Lab

# Lab 3 – Distributions and Functions: Generate Distributions

A key focus of our class this week has been “distributions.” A distribution is simply an arrangement of values of a variable such as the population size of a state. A “probability distribution” is an arrangement of all the values (potential outcomes) of a variable that reflect the frequency of those values in nature. A distribution can either be empirical, which means that it is an actual bunch of numbers, or it can be theoretical, in which case we are just imagining an ideal arrangement of numbers. The normal or “bell” curve is just such a theoretical distribution.

The R open source statistical system is great at creating empirical distributions that are made up of randomly generated numbers. In Chapter 9, the book includes several commands and explanations of randomly generated distributions. Chapter 9 decided that the distribution of the population sizes of states was “geometric,” and while this is one possibility, a better choice would be the so-called “Pareto” distribution. We can use R to generate a Pareto distribution of state populations that may be quite similar to the populations of the actual U.S. states. In other words, we can generate random numbers for the sizes of the Fictional States of America.

**Task 1**: Write, test, and submit the necessary code in R to accomplish the following:

1. Generate a normal distribution, or 1,000 samples, with a mean of 80.

2. Write a function that takes three variables – a vector, a min and a max – and returns the number of elements in the vector that are between the min and max (including the min and max).

3. Use the function to see how many of your normal distribution samples are within the range of 79 to 81.

4. Repeat 3 times (creating a normal distribution and then calling your function) to see if the results vary.

**Task 2**: Write, test, and submit the necessary code in R to accomplish the following:

1. Generate 51 random numbers in a Pareto distribution and assign them to a variable called “FSApops.”

2. Specify a “location” and a “shape” for your Pareto distribution that makes it as similar as possible to the actual distribution of state populations on page 44 of the textbook.

3. Create a histogram that shows the distribution of values in FSApops.

4. Use a command to report the actual mean and standard deviation of the 51 values stored in FSApops.

5. Use a command to report the population of your largest fictional state (i.e., your California) and your smallest fictional state (i.e., your Wyoming).

**Hints**: The necessary R command for generating random numbers in a Pareto distribution is located in a “package” called “VGAM.” The code you submit should include the two necessary commands for making this happen (they should be the first two commands in your code). In the VGAM package, you will find a command for generating random numbers that fit a Pareto distribution. You will have to look up the meaning of the location and shape parameters so that you can figure out how to set them to make your Fictional States of America as similar to the real states as possible. Note that random numbers differ substantially every time you run the command, so we don’t expect your data to be a perfect match. You do want your smallest state to be about the size of Wyoming and about 15 of your states to be under 2 million in population.