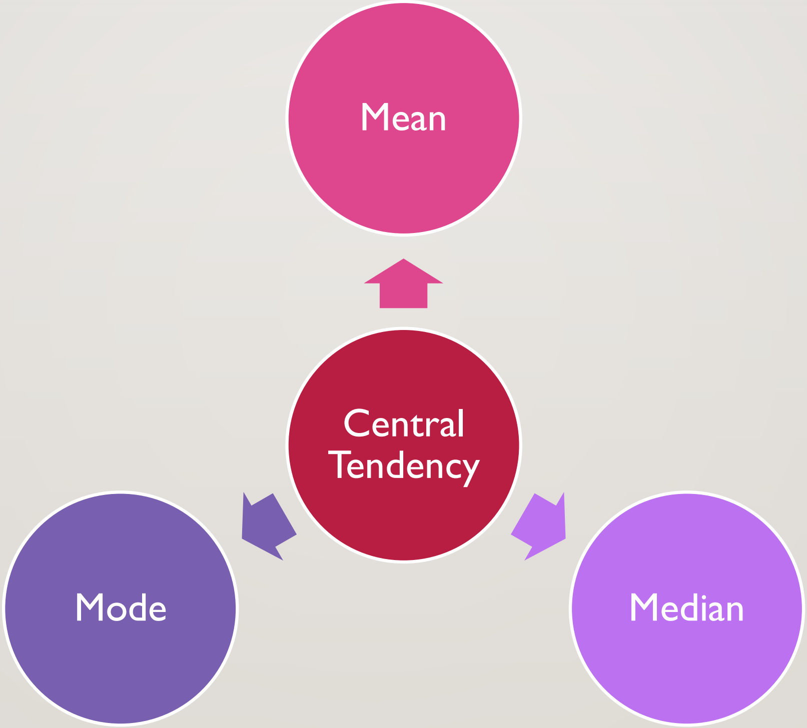
MEASURE OF CENTRAL TENDENCY

KUSHAL SHARMA – DATA SCIENCE PRACTITIONER

AGENDA

- A Quick Revision
- The most fundamental analytics
  - Mean
  - Median
  - Mode
- Demonstration based on dataset in R for statistical concepts

MEASURES OF CENTRAL TENDENCY



Central Tendency

Mean

Median

Mode

VIDEO REFERENCE

Measures of Central Tendency - https://www.youtube.com/watch?v=clqgY4TIDwY

Real Life Example - https://www.youtube.com/watch?v=_rywstqBqOs

R - MEAN, MEDIAN AND MODE

Statistical analysis in R is performed by using many in-built functions.

Most of these functions are part of the R base package.

These functions take R vector as an input along with the arguments and give the result.

MEAN

It is calculated by taking the sum of the values and dividing with the number of values in a data series.

The function mean() is used to calculate this in R.

Syntax

The basic syntax for calculating mean in R is −

mean(x, trim = 0, na.rm = FALSE, ...)

Following is the description of the parameters used −

x is the input vector.

trim is used to drop some observations from both end of the sorted vector.

na.rm is used to remove the missing values from the input vector.

EXAMPLE

```
# Create a vector.
x <- c(12,7,3,4.2,18,2,54,-21,8,-5)
# Find Mean.
result.mean <- mean(x)
print(result.mean)
```

When we execute the above code, it produces the following result −

```
[1] 8.22
```

APPLYING TRIM OPTION

When trim parameter is supplied, the values in the vector get sorted and then the required numbers of observations are dropped from calculating the mean.

When trim = 0.3, 3 values from each end will be dropped from the calculations to find mean.

In this case the sorted vector is (−21, −5, 2, 3, 4.2, 7, 8, 12, 18, 54) and the values removed from the vector for calculating mean are (−21,−5,2) from left and (12,18,54) from right.

APPLYING TRIM FUNCTION

```
# Create a vector.
x <- c(12,7,3,4.2,18,2,54,-21,8,-5)
# Find Mean.
result.mean <-  mean(x,trim = 0.3)
print(result.mean)
```

When we execute the above code, it produces the following result −

```
[1] 5.55
```

APPLYING NA OPTION

If there are missing values, then the mean function returns NA.

To drop the missing values from the calculation use na.rm = TRUE. which means remove the NA values.

APPLYING NA OPTION

```
# Create a vector.
x <- c(12,7,3,4.2,18,2,54,-21,8,-5,NA)
# Find mean.
result.mean <-  mean(x)
print(result.mean)
# Find mean dropping NA values.
result.mean <-  mean(x,na.rm = TRUE)
print(result.mean)
```

When we execute the above code, it produces the following result −

```
[1] NA
[1] 8.22
```

MEDIAN

The middle most value in a data series is called the median. The median() function is used in R to calculate this value.

Syntax

The basic syntax for calculating median in R is −

```
median(x, na.rm = FALSE)
```

Following is the description of the parameters used −

x is the input vector.

na.rm is used to remove the missing values from the input vector.

MEDIAN – EXAMPLE

```
# Create the vector.
x <- c(12,7,3,4.2,18,2,54,-21,8,-5)
# Find the median.
median.result <- median(x)
print(median.result)
```

When we execute the above code, it produces the following result −

```
[1] 5.6
```

MODE

The mode is the value that has highest number of occurrences in a set of data. Unlike mean and median, mode can have both numeric and character data.

R does not have a standard in-built function to calculate mode. So we create a user function to calculate mode of a data set in R. This function takes the vector as input and gives the mode value as output.

MODE – EXAMPLE

- # Create the function.
- getmode <- function(v) {
-    uniqv <- unique(v)
-    uniqv[which.max(tabulate(match(v, uniqv)))]
- }

- # Create the vector with numbers.
- v <- c(2,1,2,3,1,2,3,4,1,5,5,3,2,3)

- # Calculate the mode using the user function.
- result <- getmode(v)
- print(result)

- # Create the vector with characters.
- charv <- c("o","it","the","it","it")

- # Calculate the mode using the user function.
- result <- getmode(charv)
- print(result)
- When we execute the above code, it produces the following result −

- [1] 2
- [1] "it"

DISCUSSION
Q&A

THANK YOU!

KUSHAL@INDEEDINSPIRING.COM

9762203269