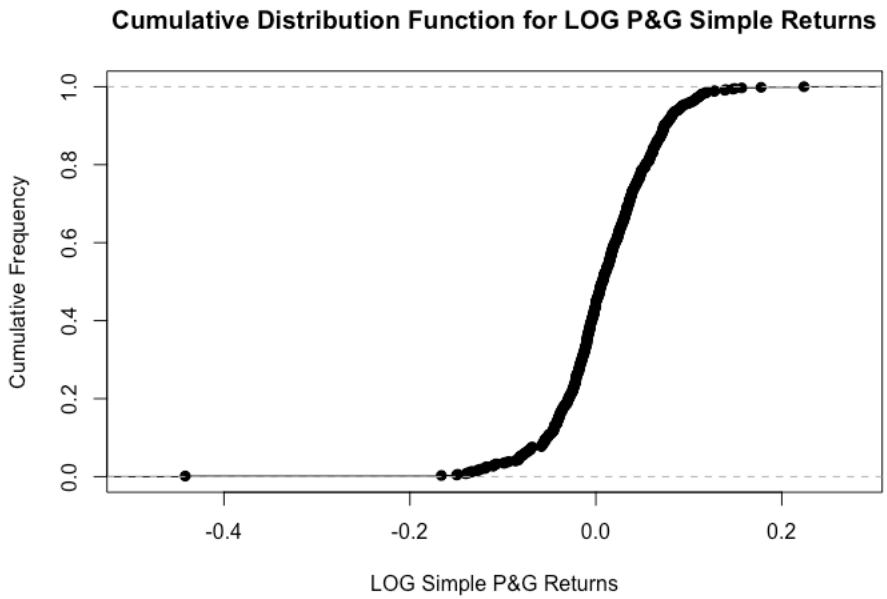
Part E: *Obtain the empirical density plot for daily log returns and the equal weighted index*

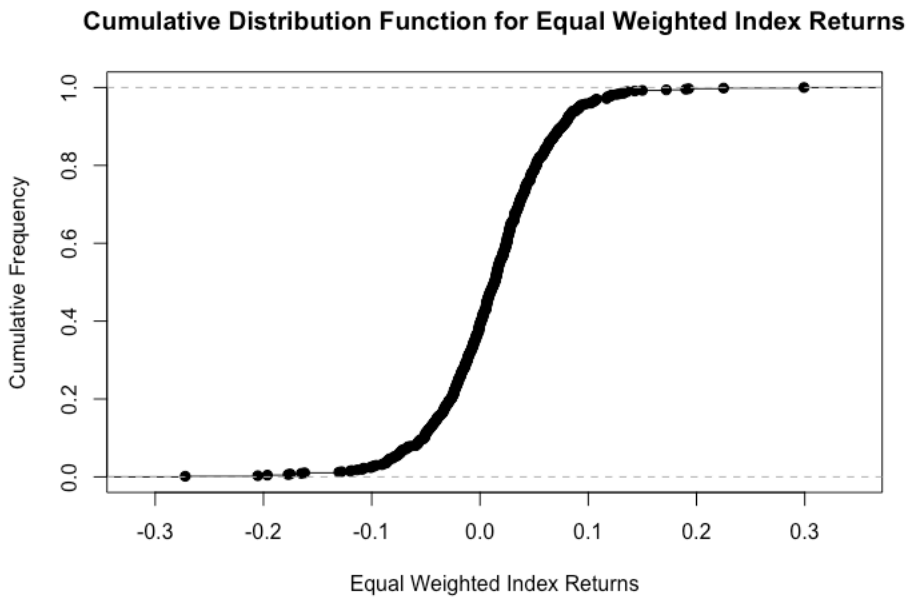
Code:

```
proclogret_cdf <- ecdf(proc_gamble$LOGRET)
ew_cdf <- ecdf(proc_gamble$ewretd)

plot(proclogret_cdf, xlab = 'LOG Simple P&G Returns',
    ylab = 'Cumulative Frequency',
    main = 'Cumulative Distribution Function for LOG P&G Simple Returns')
plot(ew_cdf, xlab = 'Equal Weighted Index Returns',
    ylab = 'Cumulative Frequency',
    main = 'Cumulative Distribution Function for Equal Weighted Index Returns')
```

Output:





Summary of Questions 1 and 2:

Overall, I wasn't too surprised with the results. At first glance in the empirical density function plots, it can be seen that the Proctor & Gamble was much more volatile than the Caterpillar stock and contained many more outliers. It was much more platykurtic with all its outliers, and so it wasn't surprising that even though it was logarithmically transformed, it was still not accepted as fully normal, while the Caterpillar returns were. However it was interesting to see that the S&P composite index was much more volatile in the distribution from 1961 to 2016, than it was from 2007 to 2016. This time period does expose the market data to more recessions, which can lead to much more volatility, and it could also explain why Proctor & Gamble was much more volatile as well. But overall, this exercise shows the importance of a diverse investment portfolio. We can see that the market is going to act less volatile than the returns of individual funds, so for maximal safety and predictability in your returns, you want as diverse assets as you can possibly obtain.

