Mathematical/Statistical Concepts for ML

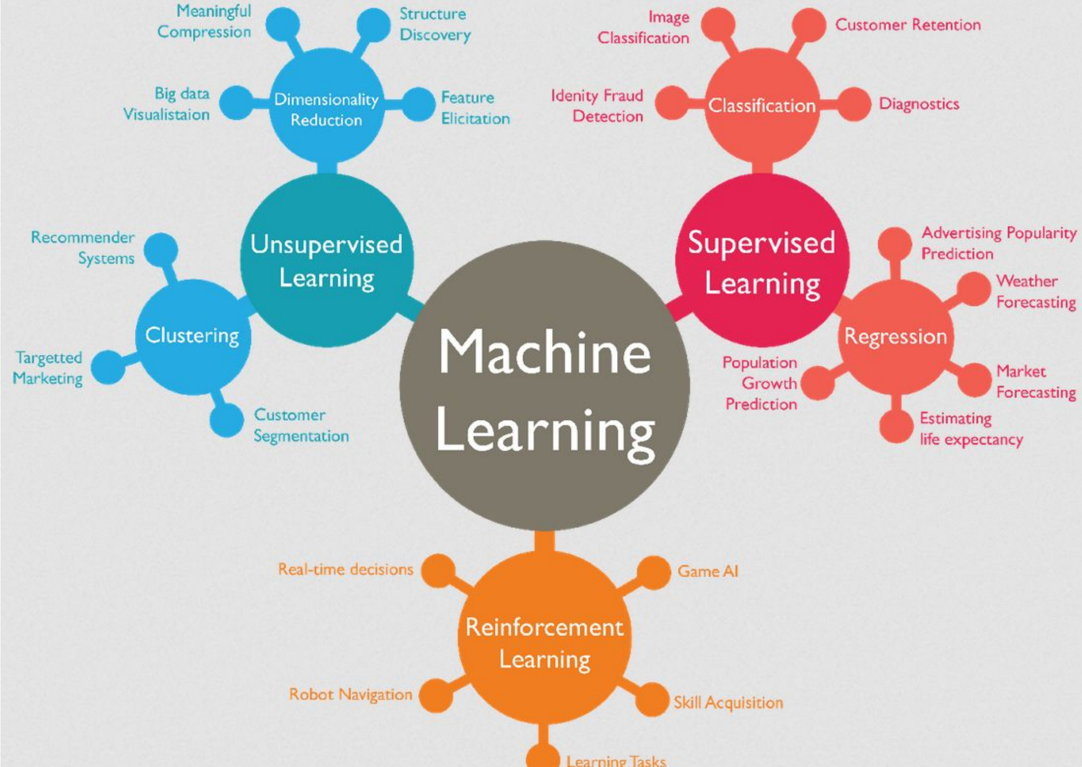
Prof: M. Sc. René Vilar Schlichter

Machine Learning

Contents

- Math and Statistics: (Today)
  - Quantitative vs Cualitative Data
  - Linear Algebra recap
    - Linear Equations
    - Vectors
    - Matrices
  - Calculus recap
    - Derivatives
    - Integrals
    - Demo on Gradient Descent
  - Statistics recap
    - Central Tendencies
    - Spread
    - Probability

