First Name: Susheela

Last Name: Polepalli

M-Number: M10727836

Homework 3

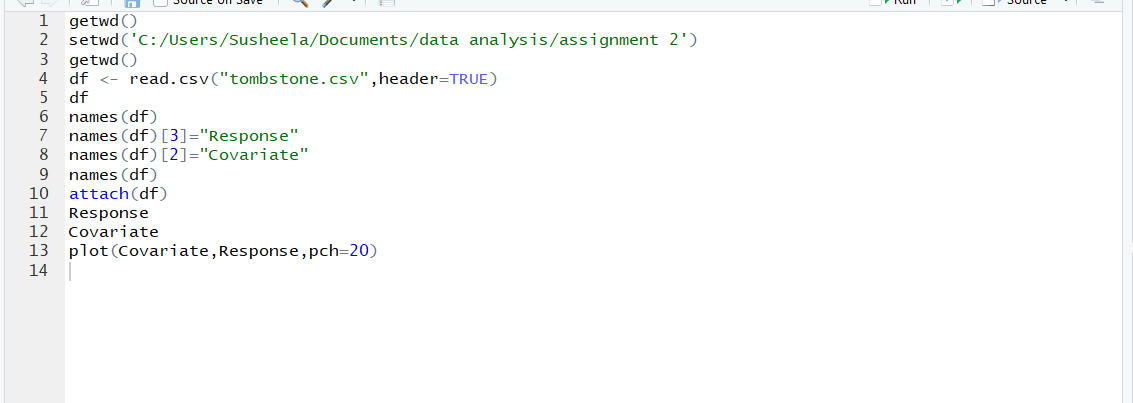
BANA7038

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:**

Solution:

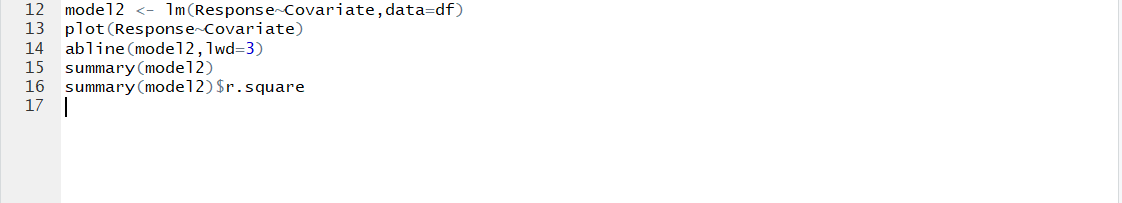
The following code is used to read the data and attach the variables to the workspace.



1. **2. Obtain , explain what it means.**

**Solution:**

**Code:**



**Output:**



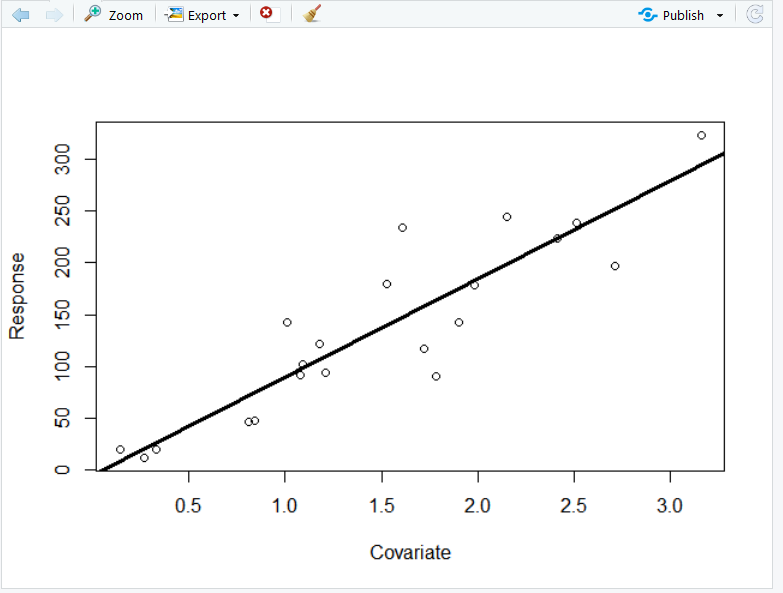
**Observations of R2:**

Since R^2 is a proportion, it is always a number between 0 and 1.

