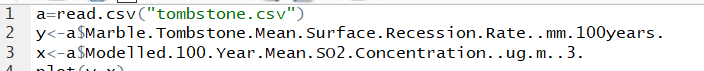
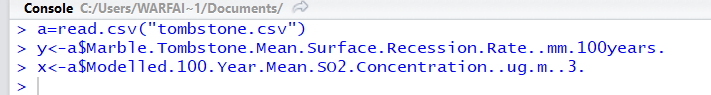
**Part A**

**Q1:** Read <tombstone.csv> into R. response variable = Marble Tombstone Mean Surface Recession Rate, and covariate = Mean SO2 concentrations over a 100 year period.

R-code:



R-output:

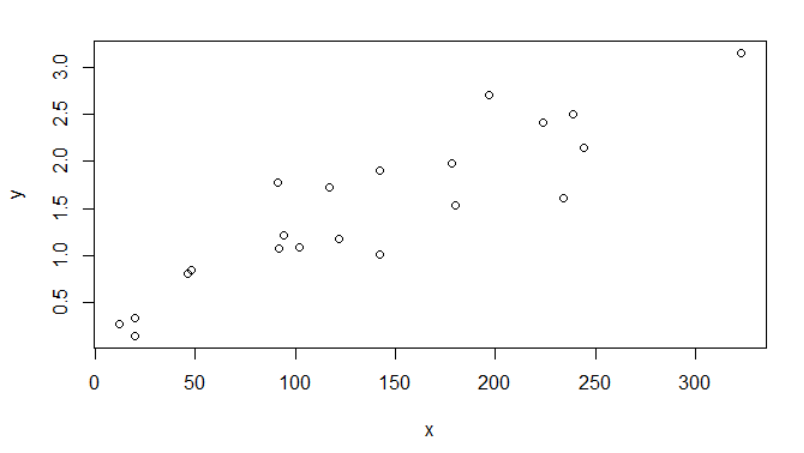


**Q2**. Plot data and briefly describe what you observe.

R-code:



R-output:

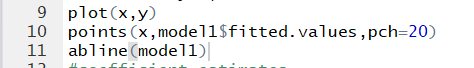


From the plot it looks like that as X increases y also increases so there is a direct(monotonous) relation between the response and covariate variable.

**Q3**. Perform linear regression using lm() function

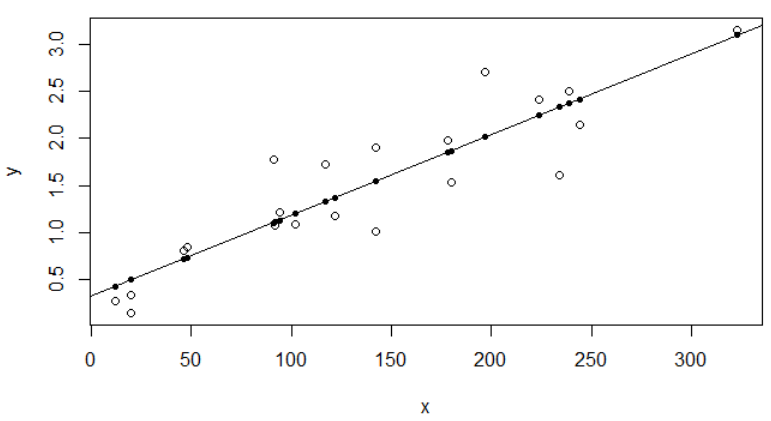
R-code:





R-output:





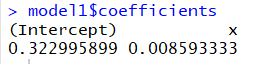
The solid line is the linear regression model of the input tombstone data

**Q3.1**. Obtain coefficient estimates , .

R-code:



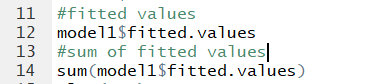
R-output:



