# DATA ANALYTICS BOOTCAMP

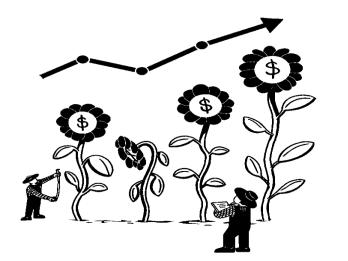
Uloh C. Armstrong



☐ MEANING AND SCOPE OF STATISTICS.

**Statistics** is a branch of mathematics and a scientific discipline concerned with collecting, organizing, analyzing, interpreting, and presenting data.

It involves methods and techniques used to gather, summarize, and draw conclusions or inferences from data.



### **Statistics**

[stə-'ti-stiks]

A branch of applied mathematics that involves the collection, description, analysis, and inference of conclusions from quantitative data.

### ☐ TYPES OF STATISTICS.

## **Descriptive statistics:**

Describes and summarizes data through measures such as mean, median, mode, etc.

## Inferential statistics:

Uses sample data to make inferences or predictions about a larger population

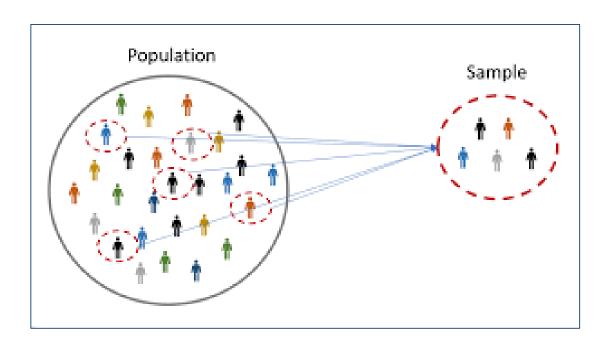
## ☐ POPULATION AND SAMPLE

## Population:

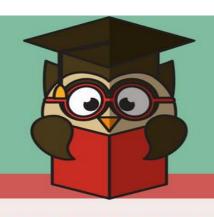
This is a set of all objects or units about which conclusions are to be drawn.

## Sample:

A subset of a population.



☐ STATISTICAL DATA.
☐ TYPES OF DATA.



# QUANTITATIVE VS QUALITATIVE DATA

### QUANTITATIVE DATA

Quantitative data can be expressed as a number or can be quantified. Simply put, quantitative data can be measured by numerical variables.

### QUALITATIVE DATA

Qualitative data can't be expressed as a number and can't be measured.

Qualitative data consist of words, pictures, and symbols, not numbers.







#### **EXAMPLES**

- Scores on tests and exams e.g. 85, 67, 90 and etc.
- The weight of a person or a subject.
- Your shoe size.
- The temperature in a room.

#### **EXAMPLES**

- Colors e.g. the color of the sea
- Your favorite holiday destination such as Hawaii, New Zealand.
- Names as John, Patricia,....
- Ethnicity such as American Indian, Asian, etc.



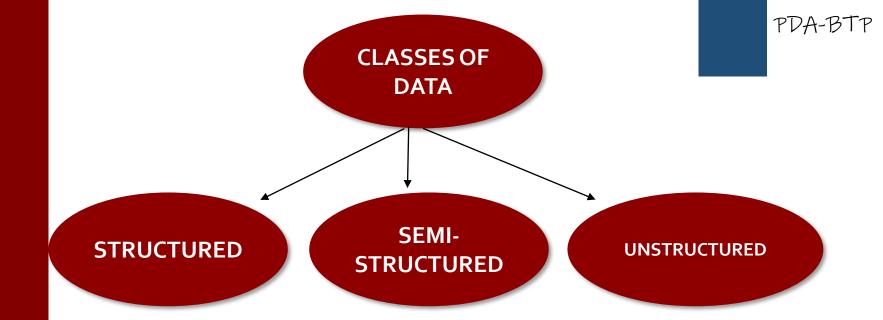


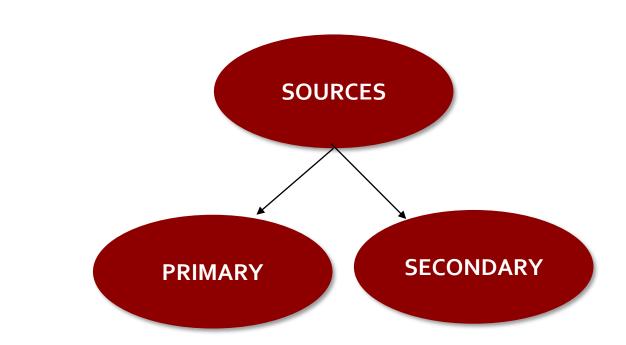


### intellspot.com

☐ Basic **STATISTICS** 

☐ CLASSES AND SOURCES OF DATA.



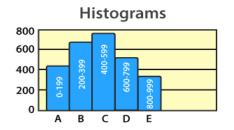


☐ GRAPHICAL

PRESENTATION OF DATA.

