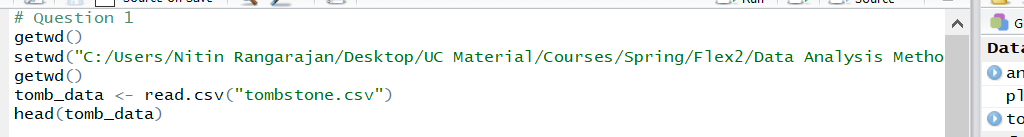
**BANA 7038 – Homework 3**

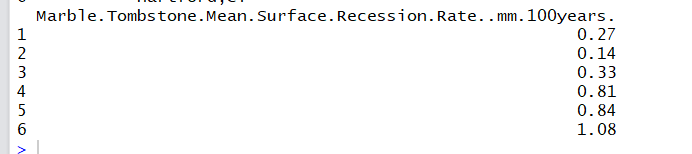
**Nitin Rangarajan (rangarnn) M12428135**

**Question 1 –** Read <tombstone.csv> into R. Use response variable = Marble Tombstone Mean Surface Recession Rate, and covariate = Mean SO2 concentrations over a 100-year period. Description: Marble Tombstone Mean Surface Recession Rates and Mean SO2 concentrations over a 100-year period

**Solution –**

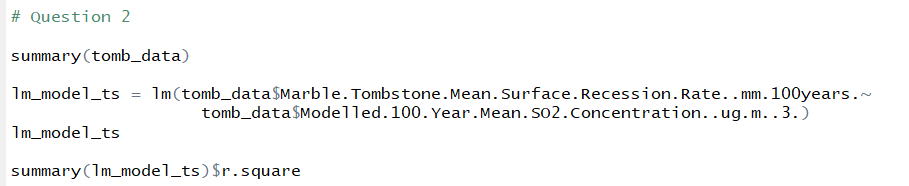


**Output:**

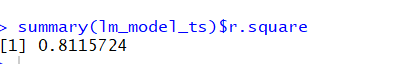
 **-------------------------------------------------------------------------------------------------------------------------------**

**Question 2** - Obtain , explain what it means.

**Solution –**



**Output –**  value is 0.8115



value is an indication of how close the data are to the fitted regression line. So, in this case there is a strong dependency (81%) between covariate and response variable.

**-------------------------------------------------------------------------------------------------------------------------------**

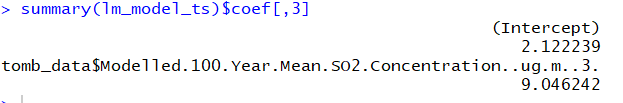
**Question 3** - Perform the following hypothesis testing and interval estimation using lm() and other related R functions.

3.1. Perform t tests, obtain t statistics and p values, interpret the results, make a conclusion (i.e. reject or not reject) and explain why. Note: please explain what the null hypothesis is.

**Solution –**



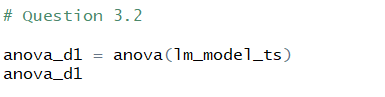
**Output –**



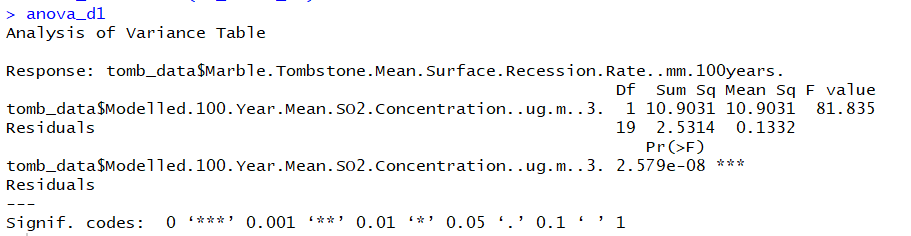
Null hypothesis is a test for determining the relationship between covariate and response variable. If we reject null hypothesis, then the variables have a linear relationship. Here, T-values are 2.12 and 9.049. Beta1 value is greater than 2 and P value is less than 0.05. Thus, we can reject null hypothesis. **----------------------------------------------------------------------------------------------------------------------------**

3.2. Perform ANOVA test (F test), obtain F statistic and p value, interpret the results, make conclusion (i.e. reject or not reject) and explain why. Note: please explain what the null hypothesis is.