If R^2 = 1, all of the data points fall perfectly on the regression line. The predictor x accounts for all of the variation in y.

If R^2 = 0, the estimated regression line is perfectly horizontal. The predictor x accounts for none of the variation in y.

The R2=0.8115724 means that the data points are close to the fitted regression line.



**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:**

**Code:**



**Output:**

**­­**

The coefficient t-value is a measure of how many standard deviations our coefficient estimate is far away from 0. We want it to be far away from zero as this would indicate we could reject the null

We reject the null hypothesis and conclude that there is a relationship between response and covariate as the t value is greater than 0 and corresponding p values are less which leads to rejection of null hypothesis.

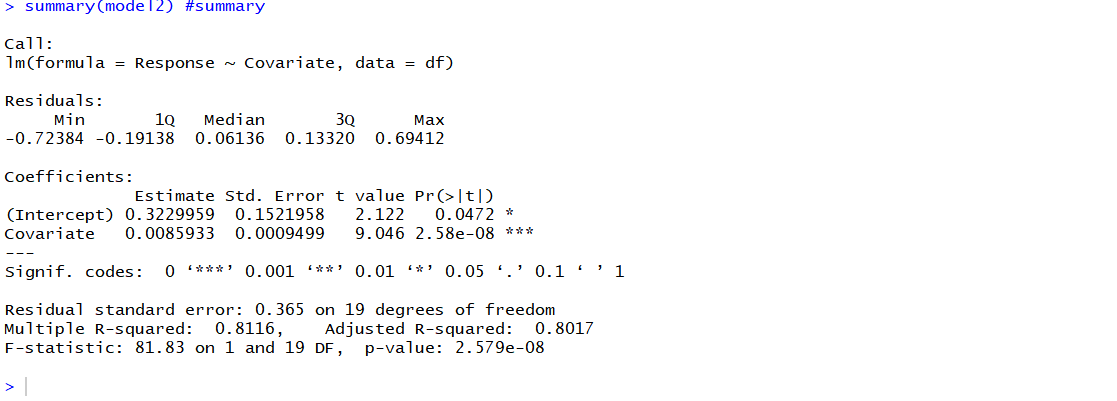
**3.2. Perform ANOVA test (F test), obtain F statistic and p value, interpret the results, make conclusion (i.e. reject or not reject) and explain why. Note: please explain what the null hypothesis is.**

**Solution:**

**Code:**



**Output:**



F statistic=81.83 and p value=2.579e-08

Null Hypothesis H0:There is no relationship between response and covariate is rejected as the F statistic is > 1.

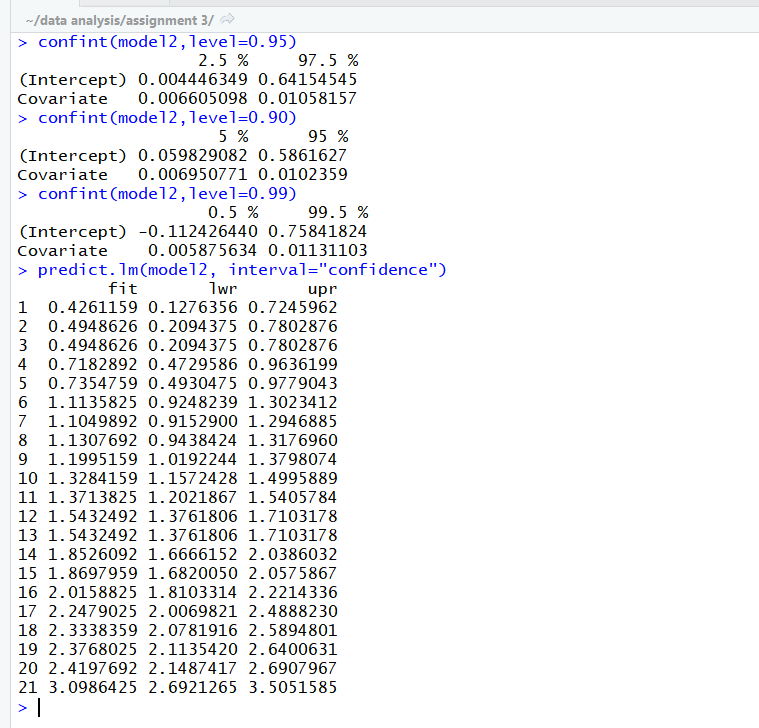
**3.3. Compute confidence interval for coefficients, fitted values (mean response), interpret the meanings of these quantities.**

