**Project 1**

Name: Neil Yejjey

TU ID: TUK12930

**Question 1. (15 points)**

The dataset “Project1\_data\_1” has information for a sample of stocks, which are randomly chosen from a large population (i.e. all the stocks in the stock market). Please use approach #1 to test if the population mean of stock earnings is different from $5 at the 5% significance level.

1. The alternative hypothesis is \_μ≠5 or MU is not equal to 5 \_\_\_\_
2. The test-statistic is \_0.4382\_\_\_\_\_
3. The critical value(s) is(are) \_-2.0518 and 2.0518\_\_\_\_\_
4. We should \_\_Not Reject\_\_\_ the null hypothesis, and conclude that \_ population mean of stuck earnings is not different from $5 at the 5% significance level\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_



1. Please screenshot the R code you used for calculation and insert the image below.

Text

Description automatically generated

**Question 2. (15 points)**

The dataset “Project1\_data\_1” has information for a sample of stocks, which are randomly chosen from a large population (i.e. all the stocks in the stock market). Please use confidence interval to test if the population mean of stock price is different from $50 at the 5% significance level.

1. The alternative hypothesis is \_\_μ≠50\_\_\_\_\_\_\_
2. The confidence interval is \_(55.63792 , 78.95493)\_\_\_\_\_\_\_\_\_
3. The hypothesized value is \_\_\_50\_\_\_\_\_\_\_
4. We should \_reject\_\_\_\_ the null hypothesis, and conclude that \_\_\_the population mean of stock price is different from $50 at 5% significance level\_\_\_\_\_\_\_\_\_\_\_\_\_\_
5. Please screenshot the R code you used for calculation and insert the image below.

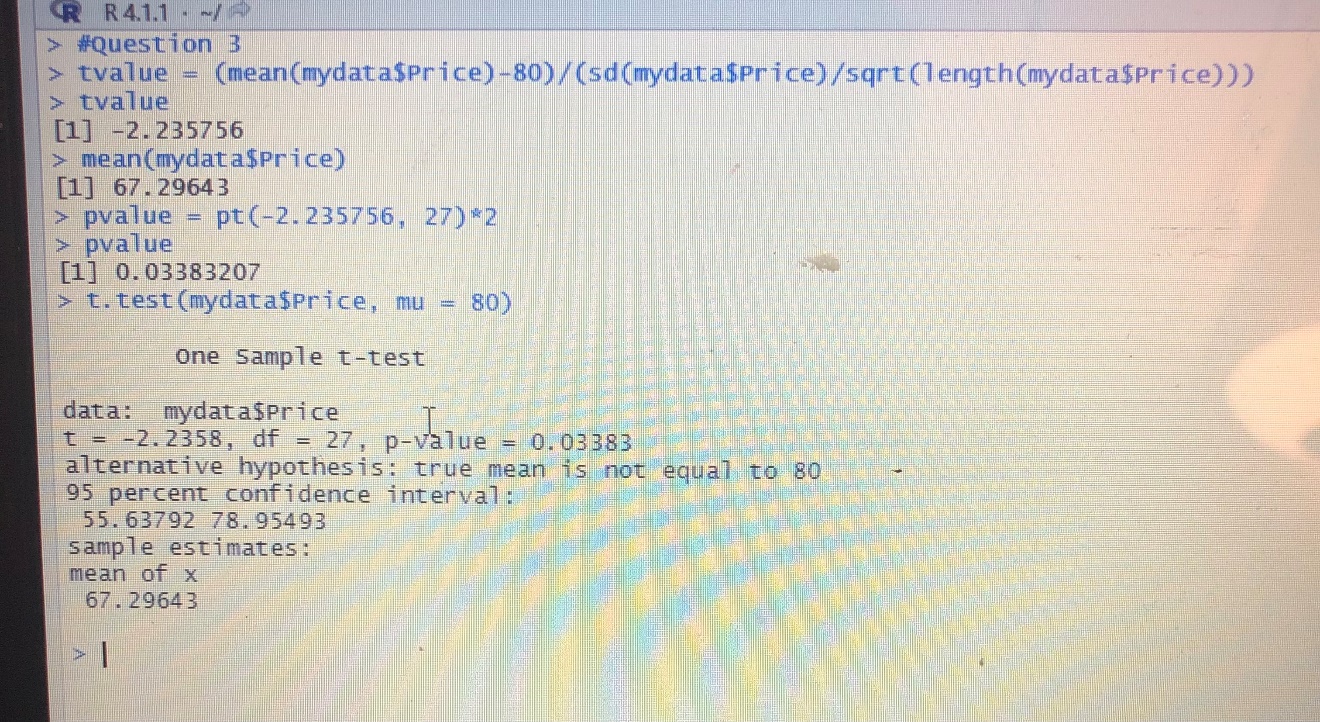
A picture containing text

Description automatically generated

**Question 3. (15 points)**

The dataset “Project1\_data\_1” has information for a sample of stocks, which are randomly chosen from a large population (i.e. all the stocks in the stock market). Please use approach #3 to test if the population mean of stock price is different from $80 at the 5% significance level.