**Solution –**



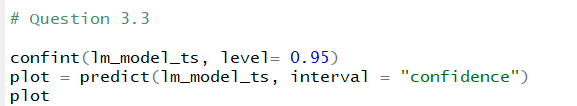
**Output –**



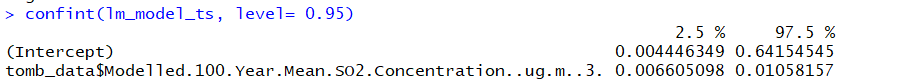
Null hypothesis is a test for determining the relationship between covariate and response variable. If we reject null hypothesis, then the variables have a linear relationship. Here, we can see that the F-value is greater than F-critical and thus, we can reject null hypothesis.   
**-------------------------------------------------------------------------------------------------------------------------------**

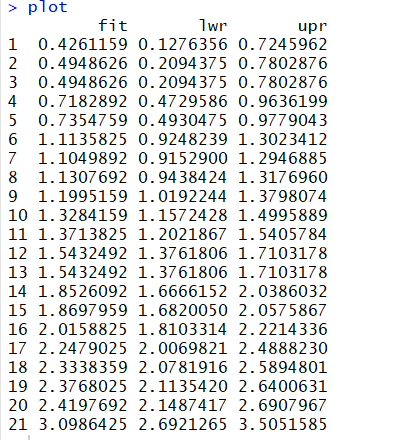
3.3. Compute confidence interval for coefficients, fitted values (mean response), interpret the meanings of these quantities.

**Solution –**



**Output –**

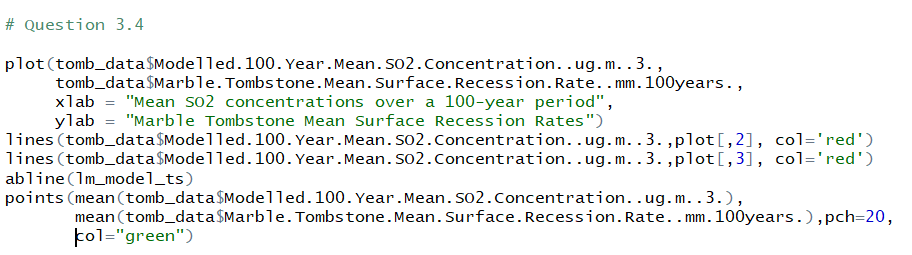




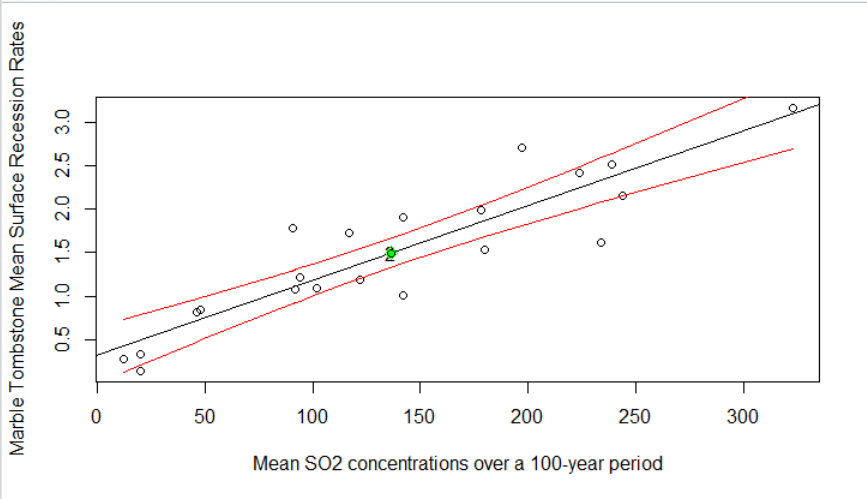
Using the predict method where we specify the confidence interval, we arrive at three sets of values – fit, lower and upper. These are the lower and upper limit of the fitted values. The fitted values are the mean of the lower and upper values.   
**-------------------------------------------------------------------------------------------------------------------------------**

3.4. Plot data points, the regression line, the confidence interval for fitted values (to show that the interval is wider on both sides and narrow in the center).

**Solution –**



**Output –**



**-------------------------------------------------------------------------------------------------------------------------------  
Question 4-** Using the output from summary(), suppose we want to test for null hypothesis of H\_0: β\_1=0.01 against the alternative hypothesis H\_1: β\_1≠0.01, what do you conclude? Reject or not reject? Explain why.

