**3.1 Descriptive Statistics**

Samples of data: Part of data out of full range of data. There is always uncertainty in it, hence, sampling needs to be done so that it represents all the data with as much accuracy possible.

Types of descriptive statistics:

* Mean
* Median
* Mode
* Range
* Variance
* Standard deviation

Measure of central tendency: mean, median, mode

Measure of dispersion(indicates the spread of the data set): range, standard deviation, variance

Histograms: show both central tendency and dispersion together. They show frequencies of occurrences on the y-axis with the range or classes on the x-axis. In R, the function to generate histogram from a data set is

* hist(data set)

Normal distribution: Frequently data tends to form a bell curve. In R, the function to generate a dataset with normal distribution is

* rnorm(number of samples, mean, standard deviation)

**3.2 Functions**

3 components of a function are:

1. Name
2. Arguments(can be zero or more): Functions takes some parameters to be able to perform operations and return some value
3. Return: output of the function

It is a bundle of code that can be used again and again without retyping.

Some functions are inbuilt in R, however we have to define functions that are not inbuilt. The syntax for writing a function that is not inbuilt in R is:

* FunctionName <- function(input\_arguments)  
  {

return(output\_value)

}

Syntax for calling a function in R is:

* FunctionName(input\_arguments)

**3.3 Data Science in the Real World**

Data Science can be used for various reasons in different scenarios. It can be used to optimize the running time for a machine by making an algorithm understand when the machine looks like it would break down in the near future. It can also be used in the supply chain, by making sure the stocking reflects the users’ needs.

**Questions from the videos:**

* Why is understanding central tendency and measure of dispersion helpful?
* Both of the qualities give meaningful insights about the data. Only knowing central tendency is not useful in terms of being able to understand the range of data and how much it spreads, whereas knowing the measure of dispersion cannot produce a singular value for performing operations. Also, i believe, knowing both the central tendency and measure of dispersion can also allow us to understand how the outliers in the data should be treated.

**Questions for Professor:**

* Where would mode be a useful parameter in practical analysis?