Statistics

Probability & Statistics

Watches Sale Probem :-

Descriptive	Predictive	Prescriptive
Describes the problem	make changes and see the outcome	Optimal Otimization
like watches not get sale out	like: Improve fashion & see the result	like broadway analysis like increasing sales might gives less profit (In same prospect).

Descriptive :-

"Describing the data without necessarily trying to built any data without any prediction or modular scheme to it."

Mean → Average of the overall

Median → Average of the middle

In case of Ages

Mean → Average of overall

Median → Age of the average person

Mean = Median ⟷ Balanced data

Mean > Median → More data on right side (rightly skewed is pushed.

Mean − Median $=$ +ve (Right Skewness)
-ve (Left Skewness)

Standard Deviation

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + x_3 \ldots)$$

$$\text{Std} = \sqrt{\frac{1}{n-1}\left[(x_1 - \bar{x})^2 + (x_n - \bar{x}) + \ldots\right]}$$

How far the data is from mean?

every data point

too far → more deviation.

less far → less deviation.

⊙ we put squares/square root to cover both the negative/positive variation.

MAD (Mean Absolute Deviation)

$$\frac{1}{n-1}\left[(x_1-\bar{x})^2 + (x_n - \bar{x})\right]$$

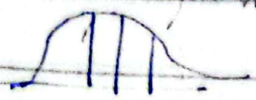
→ More Practical

How far on the average is that observation from the average?

Types of Data

① Numerical
② Categorical
- nominal (names)
- ordinal (ordered scaled data)
  - wishes, agrees, disagrees etc.

→ Distribution Analysis

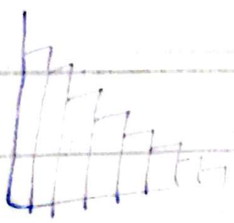


Distribution analysis is done to have some specific target. follow some specific target.

→ PAC learning.

(Probably Approximately learning)

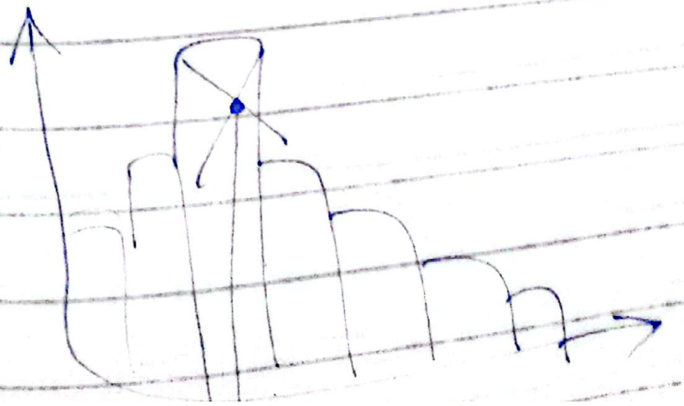
μ — population parameter mean it is unknown.



Heavy Tail Distribution

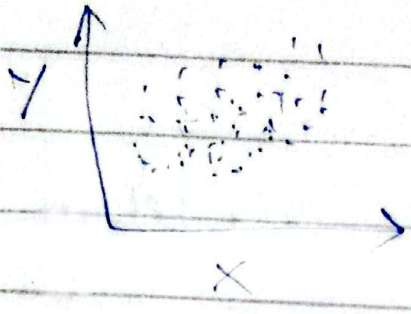
network traffic example.

Mode → Peak of the Distribution
(Most Repeated)
highest frequency in data.



Bivariable

Here exist the correlation b/w two variables.



All these points are pairs $(x_1,y_1),(x_2,y_2)\dots(x_n,y_n)$

$$\bar{x} = \frac{1}{n}\left[x_1+x_2\dots x_n\right] = \frac{1}{n}\left[\sum_{i=1}^{n} x_i\right]$$

$$\bar{y} = \frac{1}{n}\left[\sum_{i=1}^{n} y_i\right]$$

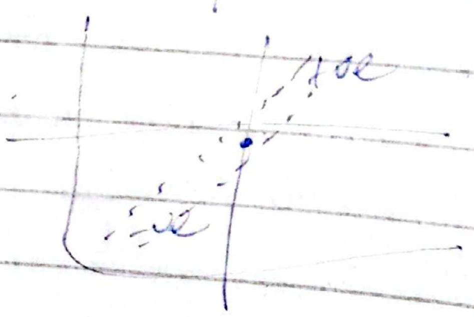
$$\left[\sum_{i=1}^{n} (x_i-\bar{x})(y_i-\bar{y})\right]$$

