**Provided Exercise**

Generate artificial data, then, while treating this data as a sample of a larger population, calculate the sample mean, median, 10% trimmed mean and 35th percentile. Finally, plot this data using a histogram and describe the shape.

**Solution**

**Part (a) – Generate Data**

Data can be generated in R using a number of distributions. This example will utilize a normal distribution and the following code:



The rnorm function above will generate 1000 points of data with a mean of 0 and a standard deviation of 1.

**Part (b) – Calculate Sample Mean**

Let x1,x2,...,xn denote n observations. The sample mean, , can be calculated using the formula below:

Using R software, we compute the sample mean as follows:



For the generated data, the mean is -0.02582443.

**Part (c) – Calculate Sample Median**

Once again, let x1,x2,...,xn denote n observations, this time arranged in increasing order of magnitude. Given this, the sample median, , can be calculated using the formula below:

In R, this can be calculated with the following code:



For the generated data, the median is -0.01313418.

**Part (d) – Calculate 10% Trimmed mean**

The 10% trimmed mean, , is calculated by trimming or removing both the largest and smallest 10% of the data before calculating the mean. In R, this can be done by adding an additional argument to the previously used mean function, as shown in the following code:



For the generated data, the 10% trimmed mean is -0.02582443.

**Part (e) – Calculate 35th Percentile**

A percentile is defined as a value on a scale of one hundred that indicates the percent of a distribution that is equal to or below it. As an example, a percentile score of 95 is a score equal to or better than 95 percent of the scores. In R, the 35th percentile can be calculated using the following code:



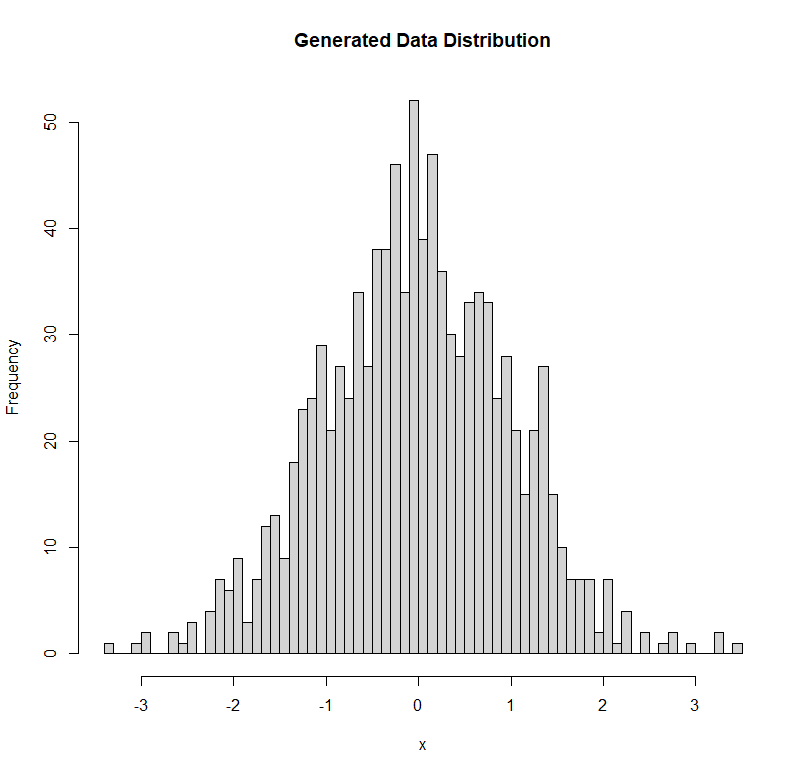
For the generated data, the 35th percentile is -0.3826663.

**Part (f) – Plot a Histogram and Describe**

A histogram is a diagram consisting of rectangles whose area is proportional to the frequency of a variable and whose width is equal to the class interval. They are used to visualize the distribution of a dataset. This is done with the following R code:



This code produced the chart below:



The chart above shows that the generated data largely follows a standard normal distribution. The mean appears to match the expected value of 0, and the frequency of generated data points noticeably decrease as the distance from this mean increases. Said another way, most of the data points are clustered towards the center of the distribution, with data becoming more and more scarce towards the tails.

**Summary**

In this exercise we generate a random dataset that followed a normal distribution with a mean of 0 and a standard deviation of 1. Using this data, we then calculated the sample mean of the dataset, -0.02582443, the sample median, -0.01313418, the 10% trimmed mean, -0.02582443, and the 35th percentile, -0.3826663. Then we generated a histogram from this dataset and confirmed that data did follow a normal distribution.