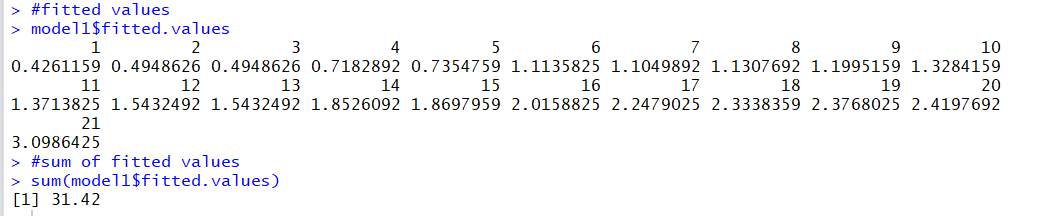
Here we found =0.32299 & =0.008593

**Q3.2**. Obtain fitted values and the sum of fitted values.

R-code:



R-output:



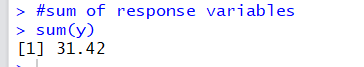
Sum of fitted values=31.42 which should be equal to sum of response values.

**Q3.3**. Obtain the sum of all values of response variable.

R-code:



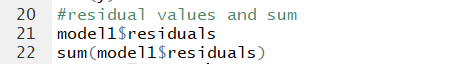
R-output:



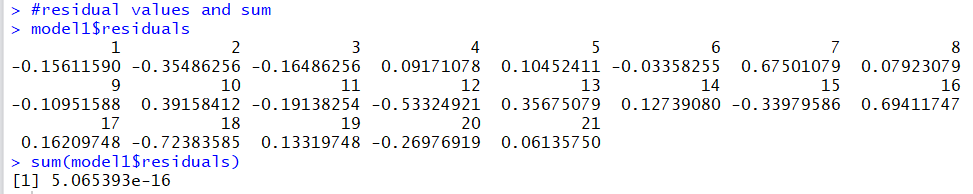
It can be clearly seen that sum of response variable is equal to sum of fitted values.

**Q3.4**. Obtain residuals and the sum of residuals.

R-code:



R-output:



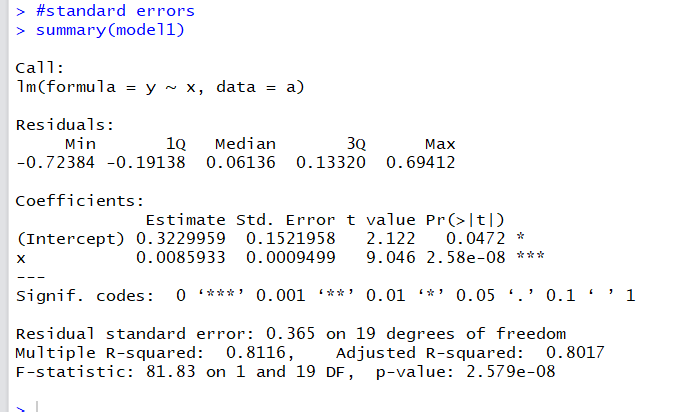
As expected sum of the residual values is approximately equal to 0.

**Q3.5**. Obtain the standard errors of , .

R-code:



R-output:



The standard error for =0.1521958 & =0.0009499

**Q4.** Suppose we increase SO2 Concentration by one unit, how does such a change influence the Surface Recession Rate?

As we know that =0.008593 so with a unit increase in SO2 concentration, there will be a 0.008593 unit increase in the Mean Surface Recession Rate.

**Q5**. Does the intercept of the linear regression have natural interpretation? If so, what does it mean?

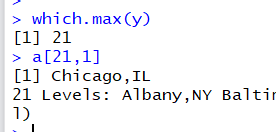
If 100-Year Mean SO2 Concentration (ug/m\*\*3)=0 then the Surface Recession Rate(mm/100 years)=0.32299

**Q6**. Which city (i.e., observation) has the highest Surface Recession Rate?

R-code:

>which.max(y) gives the row index of city with highest Surface Recession Rate

R-output:



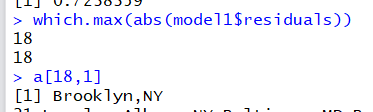
Chicago is the city with highest Surface Recession Rate

**Q7**. Which city (i.e., observation) has the largest residual (i.e., the largest absolute value) according to the linear regression you just fitted?