### **TYPES OF GRAPHICAL REPRESENTATION**

# Bar Graphs 800 600 400 200 0 A B B B C

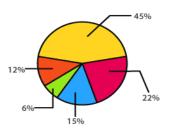


BYJU'S
The Learning App

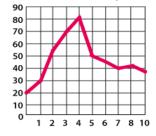
### Frequency Table

Rulers of France			
Reign (Years)	Tally	Frequency	
1-15	ווו זאע זאע זאע	18	
16-30	וזאע זאע	11	
31-45	ו זאג	6	
46-60	IIII	4	
61-75	1	1	

Circle Graph



### Line Graphs



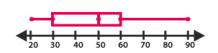
Stem and Leaf Plot

Stem	Leaf		
0	1, 1, 2, 2, 3, 4, 4, 4, 4, 5, 8		
1	0, 0, 0, 1, 1, 3, 7, 9		
2	5, 5, 7, 7, 8, 8, 9, 9		
3	0, 1, 1, 1, 2, 2, 2, 4, 5		
4	0, 4, 8, 9		
5	2, 6, 7, 7, 8		
6	3, 6		
Key: 6   3 = 63 Year			

### **Line Plot**



### Box and Whisker Plot



☐ ARITHMETIC MEAN.

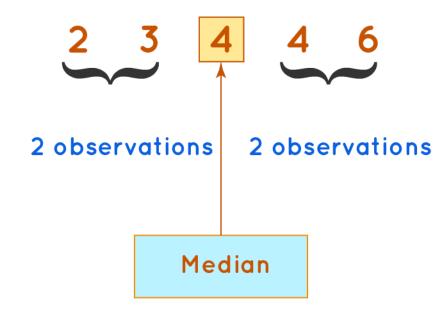
The arithmetic mean, commonly known as the average, is a measure of central tendency used to represent a set of numerical data

Arithmetic Mean = x1 + x2 + x3 + + xn / n				
	A	В	D	
2	Sr.No	Division A		
3	Student 1	56		
4	Student 2	60		
5	Student 3	56		
6	Student 4	64		
7	Student 5	70		
8	Student 6	55		
9	Student 7	50		
10				
	Number of	7		
11	Students (n)	,		
12	Arithmetic Mean	=(B3+B4+B5+B6+B7		
13		+B8+B9)/B11		

☐ MEDIAN.

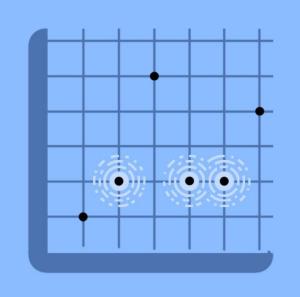
The **median** represents the middle value in a dataset when arranged in ascending or descending order.

### Median of Data



☐ MODE.

The **mode** is the value/values within a dataset that occur most frequently.



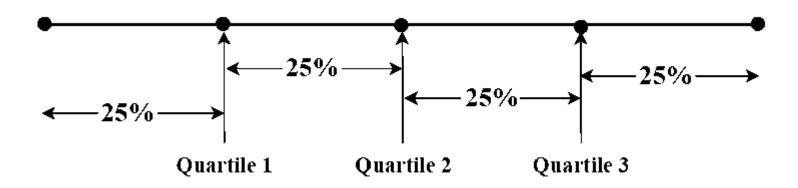
## Mode

['mōd]

The value that appears most frequently in a data set. A set of data may have one mode, more than one mode, or no mode at all.

**QUARTILES.** 

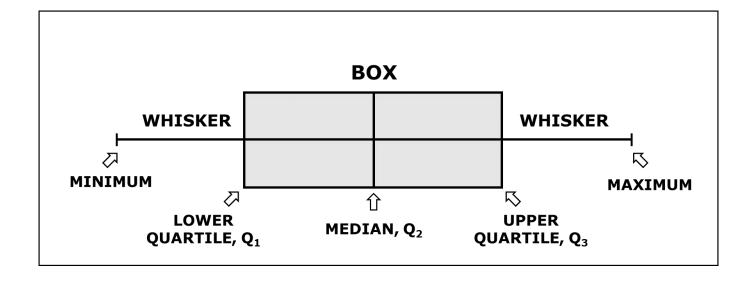
Quartiles are values that divide a dataset into four equal parts, each representing 25% of the data when arranged in ascending or descending order.



BOX & WHISKERS PLOT (BOX PLOT).

A box plot, also known as a box-andwhisker plot, is a graphical representation that displays the distribution, central tendency, and variability of numerical data through five summary statistics:

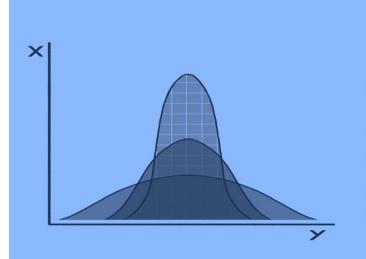
minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum



### ☐ Basic **STATISTICS**

☐ THE VARIANCE.

The **Variance** is a measure of how spread out the values in a data set are. It is closely related to the standard deviation, but it is expressed in a different unit of measurement.