Teaser



Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Data

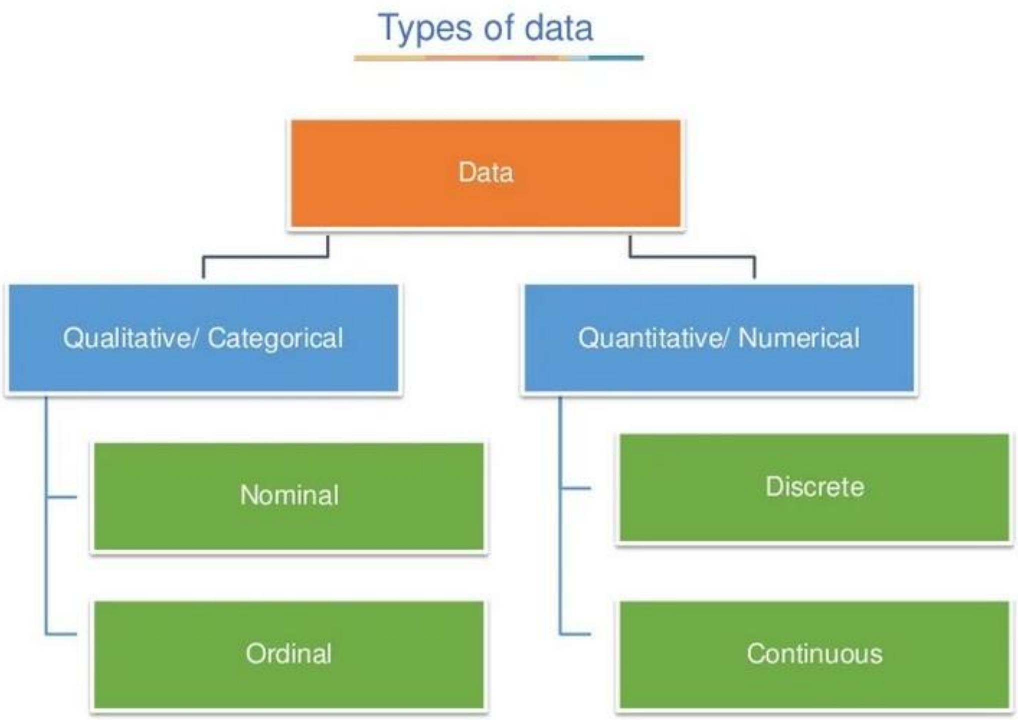
What is Data?

Data denotes the individual pieces of factual information collected from various sources. It is stored, processed and later used for analysis

Data in various forms

Performing analytics to derive insights

Types of Data

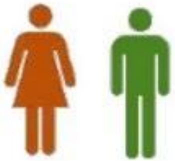


Qualitative Data

Qualitative/ Categorical Data

Nominal

They are used to label variables without providing any measurable value



E.g.: Country, gender, race, hair color, etc.

Ordinal

Categorical data with a set order or scale to it

SALARY RANGE (in dollars)	NO. OF EMPLOYEES
10k-20k	150
20k-30k	100
30k-40k	25
50k+	10
80k+	5

E.g.: Salary range, movie ratings etc.