1. The alternative hypothesis is \_\_ μ≠80 \_\_\_\_
2. The test-statistic is \_-2.235756\_\_\_\_\_
3. The P-value is \_\_0.03383207\_\_\_\_
4. We should \_\_\_reject\_\_ the null hypothesis, and conclude that \_the population mean of stock price is different from $80 at the 5% significance level\_\_\_\_\_\_
5. Please screenshot the R code you used for calculation and insert the image below.



**Question 4. (15 points)**

The dataset “Project1\_data\_1” has information for a sample of stocks, which are randomly chosen from a large population (i.e. all the stocks in the stock market). Please use test-statistic and critical value(s) to test if the population mean of stock earnings is greater than $4 at the 5% significance level.

1. The alternative hypothesis is μ > 4 or MU is greater than 4\_\_ \_\_\_\_
2. The test-statistic is \_\_ 1.973874 \_\_\_\_\_
3. The critical value(s) is(are) \_\_ 1.703288 \_\_\_\_\_\_
4. We should \_reject \_\_\_\_\_ the null hypothesis, and conclude that \_the population mean of stock earnings is greater than $4 at the 5% significance level\_\_\_\_\_\_\_\_\_\_\_\_
5. Please screenshot the R code you used for calculation and insert the image below.

A picture containing diagram

Description automatically generated

**Question 5. (15 points)**

The dataset “Project1\_data\_1” has information for a sample of stocks, which are randomly chosen from a large population (i.e. all the stocks in the stock market). Please use test-statistic and critical value(s) to test if the population variance of stock earnings is different from $2.5 at the 5% significance level.

1. The alternative hypothesis is **σ**2 ≠ 2.5\_\_\_\_\_\_\_
2. The test-statistic is \_\_160.2878\_\_\_\_\_
3. The critical value(s) is(are) \_\_43.19451 and 14.57338\_\_\_\_\_\_\_\_
4. We should \_ \_reject\_\_\_\_ the null hypothesis, and conclude that \_\_the population variance of stock earnings is NOT equal to $2.5 at the 5% significance level\_\_\_\_\_\_\_\_\_\_\_
5. Please screenshot the R code you used for calculation and insert the image below.

Text

Description automatically generated

**Question 6: (12.5 points)**

Use the dataset “Project1\_data\_2” and a simple linear regression model to analyze how the time spent on watching television shows can affect the quiz score. The data is created for you to practice and to understand how to do econometric analysis by using R. There are two variables in the data set. The first column is the independent variable, which represents the number of hours spent on watching television shows. The second column is the dependent variable, which represents the quiz score. There is no variable name in the data set, so you have to specify the header. The default names for the variables are X1 and X2, which have to be in the form of upper case. Assume the significance level is 5%.

1. Draw the scatter plot and the fitted straight line, and insert the image below Chart, scatter chart

   Description automatically generated
2. The regression equation is Y = 26.249 – 4.302X + e\_\_\_ \_ \_\_\_
3. Analyze the summary of regression results (i.e. state the estimated value of regression coefficients, interpret the use of t-value and R-squared, determine if each regression coefficient is different from zero by using t-value or P-value, and conclude whether and how watching TV show affects quiz score)

The intercept is 26.249 and the slope is -4.302. The T-Value for the slop coefficient is -10.93 which is less than the lower critical T-value, and P value is close to 0 and much smaller than the 5% significance level, both indicating that the slope coefficient is different from zero and watching TV effects exam score significantly. The R square is 0.7443 indicating that watching TV hours explains 74.43% of the change in exam score. An additional hour of watching TV reduces exam score by 10.93 points.