**Solution:**

**Code:**



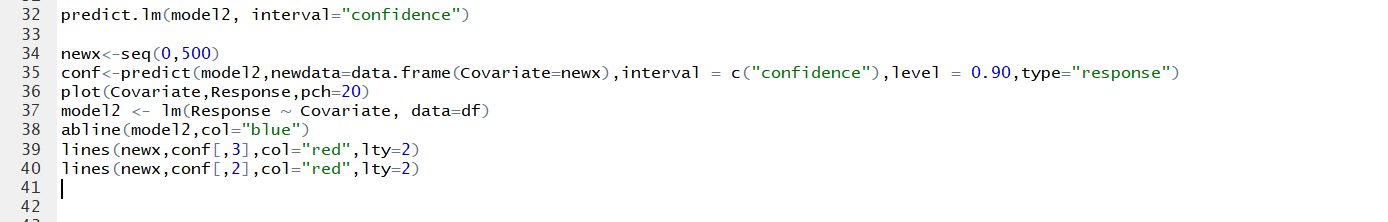
**Output:**



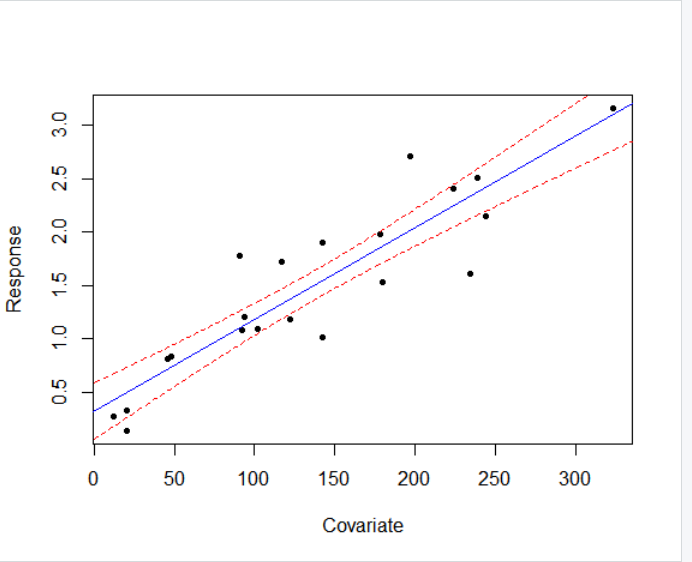
**3.4. Plot data points, the regression line, the confidence interval for fitted values (to show that the interval is wider on both sides and narrow in the center**).