Quantitative Data

Quantitative/ Numerical Data

Discrete

Data with final set of values which can be categorized

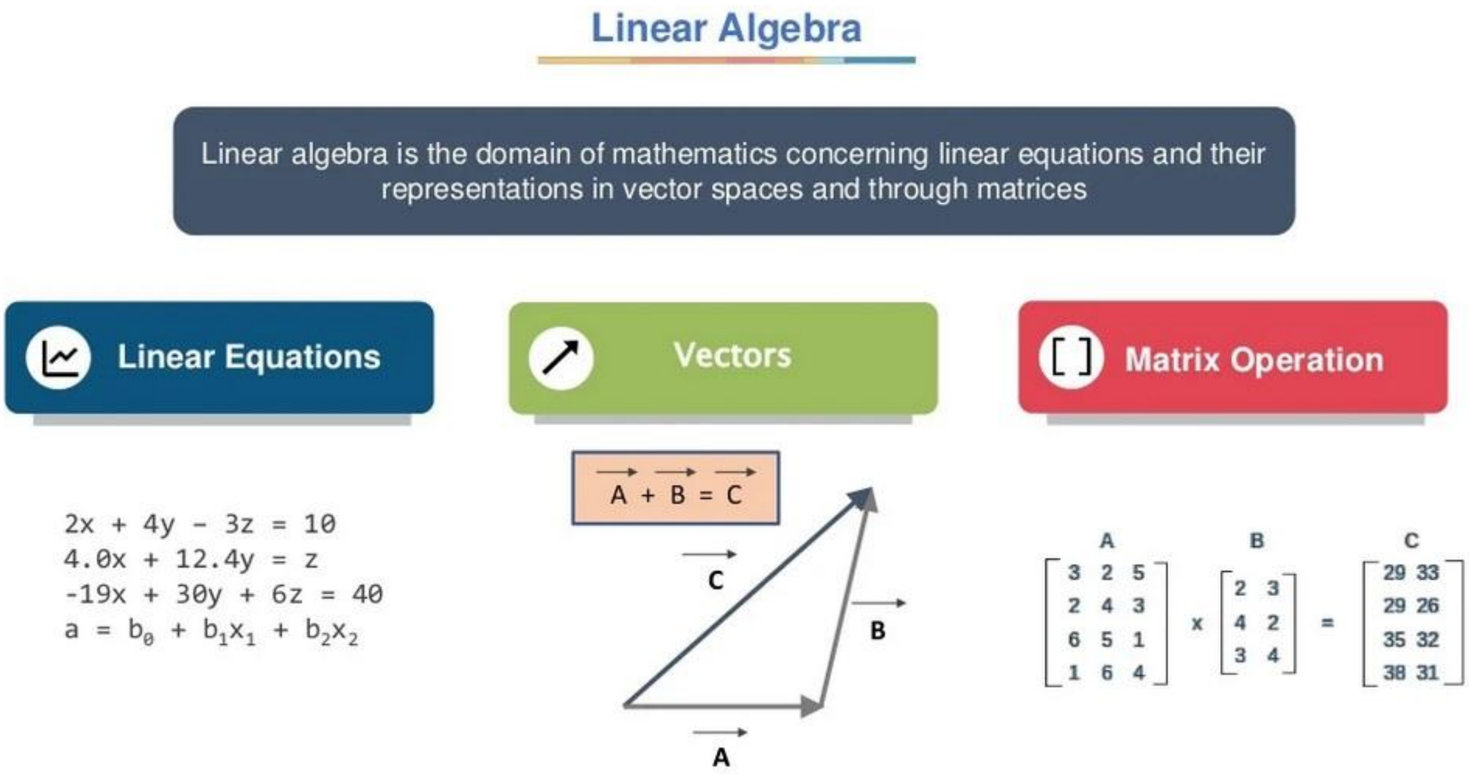
E.g.: Class strength, questions answered correctly, and runs hit in cricket

Continuous

Continuous data can take any numeric value within a range

E.g.: Water pressure, weight of a person etc.

Linear Algebra



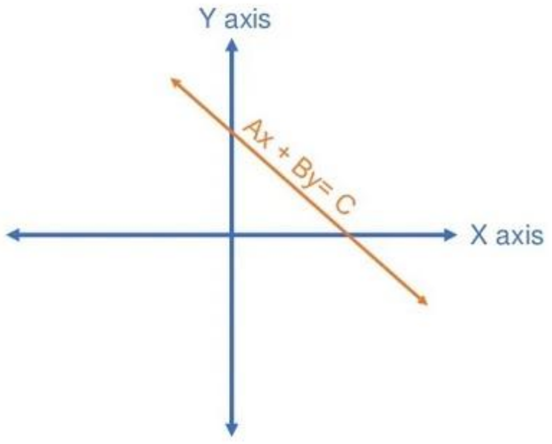
Linear Algebra: Linear Equations

Linear Equations

An equation having a **maximum order of one** is called a Linear Equation

Standard form of Linear Equation

One variable	Ax + B= C
Two variables	Ax + By= C
Three variables	Ax + By + Cz = D
N variables	Ax + By+ Cz +....= N