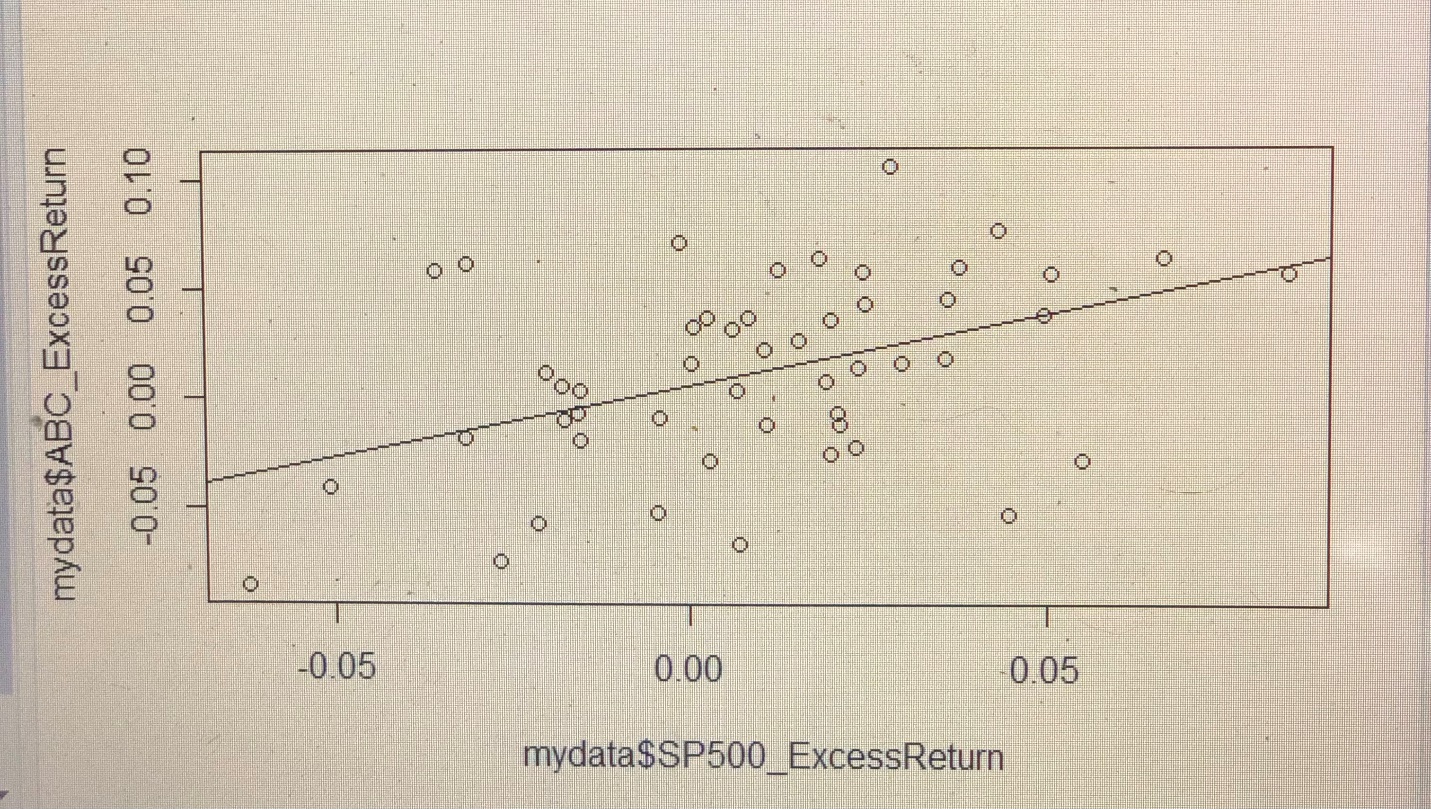
1. Please screenshot the R code you used for calculation and insert the image below.Text

   Description automatically generated Text

   Description automatically generated

**Question 7: (12.5 points)**

Use the dataset “Project1\_data\_3” and a simple linear regression model to calculate the beta of the stock. The data is created for you to practice and to understand how to do econometric analysis by using R. Please notice that you need to convert the stock price into the stock return. You can refer to the Excel document “CAPM.xlsx” to see the conversion. Assume the significance level is 5%.

1. Draw the scatter plot and the fitted straight line, and insert the image below 
2. The regression equation is \_y = 0.005506 + 0.648669 + e \_ \_\_\_
3. Analyze the summary of regression results (i.e. state the estimated value of regression coefficients, interpret the use of t-value and R-squared, determine if each regression coefficient is different from zero by using t-value or P-value, and interpret the value of beta for this stock)

The intercept is 0.005506 and the slope is 0.648669. The T-Value for the slope coefficient is 3.373 which is greater than the upper critical T-value, and P value is close to 0 and much smaller than the 5% significance level, indicating that the slope coefficient is different from zero and overall market movements effect ABC return significantly. The R square is 0.1916 indicating that watching TV hours explains 19.16% change of stock in ABC return . An additional one percent return of overall market value increases stock return 0.648669 percent.

1. Please screenshot the R code you used for calculation and insert the image below. A screenshot of a computer

   Description automatically generated with low confidence