Day-72, Feb 10, 2025 (Magh 28, 2081)

Descriptive Statistics versus Inferential Statistics

Both are statistical methods to gain different insights.

Descriptive Statistics summarize the main features of a dataset.

Visuals like graphs, tables and Summary Stats. (Summarize the data using mean (Central tendency) and measures of Dispersion (σ))

Inferential Statistics

→ Make Inferences about a dataset based on a sample of the data. Sampling, Confidence Intervals, Hypothesis Testing are the Inferential Statistics.

Measure of Central Tendency

→ Helps to understand the values that represent the center of a dataset

→ Helps to understand the values that represent the spread of a dataset.

Measures of Central Tendency

- Mean
- Median
- Mode

(10, 8, 5, 7, 70)

Mean - Average value of a dataset, (10, 8, 5, 70, 7 $\bar{x} = 20$)

Median - Middle value of a dataset (5, ~~7~~ 8, 10, 70)

Use mean, if no outliers but use median if median.

Home Prices: Mean or Median for home prices depends upon the sample data.

Mode: frequency or Repetition of data. Example Company Review Rate.

Data professionals use standard Deviation to measure variation in
→ ad revenues
→ stock prices.
→ Employee Salaries.

Measures of Dispersion

Mean for each set = 30

Set 1: 25, 30, 35

Set 2: 10, 25, 55

Set 3: 5, 10, 75

Dispersion level.

Set 3 > Set 2 > Set 1

- Range
- Standard Deviation

→ Range is the difference the largest and the smallest value in a dataset.

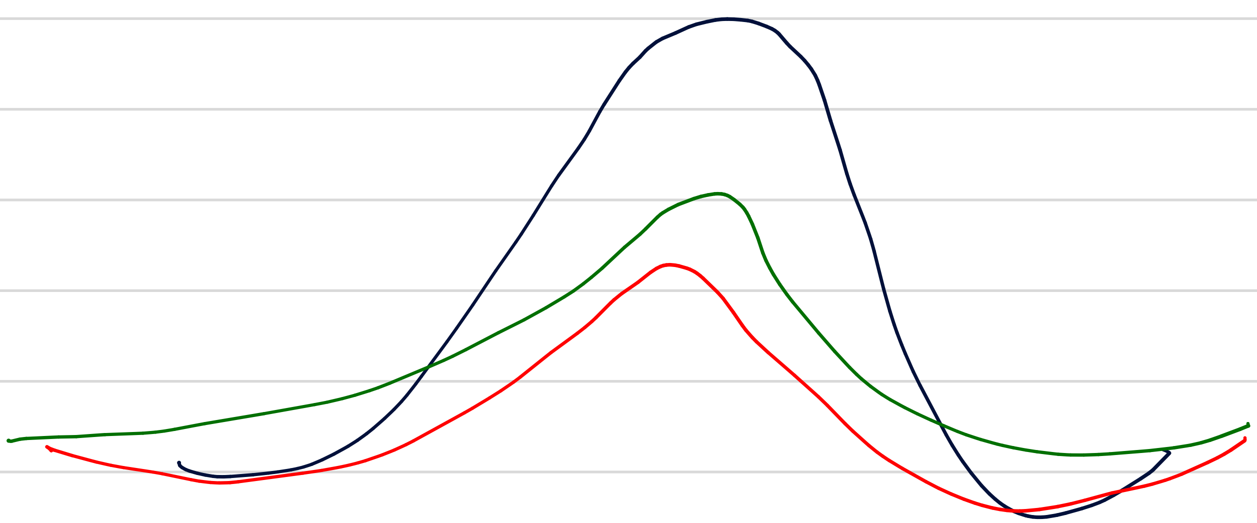
→ 77, 74, 72, 71, 67, 69, 72

Range: 77 − 67 ⇒ 10

Standard deviation (σ): Tells us how the data are spread out from the mean.

Variance: The average of the squared difference of each data point from the mean. Actually the square of σ.