Linear Algebra: Linear Equations

Linear Equations

In slope-intercept form, a linear equation can be represented as

$$y = mx + c$$

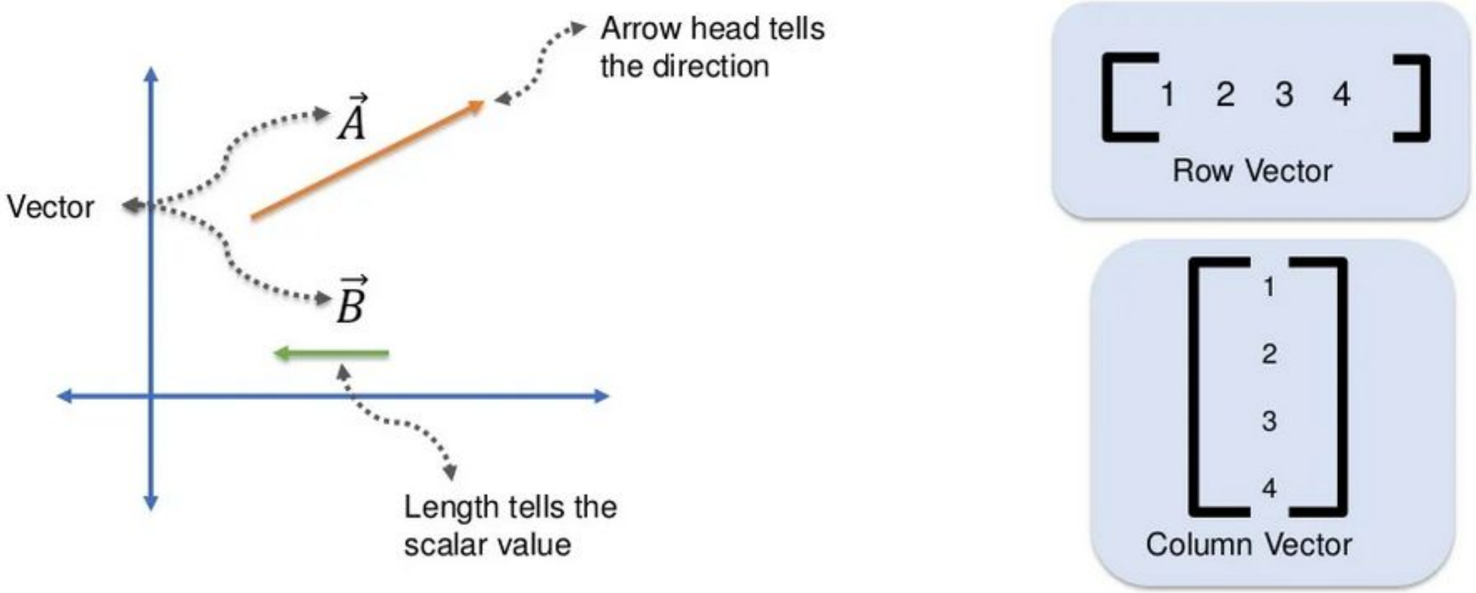
Value of Y when x=0

Slope/Gradient of the line

$$y = 2x + 3$$

Linear Algebra: Vectors

In mathematics, one dimensional matrix is called a ***vector***



Linear Algebra: Matrices

A matrix refers to a rectangular representation of an array of numbers arranged in columns and rows

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_N \\ A_{21} & \ddots & & & \vdots \\ A_{31} & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ A_M & \cdots & \cdots & \cdots & A_{MN} \end{bmatrix}$$

$M_{\text{rows}} \times N_{\text{columns}}$

Here A_{11} is the denotes the element of first row in first column

Similarly, A_{12} is the element of first row and second columns and so on

Linear Algebra: Matrix Operations

Addition

$$\begin{bmatrix} 10 & 5 \\ 11 & 16 \end{bmatrix} + \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}$$

$$\begin{bmatrix} 12 & 8 \\ 14 & 21 \end{bmatrix}$$

Subtraction

$$\begin{bmatrix} 10 & 5 \\ 11 & 16 \end{bmatrix} - \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}$$

$$\begin{bmatrix} 8 & 2 \\ 8 & 11 \end{bmatrix}$$

Linear Algebra: Matrix Operations

Multiplication

$$\begin{bmatrix} 1 & 4 \\ 6 & 3 \end{bmatrix} \times \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} 1x2 + 4x3 & 1x3 + 4x5 \\ 6x2 + 3x3 & 6x3 + 3x5 \end{bmatrix} = \begin{bmatrix} 14 & 23 \\ 21 & 33 \end{bmatrix}$$

