

Rayanna Harduarsingh

September 9th, 2020

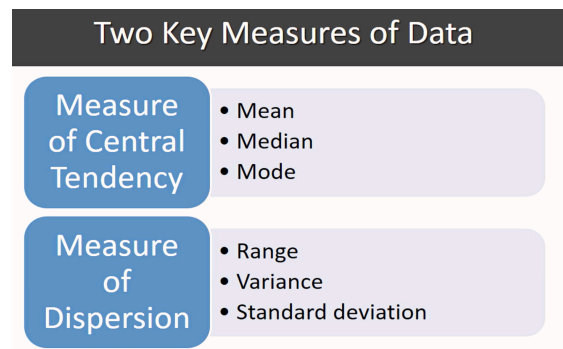
Module 3 Notes

IST 687

Lecture 3.1- Descriptive Statistics

Descriptive statistics include mean, media, range, mode, variance, and standard deviation of a data set.

- Whenever we have data, it's a sample.
- The mean adds together all the observations and divides it by number of observations.
- The media is calculated by sorting the ,list of all of the observations in the sample and then going halfway up the list. The halfway point is the median
- The range is measure of dispersion, how spread out a bunch of numbers in a sample (subtract the lowest from the highest)
- The mode is the value that occurs the most often.
- Variance is a measure of dispersion. It describes how spread out a sample of numbers is. It is obtained from all of the numbers through a simple calculation that compares each number to the mean. Calculation: obtain the deviations from the mean, then square them, add all of them up, then divide by the number of observations.
 - **R Calculation for variance:**
 - `var(myFamily$myFamilyAges)`



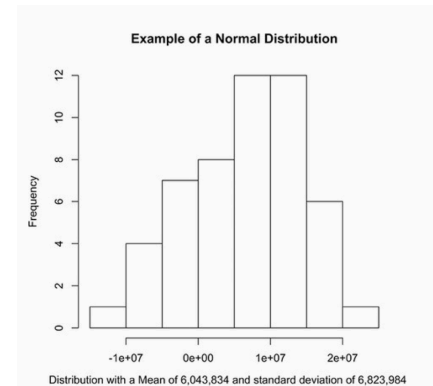
- The standard deviation is another measure of dispersion. It is the square root of the variance.

- **R Calculation for SD:**

- `sd(myFamily$myFamilyAges)`

- **Normal distribution** is used for making comparisons.

- The distribution had a characteristic (shape)
- The distribution had a center point (mean)



Lecture 3.2- Writing Functions

- Functions provide two key benefits:
 - We do not have to type the same R code over and over again (easily reuse code).
 - We do not need to understand the details of how the function was written (abstraction).

Writing a 'MyMode' function:

```
MyMode <- function(myVector) # create function MyMode
{
  return(myVector)
}
```

Annotations:

- Function name: `MyMode`
- Input argument: `myVector`
- Send back result of function: `return(myVector)`
- Curly brackets contain function code: `{ }`

From textbook:

```
MyMode(tinydata)
```

```
"MyMode <- function(myVector)
```

```
{
```

```
  uniqueValues <- unique(myVector)
```

```
  return(uniqueValues)
```

```
}
```

Lecture 3.3- Data Science in the Real World

Question: Provide some examples of Data Science in the Real World

After my journey in Applied Data Science, my desired field is to go into marketing and utilize Data science. Marketers and Advertisers handle huge amounts of digital information with their campaigns. This data as well as consumer data is crucial for them to make business decisions. From the data they collect, they can service insights and conclusions on how to better appeal to their consumers by illustrating customer personas and customize customer experiences. Some data they can deal with could be from social media campaigns such as they number of impressions they've received, how many accounts they've reached, the click through rate, and more. This is all vital information and where data scientists are needed as these huge data sets need to be compressed.