So, in the Gaussian Distribution Mean and Variance | Std. Deviation are:



std < std > std

Sample standard Deviation

$$S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

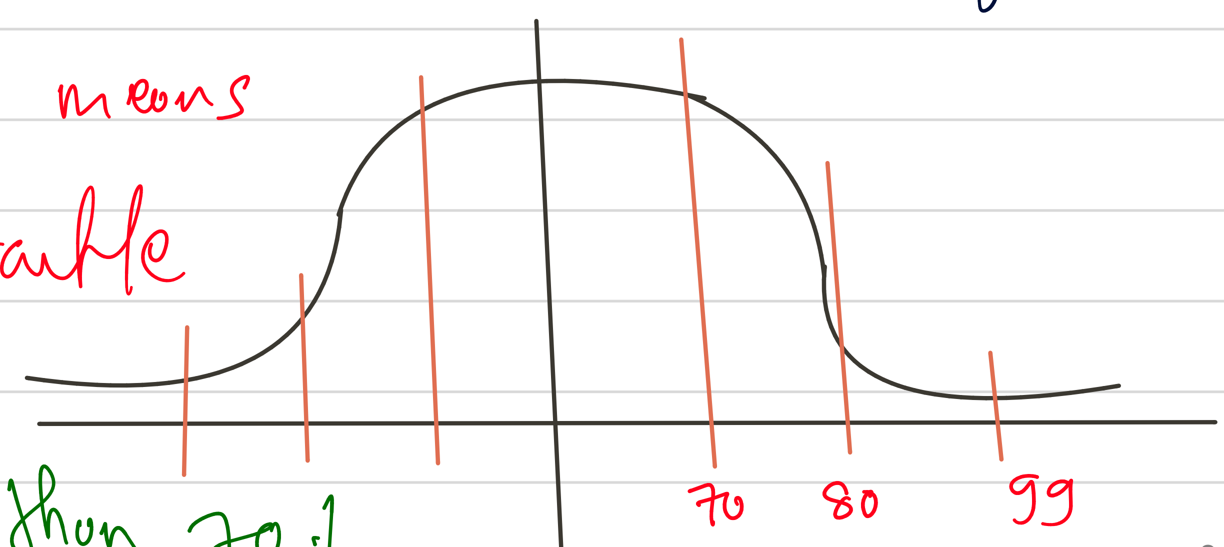
Measures of Position

→ Determine the position of a value in relation to other values in a dataset.

- Percentiles
- Quartiles
- Interquartile Range
- Five Number Summary.

Percentile: The value below which a percentage of data falls.

Suppose student receive 99 percentile means marks | test score falls in the 99th percentile higher than 99%.



Similarly if it is 70 than it is higher than 70%..

Quartile : Divides the values in a dataset into four equal parts.

Q1 Q2 Q3

25%	25%	25%	25%

Median.

$Q_1 \rightarrow 25\%$

$Q_2 \rightarrow 50\%$

$Q_3 \rightarrow 75\%$

From table 1.0, we get -

$Q_2 = 20$

$Q_3 = 25$

$Q_1 \Rightarrow 13$

Breaking Data into Quartile Helps to Analyze the Data.

Table 1.0.

Player	#7	#3	#8	#1	#2	#6	#4	#5
Goals scored	11	12	14	18	22	$\boxed{23}$	$\boxed{27}$	33

Inter Quartile Range (IQR)

→ Distance between the first Quartile (Q_1) and the third Quartile (Q_3)