Linear Algebra: Matrix Operations

Transpose

Flipping the matrix over its diagonal

$$\begin{bmatrix} 12 & 8 \\ 14 & 21 \end{bmatrix} = A$$

$$\begin{bmatrix} 12 & 14 \\ 8 & 21 \end{bmatrix} = A^T$$

Inverse

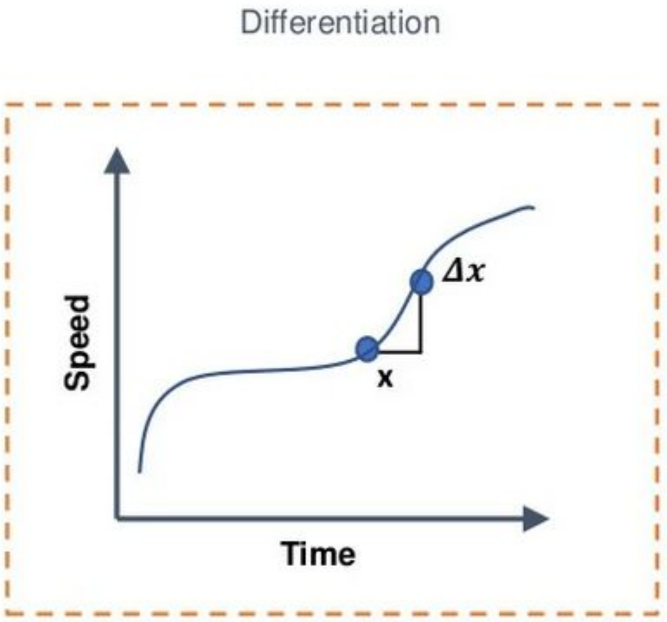
Changing the signs of values across its main diagonal

$$\begin{bmatrix} 12 & 8 \\ 14 & 22 \end{bmatrix} = A$$

$$\begin{bmatrix} -22 & 8 \\ 14 & - \end{bmatrix} = A^{-1}$$

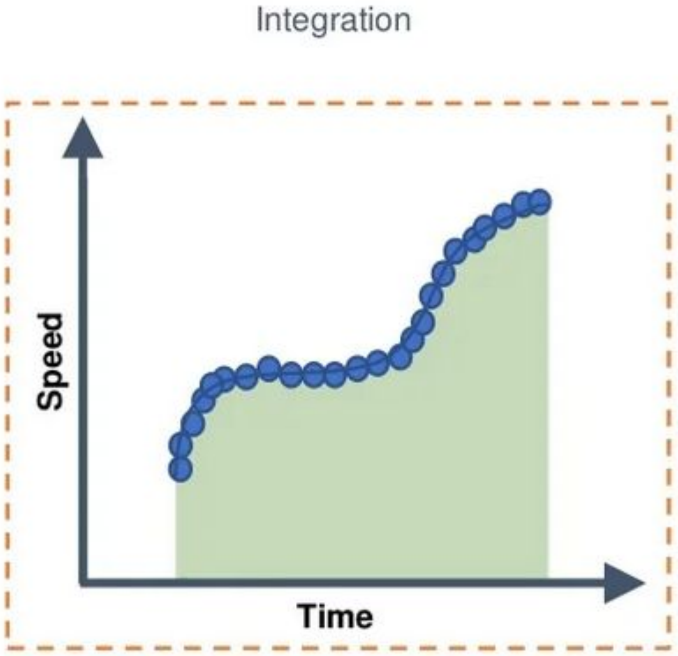
12

Calculus: Differentiation



- Helps to calculate the spontaneous rate of change
- Suppose we plot a graph of the speed of a car with respect to time
- The rate of change of speed with respect of time is nothing but acceleration
- The acceleration is the area between the start point x and end point delta Δ x