**Solution:**

**Code:**



**Output:**

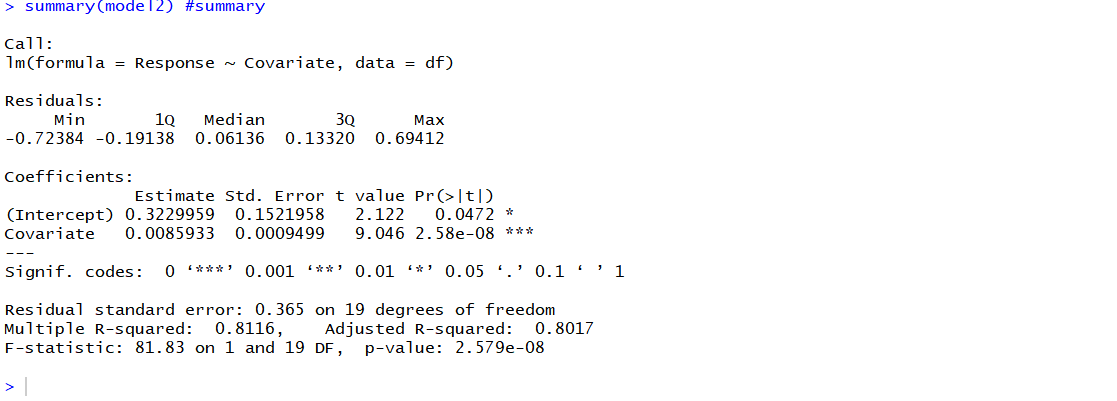


**4. Using the output from summary(), suppose we want to test for null hypothesis of against the alternative hypothesis , what do you conclude? Reject or not reject? Explain why.**

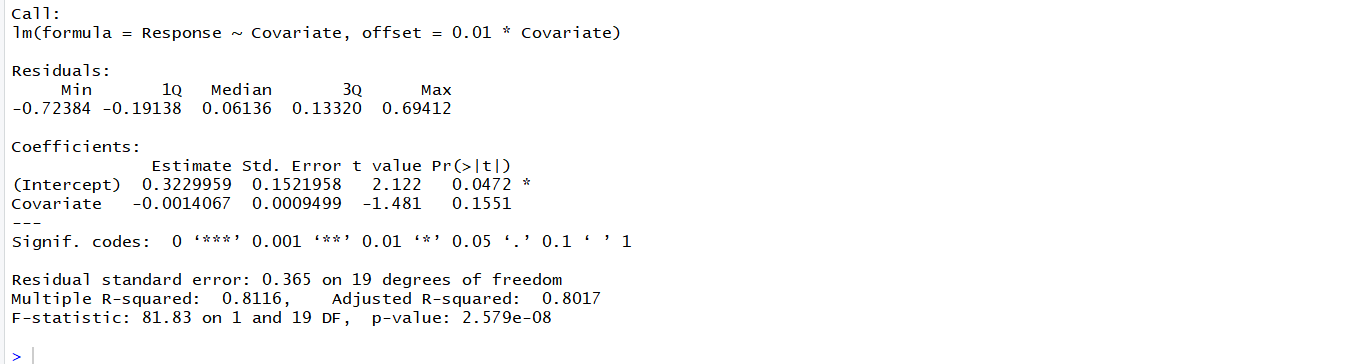
Solution:

The code:



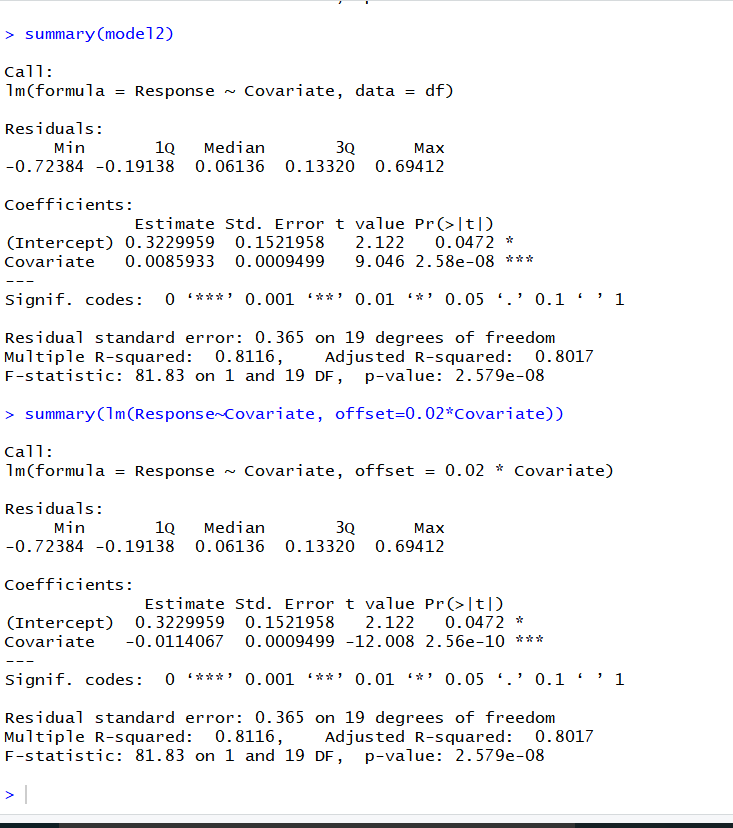






The t value -1.481 < 0 Thus we reject the hypothesis H0: **.**

**5. Using the output from summary(), suppose we want to test for null hypothesis of against the alternative hypothesis , what do you conclude? Reject or not reject? Explain why.**



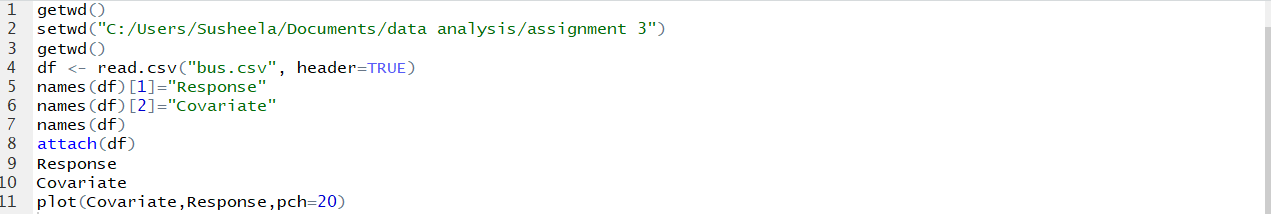
We reject the null hypothesis as the corresponding t value

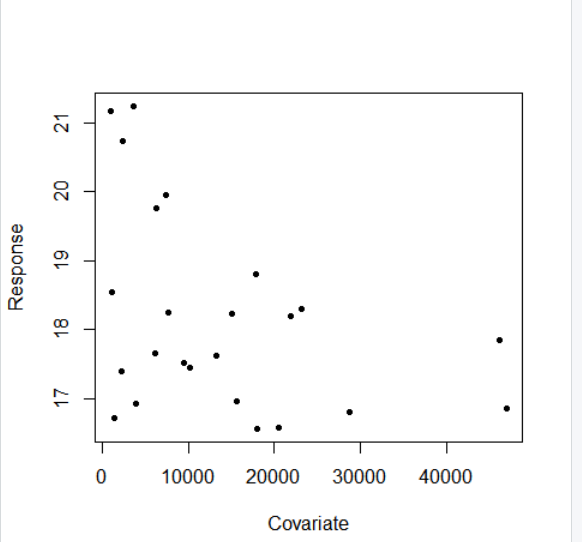
-12.008 < 0.

**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).**

**6.1. Read <bus.csv> into R. Use response variable = Expenses per car mile (pence), covariate = Car miles per year (1000s).**

**Solution:**





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

**Solution:**

**Code:**



**Output:**



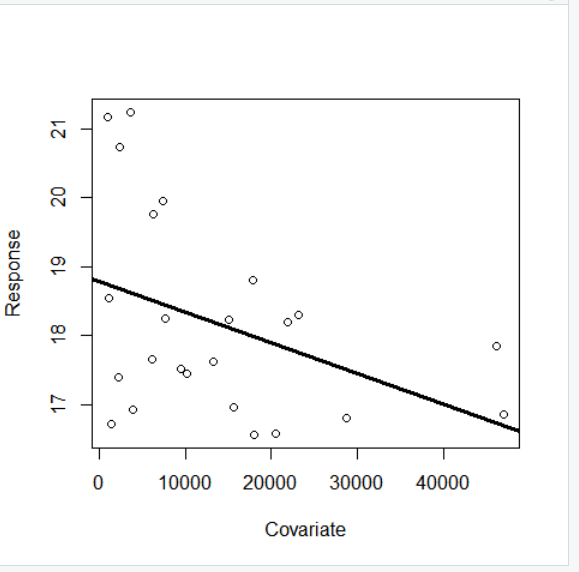
**Observations of R2 value:**

Since R^2 is a proportion, it is always a number between 0 and 1.

