Statistics

Statistics

Statistics is a branch of Mathematics that deals with

- Collection
- Organization
- Presentation
- Analysis of numerical data &
- Making inferences from it.

1.Descriptive statistics

Descriptive statistics are brief descriptive coefficients that summarize a given data set

- Measures of Frequency: Count, Percent, Frequency
- Measures of Central Tendency: Mean, Median, and Mode. ...
- Measures of Variation: Range, Variance, Standard Deviation. ...
- Measures of Position: Percentile Ranks, Quartile Ranks.

2. Inferential Statistics

Used to make judgment, observe difference between groups Draw inferences, reach conclusion beyond the immediate data.

a) Hypothesis Testing
b) Correlation

Collecting Data

Sample Vs Population

- Population - Parameter
- Sample - Statistic

How to select samples

- Simple Random Sampling

It involves picking the desired sample size and selecting observations from a population in such a way that each observation has an equal chance of selection. For example, a random selection of 20 students from a class of 50 students gives a probability of selection being 1/50.

- Stratified Sampling

This technique divides the elements of the population into key subgroups or strata. The elements are randomly selected from each of these strata. For example, males under 30, females under 30, males 30 or over, and females 30 or over. Say you want to achieve a sample size of 200, then you can pick samples of 50 from each stratum

- Cluster Sampling

Cluster sampling divides the sample into a large number of subgroups. Then some of these subgroups are selected at random, and simple random samples are then collected within these subgroups. These subgroups are called clusters. It is done ease of collection data.

Types of Data

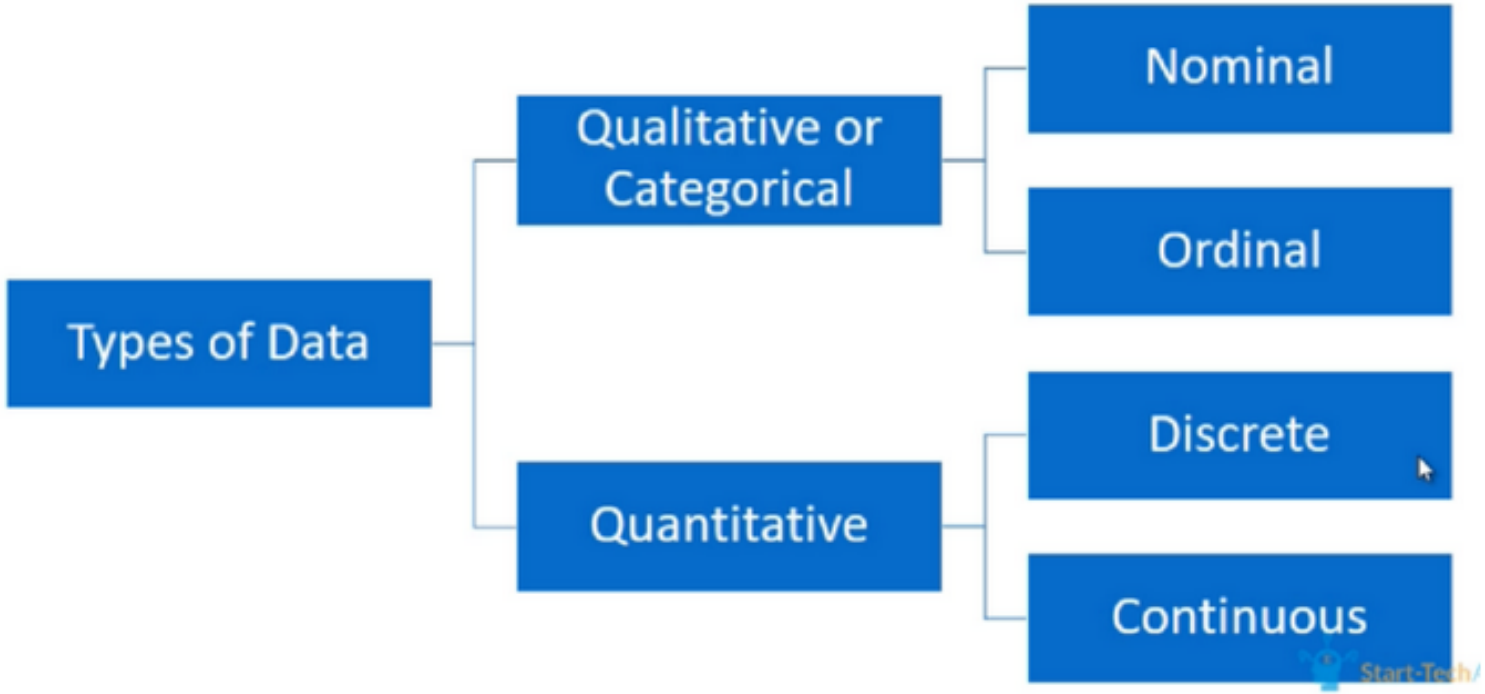

Image courtesy: https://medium.com/@rndayala/data-levels-of-measurement-4af33d9ab51a

Using the quantile() Command

```
>dat
[1] 1 2 3 4 5 6
 > quantile(dat)
0% 25% 50% 75% 100%
 1.00 2.25 3.50 4.75 6.00
```

Quantiles are cut points deviding the range of numbers

Using the quantile() Command

```
data2<-c(1, 5, 7 ,5, 3, 2, 6, 8, 5, 6)
quantile(data2, 0.3)
30%
4.4
>quantile(data2, c(0.2, 0.5, 0.8))
20% 50% 80%
 2.8 5.0 6.2
```

Cumulative Statistics

Simple Cumulative commands

cumsum(x)- The cumulative sum of a vector
cummax(x)- The cumulative maximum value
cummin(x)- The cumulative minimum value
cumprod(x)- The cumulative product

Examples

```
data2 <-  c(3, 5, 7, 5 ,3, 2, 6, 8, 5, 6, 9,4, 5, 7, 3, 4)
cumsum(data2)
 [1] 3 8 15 20 23 25 31 39 44 50 59 63 68 75 78 82
>cummax(data2)
 [1] 3 5 7 7 7 7 7 8 8 8 9 9 9 9 9 9
```

More Cumulative Commands

The seq() : generate sequences of values using from, to and interval

> seq(from = 1, to = 10, by = 2)

[1] 1 3 5 7 9

The following produces the same result:

> seq(1, 10, 2)

If you use the ‘along =‘ instruction you can create an index for a vector:

> seq(along = data2)

[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Running mean

cumsum(data2) / seq(along = data2)