Calculus: Integration



- Finding the area under the slope is the main process in the integration
- Similar, small intervals are made of smallest possible length **x + Δ x**
- Helps to find the overall acceleration by summing up all the lengths together

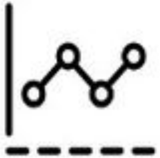
$$\int ax \, dx = a + c$$

Calculus: Applications

- It provides us the tools to build an accurate predictive model

- Multivariate calculus explains the change in our target variable in relation to the rate of change in the input variables

- In gradient descent, calculus is used to find the local and global maxima



Statistics

Statistics: Types of Sampling

Probabilistic approach

Selecting samples from a larger population using a method based on the theory of probability

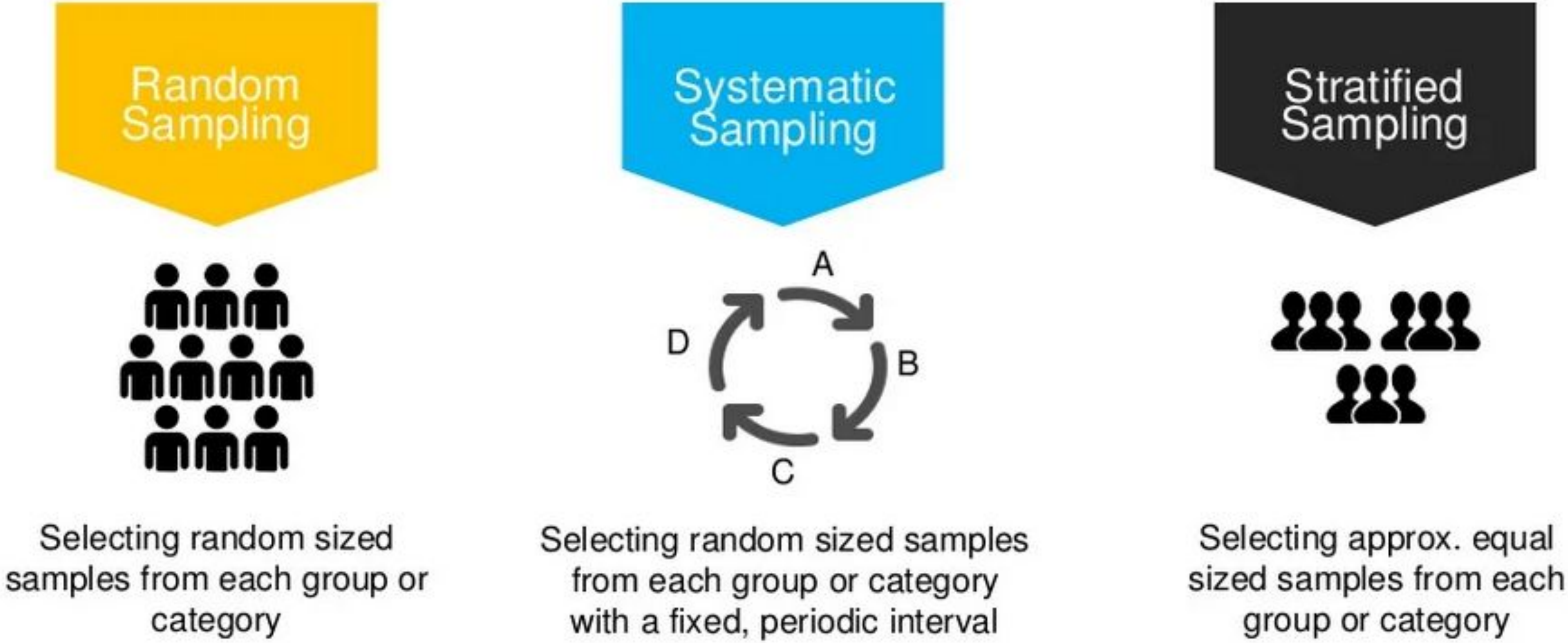
E.g. Random, Systematic, Stratified

Non-probabilistic approach

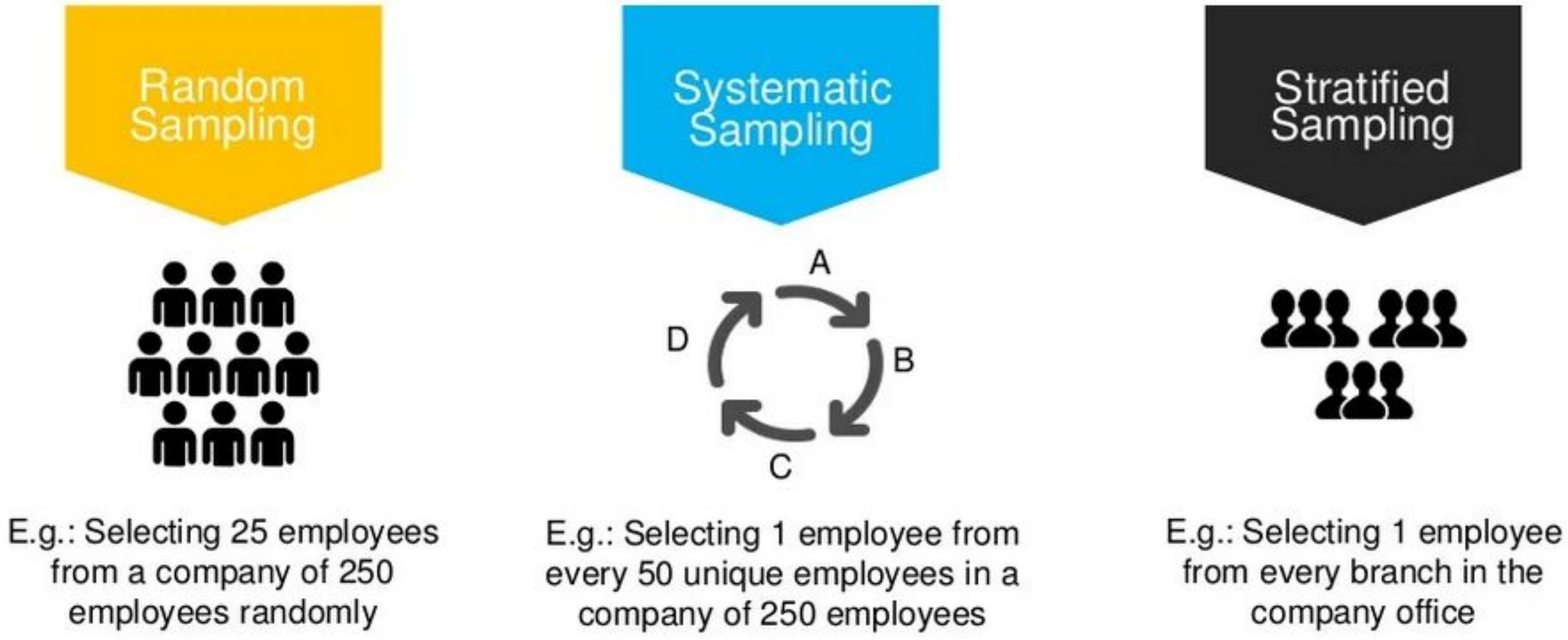
Selecting samples based on the subjective judgment of the researcher rather than random selection

E.g. Convenience, Quota, Snowball

Statistics: Probabilistic Sampling



Statistics: Probabilistic Sampling



Statistics: Types

Descriptive Statistics

It is used to describe the basic features of data and form the basis of quantitative analysis of data

Measure of central tendencies

- Mean
- Median
- Mode

Measure of Spread

- Range
- Interquartile range
- Variance
- Standard deviation

Descriptive Statistics

It is used to describe the basic features of data and form the basis of quantitative analysis of data

Measure of central tendencies

- Mean
- Median
- Mode

Measure of Spread

- Range
- Interquartile range
- Variance
- Standard deviation

Statistics: Example of Central Tendencies

Mean: Average marks of students in a classroom

$$\text{Mean} = \frac{\text{Sum of all the marks of the students}}{\text{Total no. of students}}$$