If R^2 = 1, all of the data points fall perfectly on the regression line. The predictor x accounts for all of the variation in y.

If R^2 = 0, the estimated regression line is perfectly horizontal. The predictor x accounts for none of the variation in y.

The R2=0.1582641 means that the data points are far from the fitted regression line.



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

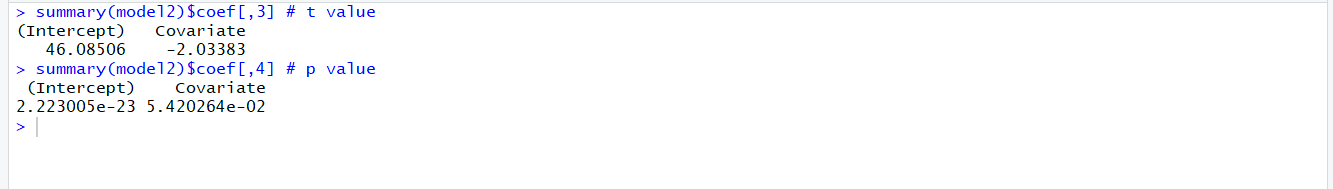
**6.3.1. Perform t tests, obtain t statistics and p values, interpret the results, make a conclusion (i.e. reject or not reject) and explain why. Note: please explain what the null hypothesis is.**

**Solution:**

**Code:**



**Output:**



The Null hypothesis H0: =0 is rejected.

The coefficient t-value is a measure of how many standard deviations our coefficient estimate is far away from 0. We want it to be far away from zero as this would indicate we could reject the null

We reject the null hypothesis and conclude that there is a relationship between response and covariate as the t value is greater than 0 and corresponding p values are less which leads to rejection of null hypothesis.

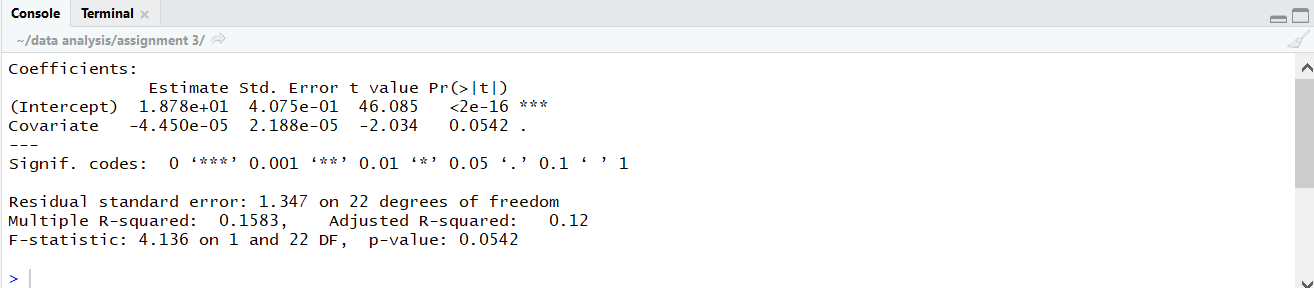
**6.3.2. Perform ANOVA test (F test), obtain F statistic and p value, interpret the results, make conclusion (i.e. reject or not reject) and explain why. Note: please explain what the null hypothesis is.**

**Solution:**

**Code:**



**Output:**



The Null hypothesis H0: =0 is rejected.

