3.1 Getting Help in R

Various ways to get in R. Following command will provide help in the Help window.

```
help(median)

## starting httpd help server ... done
```

Other option: This will search all functions and provide you various help pages.

```
??mean
```

Tip: If you have typed the term in the Console or in the Script window, place the cursor on the term and press F1. This will also open the help just like help() command.

3.2 Descriptive Statistics - Mean

To find mean of a number of items, the first step will be create a vector and then use the command mean()

```
student.height <- c(100, 106, 121, 111, 109, 111, 103, 117, 114)
mean(student.height)

## [1] 110.2222
```

Some times all values are not available. If there is no value for an item, it is stored as NA.

If there is an NA value in the vector, the mean will shown as NA.

```
student.height1 <- c(100, 106, 121, 111, 109, 111, 103, 117, 114, NA, 1047)
mean(student.height)

## [1] 110.2222
```

In case as above, you will need to remove NA from the calculation. That can be done by argument na.rm = TRUE

```
student.height1 <- c(100, 106, 121, 111, 109, 111, 103, 117, 114, NA, 1047)
mean(student.height, na.rm = TRUE)

## [1] 110.2222
```

In the above case there is an extreme value of 1047. In some cases you might want to remove these outliers from the calculation.

In R there is an option to trim the data. The argument trim = 0.1 (for example) will remove 10% data from the top and 10% from the bottom.

```
student.height1 <- c(100, 106, 121, 111, 109, 111, 103, 117, 114, NA, 1047)
mean(student.height, trim = 0.1, na.rm = TRUE)

## [1] 110.2222
```

3.2 Descriptive Statistics - Median and Mode

Median

To find the median of a number of items, use the command **median()**

```
student.height <- c(100, 106, 121, 111, 109, 111, 103, 117, 114)
median(student.height)

## [1] 111
```

Mode

There is no direct command in R to find the mode.Some external packages do have this option. But right no lets stick to R basic commands to find the mode.

In this simple case, if we create a table, it will show, which height is repeated most number of times

```
table(student.height)

## student.height
## 100 103 106 109 111 114 117 121
##   1   1   1   1   2   1   1   1
```

From this table you can see that the heigh 111 is repeated the most (2 times) number of times, hence 111 is the mode. You can sort this table to put 111 as the last item.

```
a <- table(student.height)
sort(a)

## student.height
## 100 103 106 109 114 117 121 111
##   1   1   1   1   1   1   1   2
```

Below code changes the sorting in decending order and [1] picks the first item from the table. Hence gives the mode

```
a <- table(student.height)
sort(a, decreasing = TRUE)[1]

## 111
##   2
```

3.4 Measurement of Variation

Range

Range provides the highest and the lowest value.

```
student.height <- c(100, 106, 121, 111, 109, 111, 103, 117, 114)
range(student.height)

## [1] 100 121
```

Quantile

This provide 0, 25, 50, 75 and 100 quartile. In this:

- 0% = Lowest value,

- 25% = First quartile (Q1)

- 50% = Median (Q2)

- 75% = Third quartile (Q3)

- 100% = Highest value

```
quantile(student.height)

##   0%  25%  50%  75% 100%
##  100  106  111  114  121
```

Inter Quartile Range (IQR)

```
IQR(student.height)

## [1] 8
```

Standard Deviation (Sample)

```
sd(student.height)

## [1] 6.64789
```