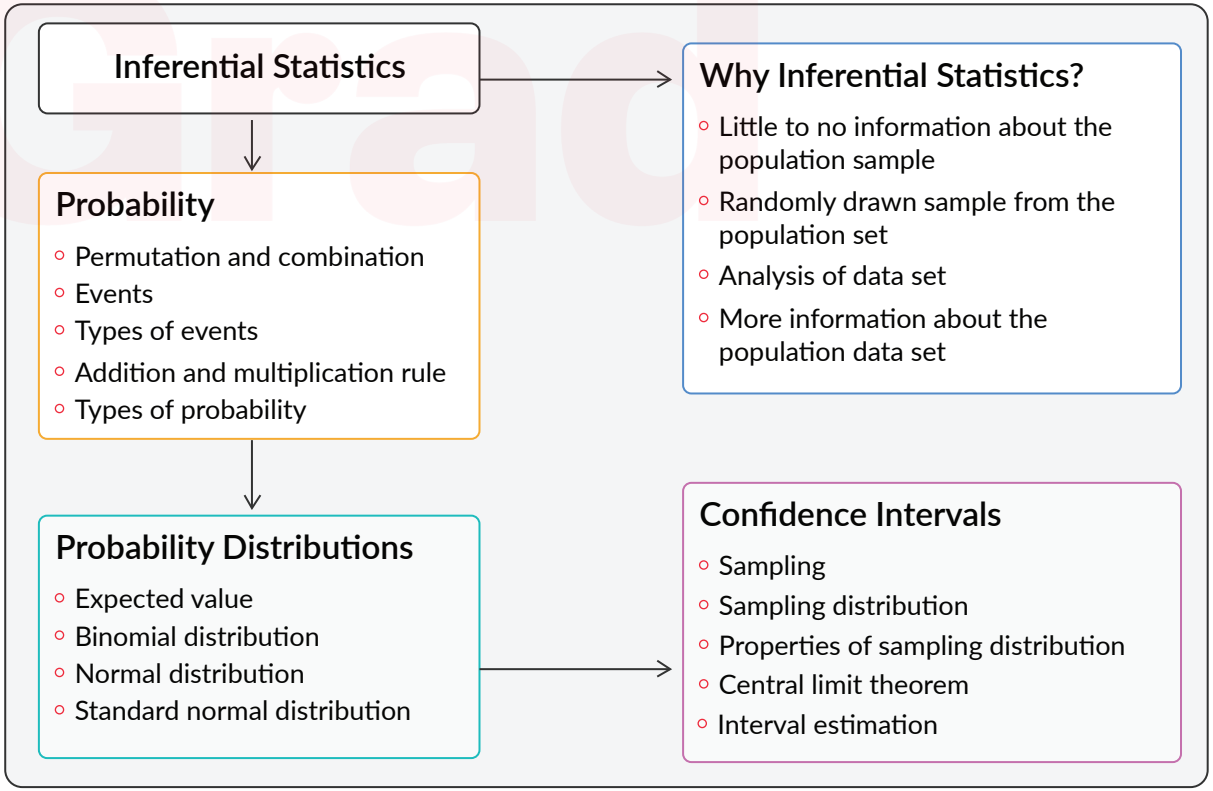
Inferential Statistics

Inferential statistics is a field that uses analytical tools to infer conclusions about a population by examining random samples.

- Use permutation, combination, and probability basics to find an estimation's likelihood.
- Understand probability distribution such as binomial distribution, normal distribution, standard normal distribution, etc., to estimate the variability of occurrence of an event.
- Understand sampling and sampling distribution to simplify the process of statistical inference when a large number of samples is drawn from the population.
- Apply the central limit theorem to safely assume that the sampling distribution of the mean will be normal in most cases.
- Apply the central limit theorem for interval estimation to calculate an interval of possible (or probable) values of an unknown population parameter.

Common Interview Questions

1. What does Inferential Statistics mean?
2. What are the different types of Distributions?
3. How can the sample data be drawn from the population?
4. What is the difference between inferential statistics and descriptive statistics?
5. Explain Confidence level, margin of error, and confidence interval
6. What does Confidence Level signify?
7. What is the use of the central limit theorem?
8. Can confidence interval be negative?
9. What are the techniques applied to gather sample data?



