Statistics Project for STA-271-03

By Henry Reichard

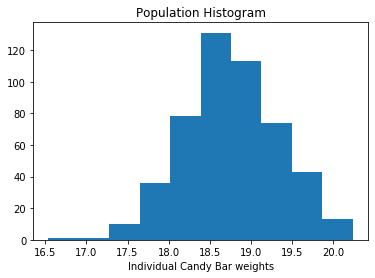
Project: Written in Python using Numpy library in Jupyter Notebook

**Introduction**

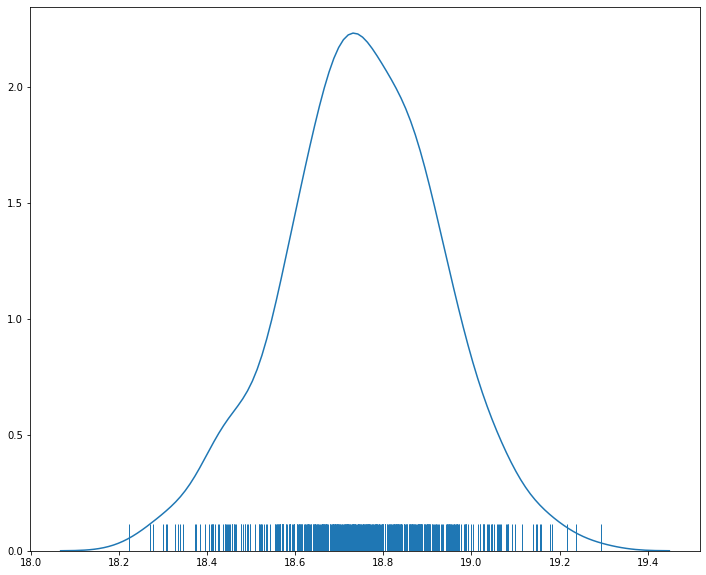
This project made a simulation of the weights of candy bars with a normal distribution and mean of 18.79 grams and standard deviation of 0.58 grams. The simulation was done in Python using the Numpy library in a Jupyter notebook. The weights of 500 candy bars are stored in an array and a histogram was made. Afterwards 500 samples of size 10 was drawn from the population data frame and was itself was stored in an array. A histogram of the sample data frame was constructed. The histograms of the population and sample was compared, and sample mean, and sample standard deviation was calculated. The population histogram, mean, and standard deviation was compared to the sample histogram, mean, and standard deviation.

**Project**

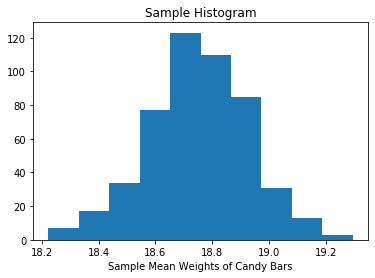
After making the simulation of the 500 candy weights in a normal distribution with the mean of 18.79 and standard deviation of 0.58, the simulation has given the population mean of 18.75 and the population standard deviation was 0.57. The population histogram looks like this:



Next thing 500 samples of size 10 was created from the population and the distribution looks normal and majority of the samples are in the curve



The samples histogram. It looks close to the population histogram, but it looks like more data is centered around the mean. By the central limit theorem, the distribution of many sample means, known as a sampling distribution, will be normally distributed. Also when the population distribution is normal, the sample distribution will be normal as well.



Additionally, the mean of the sample is 18.75. The difference between the population mean - sample means are 0.007, the difference appears not to be that significant. The reason why the sample mean is close to the population mean because sample mean is an unbiased estimate of the population mean. This means that the sample mean is not systematically smaller or larger than the population mean.

Furthermore, the standard deviation of the sample is 0.18. The difference between population standard deviation and sample deviation is 0.39, the difference appears to be big. This is because the sample standard deviation is divided by the square root of n. This will make the sample standard deviation smaller when compared to the population standard deviation.

Now let’s see if the sample mean falls within a 95% confidence interval for the population mean. The population confidence interval is (18.70,18.80) and the sample mean is 18.75. The sample mean is within the confidence interval.