

# R Practice #1

*Ryan Safner*

*September 12, 2018*

## Getting Set Up

Before we begin, start a new file with **File** → **New File** → **R Script**. As you work through this sheet in the console in R, also add (copy/paste) your commands that work into this new file. At the end, save it, and run to execute all of your commands at once.

## Creating Objects

1. Create a vector called “me” with two objects, your first name, and your last name. Then call the vector to inspect it. Confirm it is a character class vector.
2. Create a vector called “x” with all the even integers from 2 to 10.
3. Find the mean of x with `mean()`
4. Now take the following pdf of random variable  $Y$ :

$y_i$	$p_i$
2	0.50
4	0.25
6	0.25

Calculate the standard deviation “manually” using our table method. You can look at the source code of Lecture 4 for my example.

- a. Creating two vectors, one called *y.i* and one called *p.i*, with the data above.
  - b. Merge them into a data frame called *rv* with `data.frame(y.i,p.i)`. Call *rv* to inspect it.
  - c. Find the expected value of *Y* by taking the sum of each value of *y.i* multiplied by *p.i* with the `sum()` command.
  - d. Creating a new column in *rv* called *deviations*, where you subtract the mean from each *y.i* value. Call *rv* again to make sure it's now there.
  - e. Create another column in *rv* called *devsq*, where you square the deviations from part d. Call *rv* again to make sure it's now there.
  - f. Now add another column in *rv* called *weighteddevsq*, where you multiply the squared deviations in part e. by the associated probability *p.i*. Call *rv* again to make sure it's now there.
  - g. Finally, take the sum of *weighteddevsq* to get variance. Square root this to get standard deviation.
5. The mean height of adults is 65 inches, with a standard deviation of 4 inches. Use the normal distribution to find the probabilities of the following scenarios:
- a. Find the probability of someone being *at least* 60 inches tall using `pnorm()`.
  - b. Find the probability of someone being *at most* 60 inches tall.
  - c. Find the probability of someone being between 61 and 69 inches tall. Why is this number familiar?
  - d. Find the probability of someone being between 57 and 73 inches tall. Why is this number familiar?

## Playing with a Data Set

For the following questions, use the `diamonds` dataset, included as part of `ggplot2`.

6. Install `ggplot2`
7. Load `ggplot2` with the `library()` command
8. Get the structure of the `diamonds` data.frame. What are the different variables and what kind of data does each contain?
9. Get summary statistics for `carat`, `depth`, `table`, and `price`
10. `color`, `cut`, and `clarity` are categorical variables (factors). Use the `table()` command to generate frequency tables for each.
11. Now rerun the `summary()` command on the entire data frame
12. Plot a histogram of `price`.
13. Plot a boxplot of `price` by `diamond color`.

### Execute your R Script

Save the R Script you created at the beginning and (hopefully) have been pasting all of your valid commands to. This creates a `.R` file wherever you choose to save it to. Now looking at the file in the upper left pane of *R Studio* look for the button in the upper right corner that says **Run**. Sit back and watch R redo everything you've carefully worked on, all at once.