**Module 5: Additional Exercises with Answers**

**Example1: Simple Regression Model (1 Variable)**

In this example, we will consider the case of simple linear regression with one response variable and a single independent variable. The data used for this example is from a study in central Florida where 15 alligators were captured and two measurements were made on each of the alligators. The weight (in pounds) was recorded with the snout vent length (in inches – this is the distance between the back of the head to the end of the nose).

The purpose of using this data is to determine whether there is a relationship, described by a simple linear regression model, between the weight and snout vent length. We first create a data frame for this study:

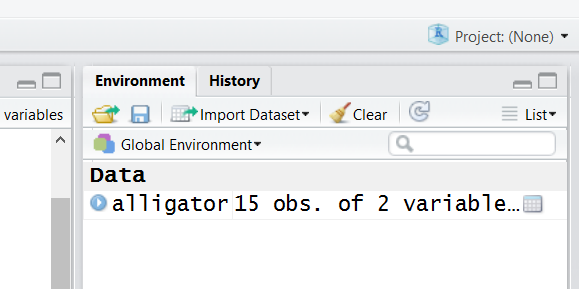
alligator = data.frame(

Length = c(47.94, 36.96, 75.94, 30.87, 45.15, 46.06, 31.81, 42.94, 33.11, 35.87, 66.02, 43.81, 40.85, 41.67, 43.816),

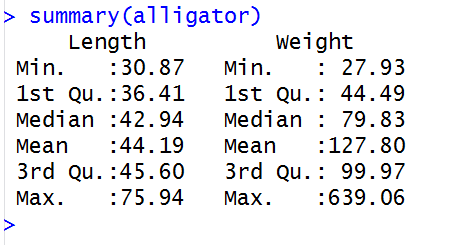
Weight = c(130.32,50.90, 639.06106, 27.93, 79.83, 109.94, 33.11, 90.01, 35.87, 38.09, 365.03, 83.93, 79.83, 83.09, 70.105)

)

By copy and pasting the above lines into the R console you should have a dataframe called alligator in your environment.



Let’s examine the data:



**Questions**

**Q1. Produce a plot of alligators’ weight against their snout vent length.**

**Q2. Build a simple linear regression model to express the alligators’ weight based on their snout vent length.**

**Q3. Based on the summary() output, what is the accuracy of this model?**

**Q4. Is Length (snout vent length) has a statistically significant relationship with weight?**

**Q5. What is the F-statistic is testing?**

**Q6. Formulate the equation of the weight of the alligators based on their snout vent length. Use the formula to predict the weight of a given alligator which has a snout vent length of 48in.**

**Q7. The coefficient of Length (snout vent length) is 12.646. What does it mean?**

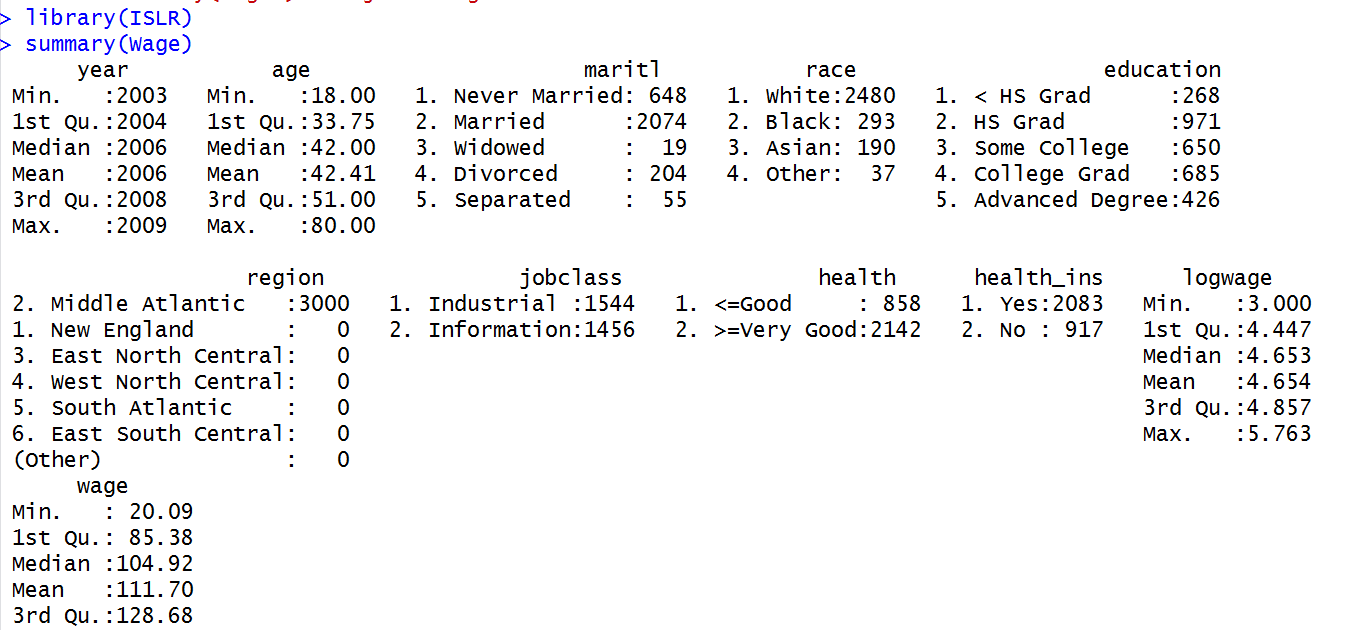
**Q8. Plot the residuals of the model against Length values. Based on this, do you think a simple linear regression model was a right choice here?**

**Example 2: Multiple Regression Model (Multiple Variables)**

Let’s see if we can build a model to predict the wage of individuals. We use the ‘Wage’ dataset available in the ‘ISLR’ package. If you have not installed the ‘ISLR’ package you can do so using

install.packages(‘ISLR’)

The following shows the summary of the dataset



**Questions**

**Q1. Build a model to predict the wage based o the age. How accurate is this model?**

**Q2. Try to improve your model by additionally including the martial staus. Does this improve the model accuracy?**

**Q3. Try to improve your model by additionally including the education. Does this improve the model accuracy?**

**Q4. What is the order of importance of variables?**

**Q5. Can you tell what the value of R2 is by simply looking at the anova output?**