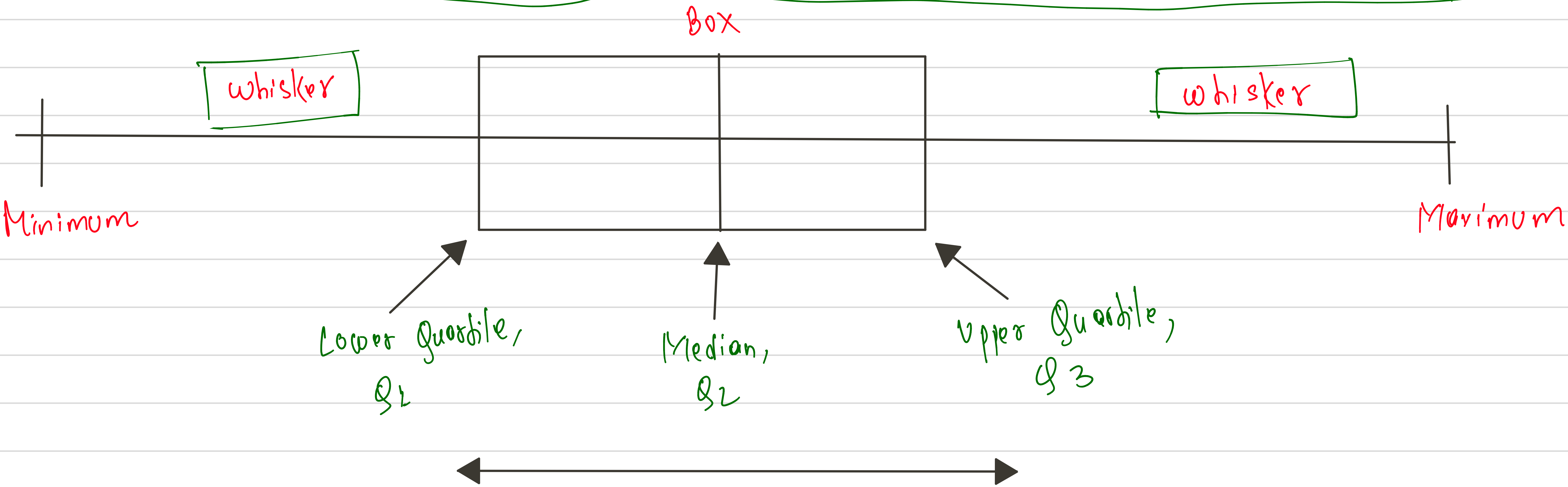
$$\text{IQR} = -Q_1 + Q_3 \Rightarrow Q_3 - Q_1$$ — Measure the middle Q

$$\text{IQR} \Rightarrow 25 - 13 \Rightarrow 12$$ from Table 1.0.

Five Number Summary

- The minimum = 11
- The first Quartile (Q_1) = 13
- The median, or Second Quartile (Q_2) = 20
- The third Quartile (Q_3) = 25
- The maximum = 38

From Extreme values to center, let's visualize it with a box plot:



$Q_3 - Q_1$ = Interquartile

$25 - 13 \Rightarrow 12$

Importance of Measures of Position to better Understand:

- Public Health Data such as Life Expectancy
- Macroeconomic data such as household income
- Business data such as Product sales.

Probability helps to measure and quantify uncertainty and make informed decisions about uncertain outcomes.

[Probability use Cases.]

- A Company will sell a certain amount of product in a given time period.
- Positive Return in financial investment.
- A political candidate'll win an election.
- A medical test'll be accurate.

Objective | Subjective Probability.

→ Based on Statistics, experiments, and mathematical measurements.

→ Based on personal feelings, experience, or judgement.

Objective Probability:

→ Based on formal reasoning about events with equally likely outcomes.

$$\text{Classical Probability} \Rightarrow \frac{\text{No. of desired outcomes}}{\text{Total number of possible outcomes}}$$

$\Rightarrow \frac{1}{2}$ in flipping Coin

Empirical Probability

→ Based on experimental or historical data. that means it
→ represents the likelihood of an event occuring based on the previous results of an experiment or past events.

Empirical Probability = $\dfrac{\text{number of times a Specific event occurs.}}{\text{total number of events.}}$

The Principles of Probability

Events in Probability –

1. Event Equals 0 means 0 % Chance that the event will occur
2. If $P(E) == 1$
   then 100 % Chance event'll occur.
3. else if $P(E) \approx 0$
   Small chance that it'll not occur.
4. else if $P(E) \approx 1$
   then strong chance it'll occur

Events in Probability : Measures the likelihood of event in a randomized condition.

- $0.05 = 5\%$
- $0.95 = 95\%$

Random Experiment.