Median:

0 1 2 3 4 5 6 7 8 9 10

Mode: What mark was scored by most of the students in a test

Descriptive Statistics

It is used to describe the basic features of data and form the basis of quantitative analysis of data

Measure of central tendencies

- Mean
- Median
- Mode

Measure of Spread

- Range
- Interquartile range
- Variance
- Standard deviation

Statistics: Example of Variance

Marks of Students (out of 100)
50
45
85
90
56
71
63

To understand the Variance and Standard Deviation we first need to find out mean

Mean

$$= \frac{45 + 50 + 56 + 63 + 71 + 85 + 90}{7}$$

$$= 65.71 = 66 \text{ (approx.)}$$

Statistics: Example of Variance

- **Variance** : measures how far each number in the set is from the mean and therefore from every other number in the set
- **Standard deviation** : it is the measure of the variation or dispersion of a set of values from the mean

Statistics: Example of Variance

Marks of Students (out of 100)	$(\text{Marks - Mean})^2$
50	50-66 =256
45	45-66 = 441
85	85-66 = 361
90	90-66 = 576
56	56-66 = 100
71	71-66 = 25
63	63-66 = 9

Variance

$$= \frac{(\text{Marks - Mean})^2}{\text{Total observations}}$$

$$= \frac{256 + 441 + 361 + 576 + 100 + 25 + 9}{7}$$

$$= \mathbf{253\ (approx.)}$$

Standard Deviation

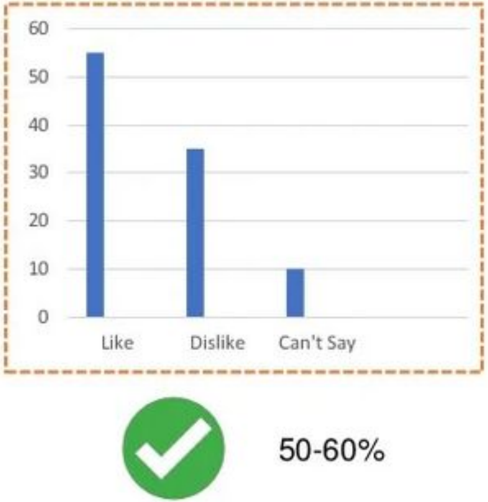
$$= \sqrt{\textbf{Variance}}$$

$$= \mathbf{16\ (approx.)}$$

Inferential Statistics

Inferential statistics allows you to make predictions or inferences from data

Movie Ratings

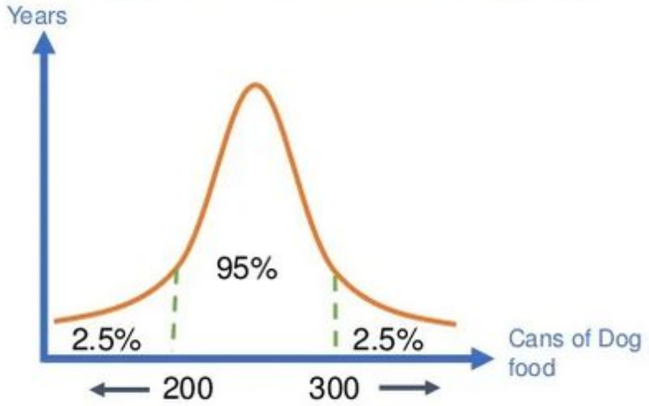


Statistics: Confidence Interval

A Confidence Interval is a range of values we are sure our true values of our observations lies in

Let's say you asked dog owners around you and asked them how many cans of food do they buy per year

Through calculations you got to know that the on an average around 95% of the people bought around 200-300 cans of food.



Hence, we can say that we have a confidence interval of (200, 300) where 95% of our values lie

Statistics: Bell Curve or Normal Distribution

The distribution of the sample means will be approximately normally distributed if you take large random samples from the population with mean μ and standard deviation σ ,with replacement