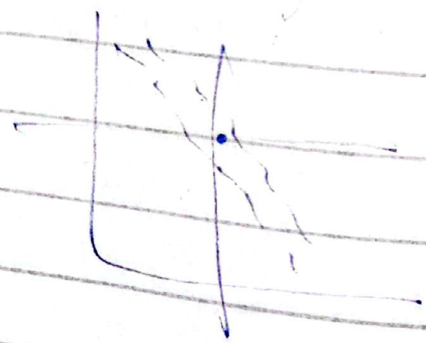
+ve

x and y are above average

-ve

x or y is below average





Covariance

Dimension Reduction

Portfolio management

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

Covariance(x,y)

$$= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \left[S\cdot D(x)\right]^2$$

= Variance(x)

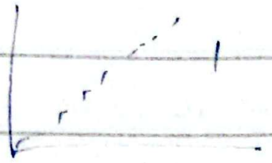
Covariance scenarios

- +ve (Moving in same direction)
- -ve (Moving in different directions)
- 0 (there can be different scenarios)

$\frac{1}{n-1}$ — gives the nature of relationship

$$\frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\bar{x})(y_i-\bar{y})}{sd(x)\,sd(y)}$$ Correlation (X,Y)

$-1 \le corr \le 1$



correlation

"It is the measure of linear relationship between X and Y."

Correlation

- +ve (Both X and Y are above average)
- -ve (Both are below average)
- 0 (No relationship)

$bmi = \frac{h}{(w)^2}$ = Two variable are converged into one (Dimensionality Reduction).

Heatmap → gives a bigger picture of the data.

Intercept is not the part of correlation, as we are talking just about the Descriptive analysis.

Descriptive Statistics

* Used for Inferencials
* Critical Analytics
* Prespective Analysis
* Visualization
* Descriptive Visualization

→ Univariant Data
  └→ One variable.

Following points tells what is the data about?

location → mean, median, Q_3, Q_1
variation → Standard Deviation, Range, IQR
Skewness

→ Bivariant Data
  └→ 2 variables

Covariance, Correlation.
  └ Universion version (Variance)

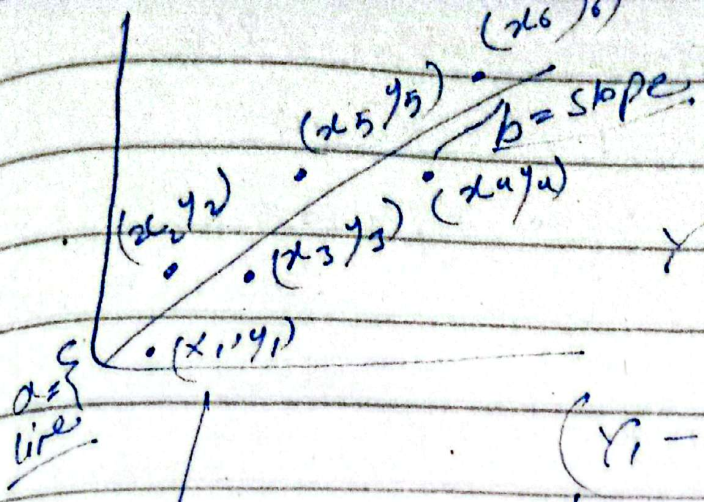
→ Multivariant

Linear Regression
  └ can be used for prescription and prediction as well.

Plots

* Histogram
* Boxplot
* Pairs
* Scatter Plots

} Human eye can understand / see them easily



$Y = a + bx$

$$\left(Y_1 - [a+bx_1]\right)^2 + \left(Y_2 - a - bX_2\right)^2$$

+ .... 5 lines for every point

larger — Farther is the line from data

Smaller — Closer is the line from data.

* These data points are the actual
* $Y = a+bx$ is the prediction.
* Actually data points are not the actual, actual data will come into the Inference / Model.