Question 3:
Part A: *A 100 basis point change in Treasury results in <u>a 2.5269% decrease</u> in the Commercial Bank Sector.*

Part B: *The empirical duration is the change in the price given a 1% change in bond yields.*

Part C:

3c) 10% change in treasury yield

$\rightarrow$ 1% = 100 basis points = 1x change

$\rightarrow$ 10% = $10(-2.527) = \boxed{-25.27\%}$

A 10% change in the Treasury Yield results in a 25.268% decrease in the Commercial Bank Sector.

Part D: *The t-statistic given for coefficient of the treasury yield change is -1.3083, which corresponds to a p value of 0.1926. This falls in the 95% confidence interval, so the null hypothesis that the coefficient of the treasury yield change equals 0 cannot be rejected.*

I am assuming the t value and the p value given are for the hypothesis that the mean = 0

Part E:

3e) $\beta_1 = -2.5269$ $\quad p = 0.1926$

t value = -1.3083

$$\frac{\beta - \mu = 0}{\text{standard error}} = -1.3083 = \frac{-2.5269}{\text{Standard error}}$$

Standard error = 1.931438

Part F:

3f) $$\frac{-2.5269 - (-2)}{\text{standard error}} = \text{t value} = \frac{-0.5269}{1.931438} = \boxed{-0.2728019}$$

From the t and p values given in the table and the t table provided in the back of the textbook, it can be assumed that we are dealing with about 30 degrees of freedom.. This is well within the confidence interval, meaning we can accept the null hypothesis that the change in treasury yield coefficient is not different from -2.

The code:
2 * pt(-0.2728019, df = 30, lower.tail = TRUE)

This shows the p value is 0.7868746 given that the t statistic is -0.2728091

This proves we cannot reject the null hypothesis and we must accept that the value -2 is not significantly different from the coefficient of the change in treasury yield.

Part G:

Code:
t_stat <- (-2.5269 - (-1))/1.931438
pt(t_stat, df = 30, lower.tail = TRUE)

This shows a p value of 0.218. This is within the confidence interval, and we cannot reject the null hypothesis, showing -1 is not significantly less than the coefficient of the treasury yield.

Part H:
The F-value in this case is 1.7116, with a p value of 0.1926. This is considered to be within the confidence interval if we are performing a test with 95% confidence, so the null hypothesis that both the slope and the intercept are not different from 0 holds to be true.
Part I:
We would want a p value less than 0.05 to reject the null hypothesis and assume that the slope and the intercept are jointly different than 0.
Part J:
R^2 = 0.0101. So 0.000101% of the total variation in the model is explained by the model itself.

Question 4:

Part A:

4a) Estimated coefficient = t statistic · standard error

$= 25 \cdot 0.03 = 0.75$

Part B:

4b) $(\beta_1 + \beta_2 D)\, x_i = (0.75 + 0.12 \cdot 1)\, x_i$

$= 0.87 \cdot 100 = 87$

A 100 unit increase in x would result in an 87 unit increase in y.

Part C:

A 0 unit change in x results in a 0 unit change in y during a down market.

Part D:

4d → $y_i = \alpha_i + (\beta_1 + \beta_2 D)\, x_i = -0.23 + (0.8 + 0.2)\, x_i$

$= -0.23 + 1$

$= 0.77$

The conditional expectation of y is 0.77 given X = 1 in an up market.

Part E:

4e → $y_i = \alpha_i + (\beta_1 + \beta_2 D)\, x_i = -0.23 + (0.8 + 0.2 \cdot 0)\, x_i$

$= -0.23 + 0.8 \cdot 1$

$= 0.57$

The conditional expectation of y is 0.57 given X = 1 in a down market.