R-code:



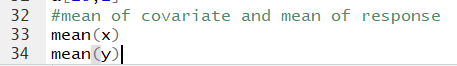
R-output:



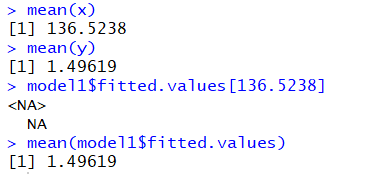
Brooklyn is the city with largest residual value of 0.723

**Q8**. Calculate the mean of covariate and mean of response. Verify the fact that the fitted regression line go through the point .

R-code



R-output



**Part B**

**Q1:** Read <bus.csv> into R. Use response variable = Expenses per car mile (pence), covariate = Car miles per year (1000s).

R-code:

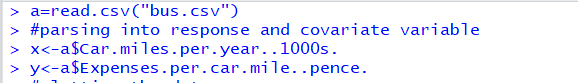
a=read.csv("bus.csv")

#parsing into response and covariate variable

x<-a$Car.miles.per.year..1000s.

y<-a$Expenses.per.car.mile..pence.

R-output:

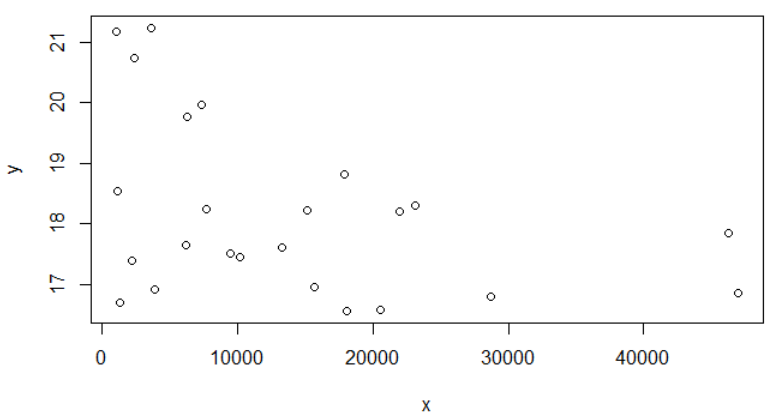


**Q2**. Plot data and briefly describe what you observe.

R-code:



R-output:

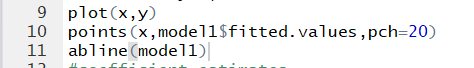


From the plot it looks like that as X decreases y also decreases so there is a direct(monotonous) relation between the response and covariate variable.

**Q3**. Perform linear regression using lm() function

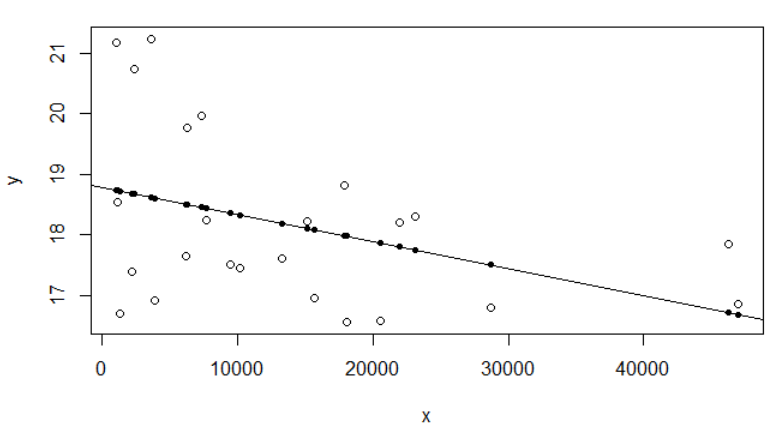
R-code:





R-output:





The solid line is the linear regression model of the input bus data

**Q3.1**. Obtain coefficient estimates , .

R-code:



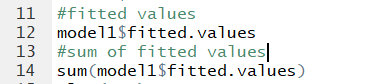
R-output:



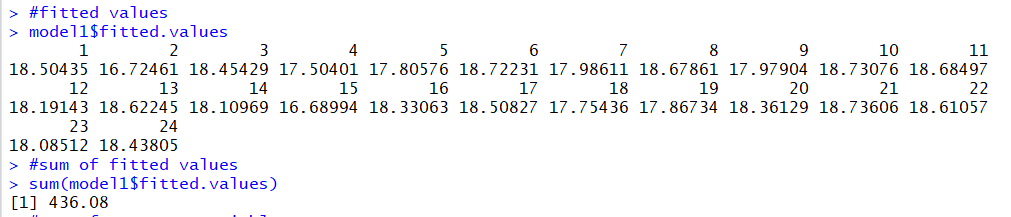
Here we found =18.7818 & = -0.00004449914

**Q3.2**. Obtain fitted values and the sum of fitted values.

R-code:



R-output:



Sum of fitted values=436.08 which should be equal to sum of response values.

**Q3.3**. Obtain the sum of all values of response variable.

R-code:



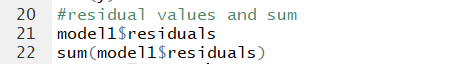
R-output:



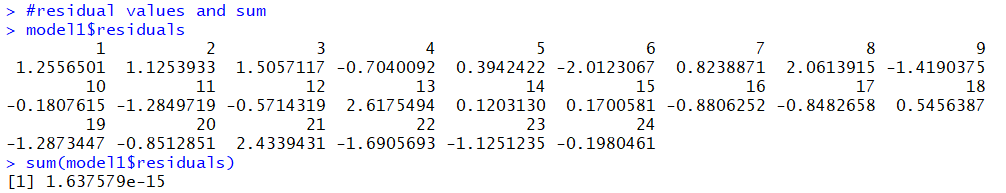
It can be clearly seen that sum of response variable is equal to sum of fitted values.

**Q3.4**. Obtain residuals and the sum of residuals.

R-code:



R-output:



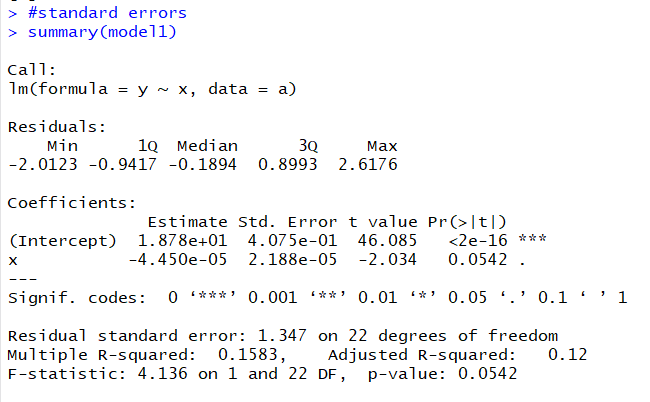
As expected sum of the residual values is approximately equal to 0.

**Q3.5**. Obtain the standard errors of , .

R-code:



R-output:



The standard error for = 0.4075 & = 2.188e-05

**Q4.** Suppose we increase Car miles per year by one unit, how does such a change influence the Expense per Car mile?

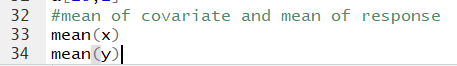
As we know that =-0.00004449914 so with a unit increase in Car miles per year, there will be a 0.00004449914 unit decrease in the Expense per car mile.

**Q5**. Does the intercept of the linear regression have natural interpretation? If so, what does it mean?

If Car miles per year (1000s) =0 then the Expenses per car mile (pence) =18.7818

**Q6**. Calculate the mean of covariate and mean of response. Verify the fact that the fitted regression line go through the point .

R-code



R-output