**Solution –** β\_1=0.01

T = (0.0085933 – 0.01)/0.0009499 = 1.48. Absolute Value |T| = 1.48

Since the absolute value of T is less than 2, we cannot reject null hypothesis since there is no linear relation between covariate and response variable.

**-------------------------------------------------------------------------------------------------------------------------------**

**Question 5** - Using the output from summary(), suppose we want to test for null hypothesis of H\_0: β\_1=0.02 against the alternative hypothesis H\_1: β\_1≠0.02, what do you conclude? Reject or not reject? Explain why.

**Solution –** β\_1=0.02

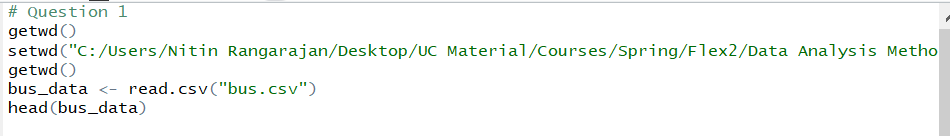
T = (0.0085933 – 0.01)/0.0009499 = 1.48. Absolute Value |T| = 12.008

Since the absolute value of T is greater than 2, we can reject null hypothesis since there is a linear relation between covariate and response variable

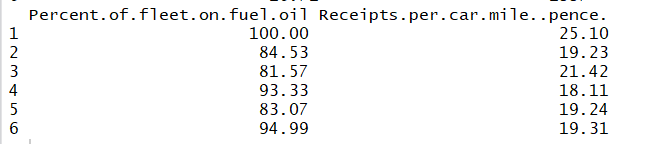
**-------------------------------------------------------------------------------------------------------------------------------**

**Question 6** -. Repeat the same questions (1-3) for the date set <bus.csv>. Description: Cross-sectional analysis of 24 British bus companies (1951). Use response variable = Expenses per car mile (pence), covariate = Car miles per year (1000s).

**Question 6.1 – Solution –**



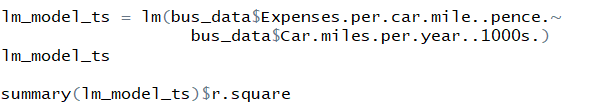
**Output:**



**-------------------------------------------------------------------------------------------------------------------------------**

**Question 6.2** - Obtain , explain what it means.

**Solution –**



**Output –**  value is 0.1582



value is an indication of how close the data are to the fitted regression line. So, in this case there is a weak dependency (15.8%) between covariate and response variable.

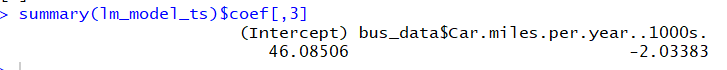
**-------------------------------------------------------------------------------------------------------------------------------  
Question 6.3** - Perform the following hypothesis testing and interval estimation using lm() and other related R functions.

6.3.1. Perform t tests, obtain t statistics and p values, interpret the results, make a conclusion (i.e. reject or not reject) and explain why. Note: please explain what the null hypothesis is.

**Solution –**



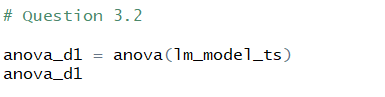
**Output –**



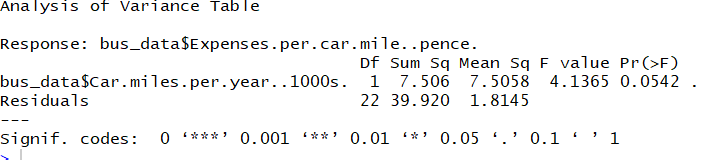
Null hypothesis is a test for determining the relationship between covariate and response variable. If we reject null hypothesis, then the variables have a linear relationship. Here, T-values are 46.08 and -2.03. Beta1 value is equal to 2 and P value is greater than 0.05. Thus, we cannot reject null hypothesis. **----------------------------------------------------------------------------------------------------------------------------**

6.3.2. Perform ANOVA test (F test), obtain F statistic and p value, interpret the results, make conclusion (i.e. reject or not reject) and explain why. Note: please explain what the null hypothesis is.