Part F:

In order to test if alpha changes between up and down markets, we need to see the multicollinearity of the equation. This can be calculated via the Variance inflation factor (VIF)

$$VIF = \frac{1}{(1 - R_i^2)}$$

Where R_i^2 = the R^2 of the correlation between the alpha variable and every other independent variable in the model. If we find the VIF between alpha and $D_i * \beta_2 * x_t$, this would explain the relationship between the two independent variables. If the VIF is between 1 and 5, it shows a

slight correlation. If the VIF is greater than 5, it shows a heavy correlation. If the correlation is at its minimum value of 1, there is no correlation.

Part G:

4g) p value = 0.976

smallest level of significance where we can reject the null hypothesis = (1 − 0.976) = 0.024

Part H:

4h) standard error = $\frac{\text{estimate}}{\text{t statistic}} = \frac{-0.23}{-2.3} = 0.1$

$$\frac{-0.23-(-0.4)}{0.1} = 1.7$$

Looking at p and t values in the table, this data has infinite degrees of freedom

→ t statistic 1.7 is between a p level of 0.05 and 0.025

so the null hypothesis cannot be rejected.

The null hypothesis cannot be rejected, -0.4 can be considered significantly the same.

Part I:

For Fund A, the R^2 value is 0.93, while for fund b it is 0.83.

R^2 = (Explained Sum of Squares) / (Total Sum of Squares).

So the fund with the larger R^2 will have a higher portion of their fund's variance explained by the model. In this case, fund A has a higher R^2, and so a higher portion of fund A's variance is explained in the model.

Part J:

4j → $R^2 = \frac{ESS}{TSS}$ → $0.93 = \frac{ESS}{70}$ → $ESS = 65.1$

Part K:

4k → $R^2 = \frac{ESS}{TSS}$ ⇒ $0.83 = \frac{166}{TSS}$

→ $TSS = \frac{166}{0.83} = 200$

Question 5:

Part A:

5a) $R^2 = 0.65^2 = 0.4225$

Part B:

5b) $R^2 = \frac{ESS}{TSS} = 0.4225 = \frac{ESS}{100}$

$\rightarrow$ ESS (Residual Error SS) $= 42.25$

Part C:

5c) Regression SS $= TSS - ESS = 100 - 42.25$

$= 57.75$

Part D:

5d) $s^2 = \frac{RSS}{T-k} = \frac{57.75}{40-3} = 1.561$

Question 6:

Part A:

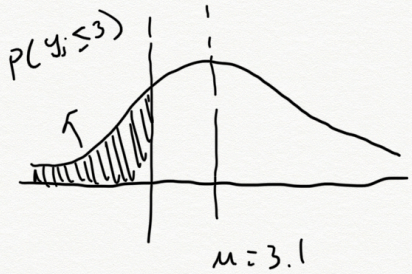
6a) → $y_i = 1.4 + 1.7 \cdot 1 + N(0, 0.3)$

$= 3.1 + N(0, 0.3)$

$= N(3.1, 0.3)$

→ mean of $y_i = \boxed{3.1}$

Standard deviation of $y_i = \sqrt{0.3} = \boxed{1.7321}$



→ given that $x = 1$

Code:

```
pnorm(3, mean = 3.1, sd = sqrt(0.3), lower.tail = TRUE)
```

Output:

The probability (Y is less than or equal to 3 given that X = 1) = 0.427566

Part B:

6b →

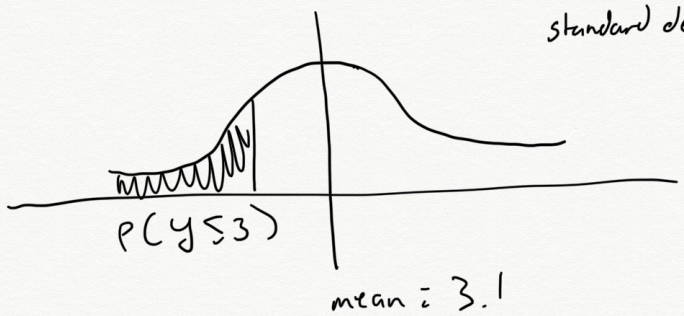
$$y_i = 1.4 + 1.7 \cdot N(1, 0.7) + N(0, 0.3)$$

$$= N(3.1, 1.7^2 \cdot 0.7 + 0.3)$$

$$= N(3.1, 2.323)$$

standard deviation $= \sqrt{2.323}$

$= 1.5242$



Code:

```
pnorm(3, mean = 3.1, sd = sqrt(2.323), lower.tail = TRUE)
```

Output:

The probability (Y is less than or equal to 3) = 0.4738438