- Can have more than one possible outcome
- Can Represent each possible outcome (in advance)
- The outcome of the experiment depends on chance.

Classical Probability $= \dfrac{\text{Number of Desired Outcomes}}{\text{total number of possible outcomes}}$

Coin toss

$\frac{1}{2} \Rightarrow 0.5\%$

Dice Roll.

$\frac{1}{6} \Rightarrow 0.1666 \Rightarrow 16.7\%$

The Basic Rules of Probability and Events:

Rules:

1. Complement Rule
2. Addition Rule
3. Multiplication Rule

Types of Events:
1. Mutually Exclusive Events
2. Independent Event

Probability Notation:

$P \to$ probability of event A $= P(A)$

$P(B) \to$ " " " " B

$P(A') \Rightarrow$ probability of not A

the event of occurring or not occurring.

Complement of Probability

$$P(A') = 1 - P(A)$$

Total probability = Probability of Not Occur + Probability of Event Occur.

$P(\text{not Rain}) = 1 - P(\text{rain}) \Rightarrow 1 - 0.3 \Rightarrow 0.7 \Rightarrow 70\%$.

Mutually Exclusive Events

→ Two events are mutually expensive if they cannot occur at the same time.

→ Eg: you cannot go left & Right at a time.

Eg: Six Sided Dice:

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

$$P(\text{roll } 2 \text{ or roll } 4) = P(\text{roll } 2) + P(\text{roll } 4)$$

$$\Rightarrow 1/6 + 1/6 \Rightarrow 1/3$$

Independent Events:

→ Two events are independent if the occurrence of one event doesn't change the probability of the other event.

→ One does not effect the outcome of another outcome.

Multiplication Rule (for Independent Events)

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

P(1st toss tails and 2nd toss heads) $=$ P(1st toss tails) $*$ P(2nd toss heads)

$$P(\text{tails and heads}) = \frac{1}{2} * 1/2$$

$$\Rightarrow 0.25$$

Addition Rule (applies) helps in Mutually Exclusive Events.
Multiplication Rule helps/ applies to Independent Events.
So, these two rules helps to outcome the Probability Event.

Conditional Probability

The probability of an event occuring given that another event has already occured.

$P(A/B)$ mean probability of event A given that probability of event B has already occured.

Baye's Theorem | Baye's Rule

$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)}$$

Named After Thomas Bayes.

Prior Probability

The probability of an event before new data is collected.

Posterior Probability

The updated probability of an event based on new data.

Bayesian Statistics (Bayesian Inference)

A powerful method for analyzing and interpreting data in modern data analytics.

Applications are found in

- Artificial Intelligence
- Medical Testing
- Financial Institutions
- Online Retailers
- Marketers

$$\therefore P(A/B) = \frac{P(B/A) * P(A)}{P(B)}$$

- $P(A)$ = Prior Probability
- $P(A/B)$ = Posterior Probability

$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)}$$

if we know one of the conditional probability $P(A/B)$ or $P(B/A)$ we can compute Bayes theorem.

Example: Weather Information

- Overall chance of Rain = 10 %.
- All days start off cloudy = 40 %.
- Rainy days start of cloudy = 50 %.

Probabilities

- Prior probability = the probability of Rainy day
- Posterior probability = the probability that it'll rain given that it's cloudy

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

- Event A = Rain
- Event B = cloudy

$$P(\text{Rain} | \text{Cloudy}) = \frac{P(\text{Cloudy} | \text{Rain}) * P(\text{Rain})}{P(\text{cloudy})}$$

$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)}$$

$$P(\text{Rain} \mid \text{Cloudy}) = \frac{P(\text{Rain}) * P(\text{Cloudy} \mid \text{Rain})}{P(\text{Cloudy})}$$

$$P(\text{Rain} \mid \text{cloudy}) \Rightarrow \frac{(0.1 * 0.5)}{0.4}$$

$$\Rightarrow 0.125 \qquad \Rightarrow 12.5\%$$

So, the Bayes' Theorem helps us to calculate Posterior Probability.

Sources: The Power of Statistics Coursera Course offered by Google.