### Variance

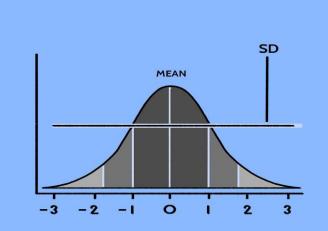
['ver- $\bar{e}$ - $\partial n(t)s$ ]

A measurement of how far each number in a data set is from the mean (average), and thus from every other number in the set.

### □ Basic **STATISTICS**

☐ STANDARD DEVIATION.

The **Standard deviation** is a measure of how spread out the values in a data set are. It is a way of quantifying the amount of variation in a set of data

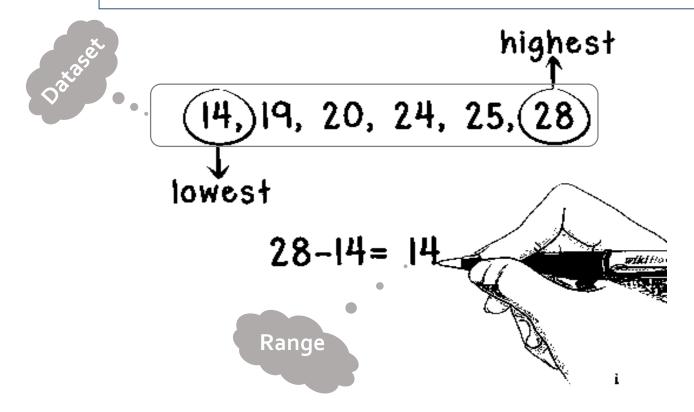


### **Standard Deviation**

['stan-dərd dē-vē-'ā-shən]

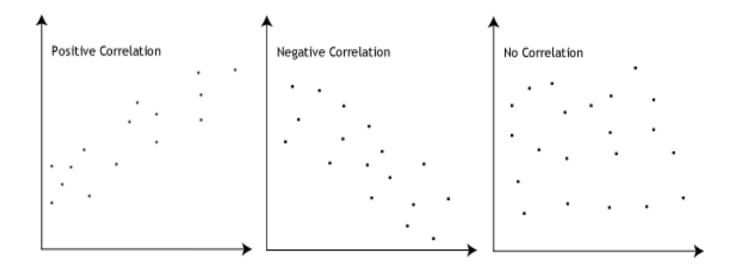
A statistic that measures the dispersion of a dataset relative to its mean and is calculated as the square root of the variance. ☐ RANGE.

The **range** in statistics refers to the difference between the highest and lowest values within a dataset.



☐ CORRELATION ANALYSIS.

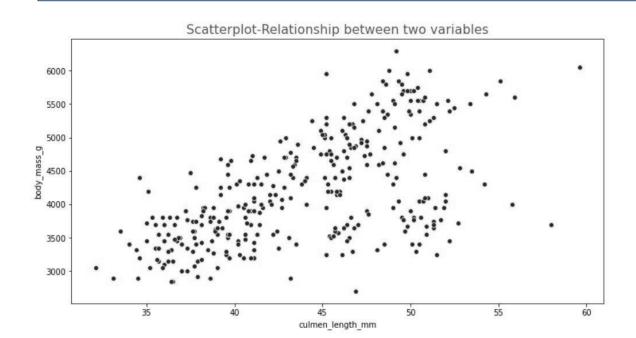
Correlation analysis is a statistical technique used to measure and describe the strength and direction of the relationship between two variables.



- ☐ Basic **STATISTICS**
- ☐ Meaning and scope of statistics.
  - ☐ Types of statistics.
    - Statistical data.
    - ☐ Types of data.
  - Classes and sources of data.
- ☐ Graphical presentation of data.
  - Arithmetic mean.
    - ☐ Median.
    - ☐ Mode.
    - Quartiles.
- Box & Whiskers plot (Box plot).
  - ☐ The Variance
  - Standard deviation.
    - ☐ Range.
  - ☐ Correlation analysis.
  - ☐ SCATTER DIAGRAM.
    - ☐ Regression Analysis

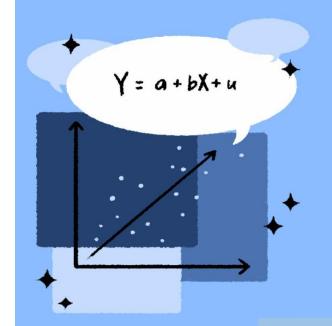
A scatter plot, also known as a scatter diagram or scatter graph, is a type of data visualization that displays the relationship or association between two numerical variables.

It helps to visualize how changes in one variable are related to changes in another variable.



☐ Regression Analysis

Regresssion analysis is a statistical method used to model relationship between two or more variables. It is a way of determining how a change in one variable is associated with a change in another variable.



## Regression

[ri-'gre-shan]

A statistical method used in finance, investing, and other disciplines that attempts to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables).

# DATA ANALYTICS BOOTCAMP

Uloh C. Armstrong