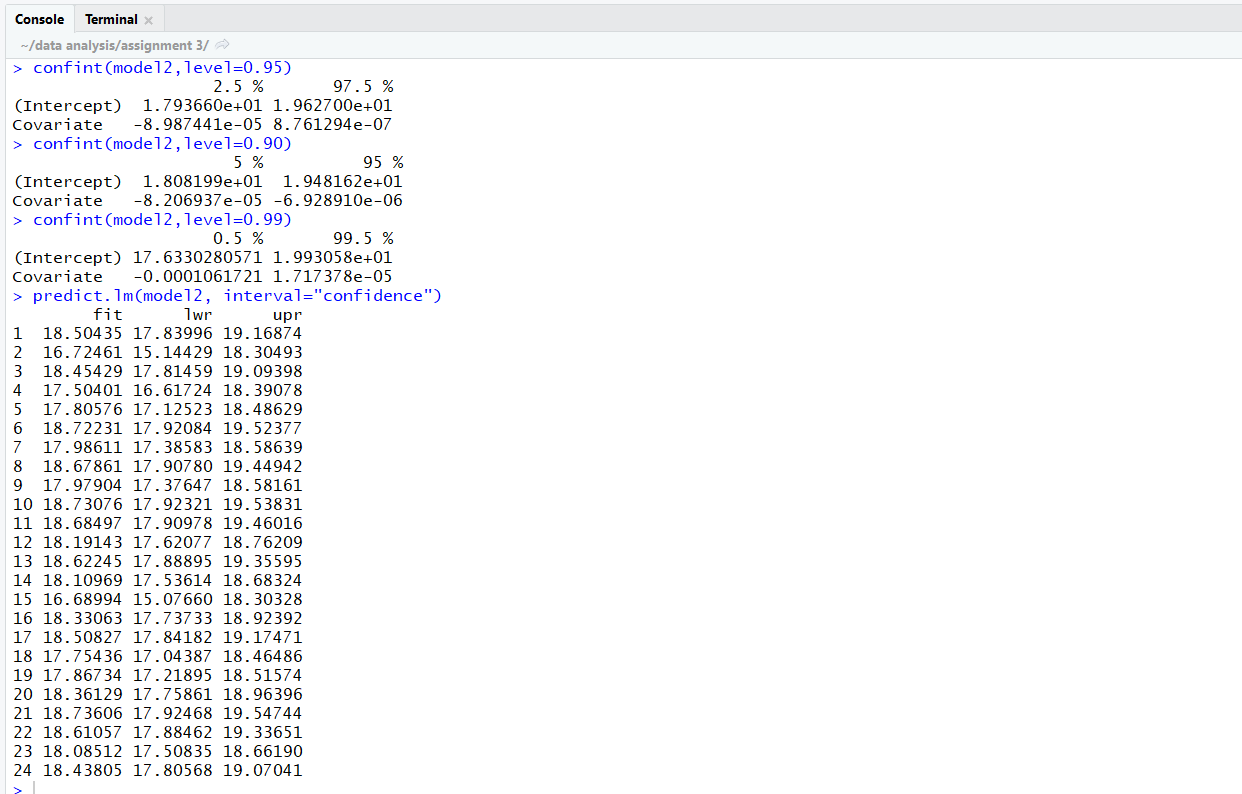
**6.3.3. Compute confidence interval for coefficients, fitted values (mean response), interpret the meanings of these quantities.**

**Solution:**

**Code:**



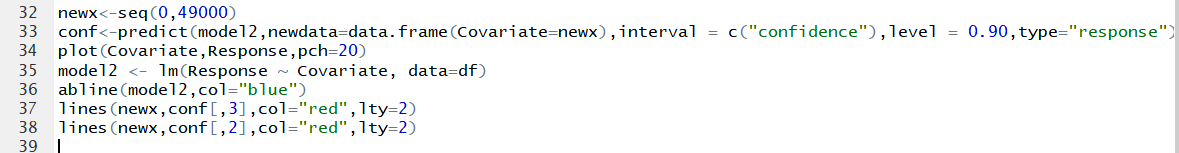
**Output:**



**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:**

**Code:**



**Output:**