**Solution –**



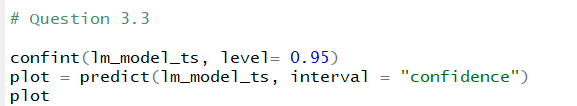
**Output –**



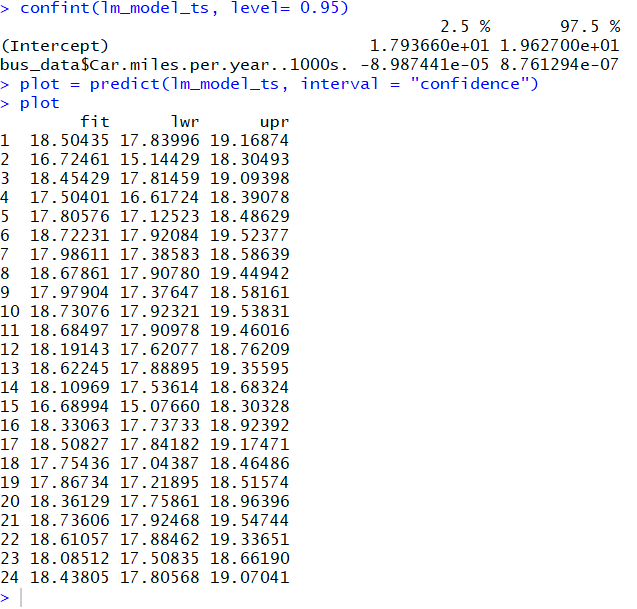
Null hypothesis is a test for determining the relationship between covariate and response variable. If we reject null hypothesis, then the variables have a linear relationship. Here, we can see that the F-value is lesser than F-critical and thus, we cannot reject null hypothesis.   
**-------------------------------------------------------------------------------------------------------------------------------**

6.3.3. Compute confidence interval for coefficients, fitted values (mean response), interpret the meanings of these quantities.

**Solution –**



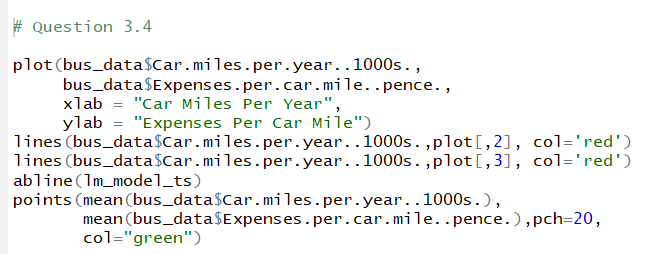
**Output –**



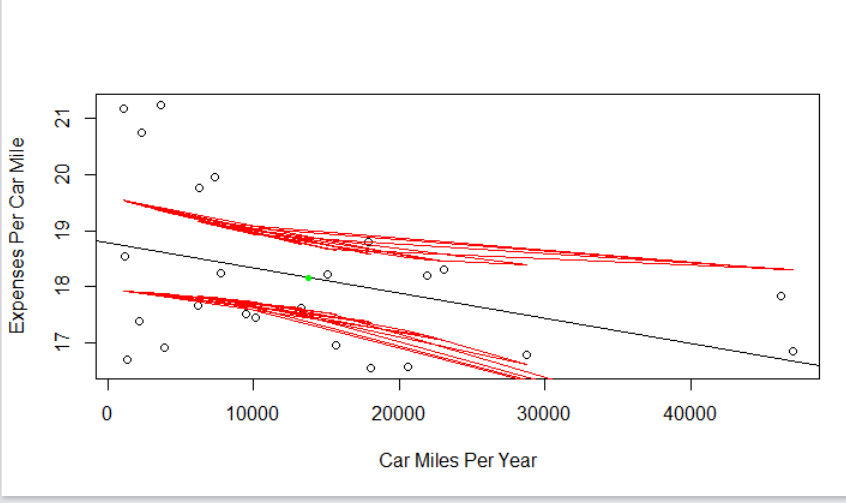
Using the predict method where we specify the confidence interval, we arrive at three sets of values – fit, lower and upper. These are the lower and upper limit of the fitted values. The fitted values are the mean of the lower and upper values.   
**-------------------------------------------------------------------------------------------------------------------------------**

6.3.4. Plot data points, the regression line, the confidence interval for fitted values (to show that the interval is wider on both sides and narrow in the center).

**Solution –**



**Output –**



**-------------------------------------------------------------------------------------------------------------------------------**