Statistical learning

Probability

Probability help us to measure the uncertainity and perform associated analyses that are essential in making effective business decision.

Uncertainity

→ Not know the corresponding outcome

* Probability refers to chance or likelihood of a particular event taking place
* An event is an outcome of an experiment.

① An experiment is a process that is perform to understand and observe possible outcomes.
② Set of all outcomes of an experiment is called sample space.

→ Experical Probability :-

Example
3 parts to assemble
① Find the probability in which atleast 2 parts are defective
② All possible sample set

② - GGG GGD GDG DGG
GDD DGD DDG DDD

① - P(GDD or DGD or DDG)
P(GDD) + P(DGD) + P(DDG)

Mutually Disjoint - P(G)×P(D)×P(D) + P(D)×P(G)×P(D) + P(D)×P(D)×P(G)

→ Only One exis at one time.
Multiplication represent that everything is Independent.

$\binom{3}{2} \times (1)^2 \times (0.9)^1$ ← 0.9×1×1 + 1×1×1 + 1×1×0.9

2 Defective parts. Chance of Bad ↑ Chance of Good ↑

$3 \times 1 \times 1 \times 0.9$

$\boxed{2.437} \rightarrow 2.43\%$

→ Probability

Probability of an event is defined as the ratio of two numbers m an n.

$P(A) = m/n$

m = Number of ways that are favourable to the occurence of A

n = Total Number of outcomes.

P(A) is always >= 0 and always <= 1

1 — Possible   0 → Uncertain

100% Chance or Certainity
50% Equally likely
0% Impossibility

} Extreme values of possibility

→ Independent Events

Must not have effect on each other.

→ Mutually Exclusive Events

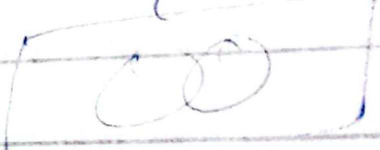
Rules for Computing Probability

Rules for Mutually Exclusive Events

① Addition Rule

$P(A \cup B) = P(A) + P(B)$

Union 

Inter-section 

$P(A \text{ or } B) = P(A) + P(B)$

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

⓪ if case of disjoint

Multiplication Rule

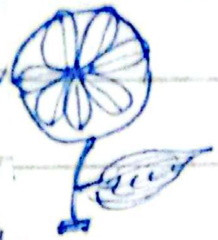
$$P(A \cap B) = P(A) \times P(B/A)$$

Events are not independent.

$$P(A \cap B) = P(A) \times P(B \text{ given } A)$$

OR

$$P(B \text{ given } A) = \frac{P(A \text{ and } B)}{P(A)}$$



If Events are independent

$$P(B \text{ given } A) = P(B)$$

Marginal Probability

The term marginal is used to indicate that the probabilities are calculated using a contingency table (joint probability table)

Contingency table consists of rows & colours of two attribute at different levels with frequencies or numbers in each of the cells. It is a matrix of frequencies assigned to rows & cols.

Conditional Probability

Based on some condition.

Bayes Theorem

Is used to revise previously calculated probabilities based on new information.

Extension of Conditional probability.

Baysian Spam Filtering

$\rightarrow$ Email Filtering

$$P(A \text{ and } B) \begin{cases} P(B) \times P(A|B) \\ P(A) \times P(B|A) \end{cases}$$

P(spam and word)

① P(spam/words) × P(word)

② P(word/spam) × P(spam)

$$P\left(\frac{spam}{word}\right) = \frac{②}{①}$$

Bionomial Distribution

Quality control Quality Assurance

Probability Distribution of a discrete random variable.

Reducing number of defectives using the porportion defective control chart.

Bionomial Possibility function

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Number of ways in which x successes can take place

$n \rightarrow$ trial

$p \rightarrow$ success probability of each trial

P(x success)

Poisson Distribution

- Another discrete distribution method.
- Possibility of getting exactly one success in continuous interval such as length, area, time and the like is constant.
- The probability of success in any of one interval is independent of the probability of success occurring in any other interval.
- Probability of getting more than one success in an interval is 0

$$P(X) = \frac{e^{-\lambda} \lambda^{x}}{x!}$$

$\lambda \rightarrow$ Average Number of successes
$e \rightarrow 2.71828$
$x \rightarrow$ Success per unit