Unit 1 Application

## For the next few questions, we’ll be using datasets from your book that are included in the ISLR package. Make sure you have it installed.

## Question 1

Use ?? to find out information about the R functions, mean, var, median, and sd.

1. What are the arguments for the functions?
2. What data types work with these funcitons?
3. How are NA’s handled?

### Answer (text only for this response)

## Question 2

Create a random numeric vector of size 100 using the function rnorm(100,1,2). a. Calculate the mean and standard deviation of your vector. b. Do your answers to a. match the arguments of rnorm()? Are they close?

### Answer

## Question 3

Using the dataset Auto, do the following:

1. Use glimpse to determine the types of the variables in Auto. Make note of which variables are factors.
2. Select only the variables mpg, horsepower, and displacement.
3. Create a scatter plot matrix using
   1. Base R with pairs
   2. library GGally and ggpairs
4. Do you notice any patterns?

### Answer

## Question 4

Using the full dataset Auto, do the following:

1. Use group\_by and summarize to find the mean and standard deviation of mpg by cylinders.
2. What pattern do you see? Are there any groups that break that pattern?

### Answer

## Question 5

Using the full dataset Auto, do the following:

1. Use filter to find all the autos cylinders greater than 3.
2. Create a scatterplot of mpg and displacement using the filtered data.
3. Change the color of the points to represent # of cylinders. (Hint: take a look a the data type of cylinders)

### Answer

## Question 6

Using the full dataset Auto, do the following:

1. Use lm() to perform a simple linear regression with mpg as the output and displacement as the input.
   1. Is there a relationship between the displacement and mpg?
   2. Is the relationship statistically significant?
   3. Interpret the parameter estimate associated with displacement.
2. Use augment from the broom package to add the residuals and other diagnostics.
3. Create a scatter plot of the residuals with respect to displacement.
   1. Do you suspect any high leverage points or outliers?
   2. Use filter to see if there are any observations with a std.resid greater than 3.

### Answer

## Question 7

Using the full dataset Auto, do the following:

1. Create four models for mpg as a function of displacement (X) with
   * X
   * log(X)
2. Using the modelr package and gather\_residuals() find the mse and mae for each of the above models.

### Answer

## Question 8

Using the full dataset Auto, do the following:

1. Use lm() to perform a multiple linear regression with mpg as the output and displacement, horsepower, and weight as the inputs.
   1. Are there relationship between displacement, horsepower, and weight and mpg?
   2. Are the relationships statistically significant?
   3. Interpret the parameter estimates.
2. Use augment from the broom package to add the residuals and other diagnostics.
3. Create a scatter plot of the residuals with respect to displacement.
   1. Do you suspect any high leverage points or outliers?
   2. Use filter to see if there are any observations with a std.resid greater than 3.
   3. Using the function vif from the car package (make sure you install it) calculate the variance inflation factors for displacement, horsepower, and weight.
   4. Is multicolinearity a potential issue? What effects would this have on parameter estimates?

### Answer