Inferential Statistics

Permutation:

- Arranging *r* objects out of *n* distinct objects in *nPr* ways
- Ordering has significance $_nP_r = \frac{n!}{(n-r)!}$

Combination:

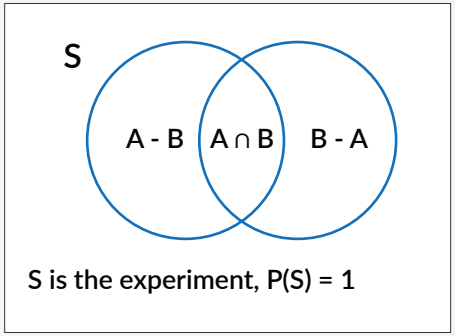
- Selecting *r* objects out of *n* distinct objects in *nCr* ways
- Ordering is not important $_nC_r = \frac{n!}{r!(n-r)!}$

Example: Four-letter words formed using the letters of the word UPGRAD

Step 1: Select 4 letters from UPGRAD
6C4 = 15 ways (Combination)

Step 2: Arrange the selected 4 letters
4P4 = 24 ways (Permutation)

Ans: Number of four-lettered words is 15x24 = 360



Probability:

Probability refers to the chances of occurrence of a given event

Probability of Event A=

$$P(A) = \frac{\text{No. of ways an event can occur}}{\text{No. of all possible outcomes}}$$

Important Terminology:

- **Experiment:** Results in well-defined outcomes. Ex.Tossing a coin
- **Event:** Any collection of outcomes of an experiment
- **Random experiment:** Do not know the exact outcome but know the set of all possible outcomes

Important Terminology:

1. $0 \leq P(E) \leq 1$
2. Sum of all possible outcomes is 1, i.e., $(\Sigma P(Ei) = 1)$
3. Probability of an impossible event is 0.
4. Probability of an event can never be negative.

Rules of Probability:

- **Addition:** P(A U B) = P(A) + P(B) - P(A ∩ B)
- **Complement:** P(AC) = 1 - P(A)
- **Conditional:** P(B|A) = P(A ∩ B) / P(A)
- **Multiplication:** P(A) = P(B|A) x P(A ∩ B)
- If A and B are independent, P(A ∩ B) = P(A)xP(B)

Types of Events

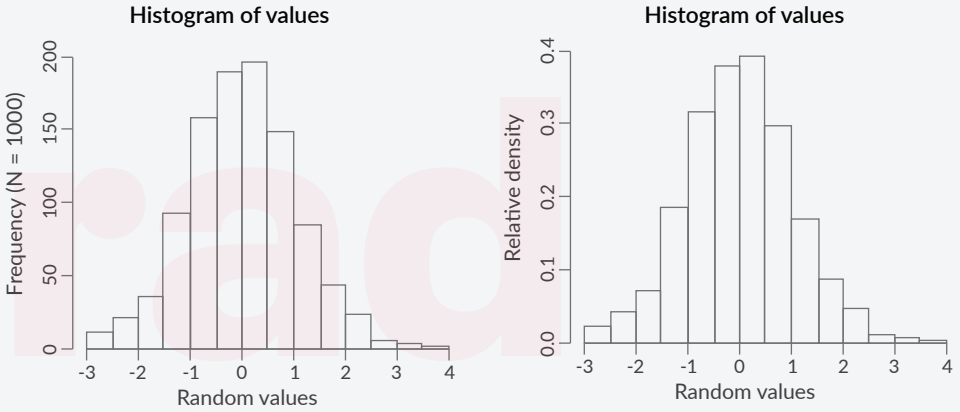
- **Independent event:** Probability of occurrence of two or more events is not affected by each other
- **Disjoint/Mutually exclusive event:** Events cannot occur at the same time

Bayes' Theorem: Helps to calculate the probability of one event when other one already occured

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \; ; \; P(B) \neq 0$$

Probability Distribution:

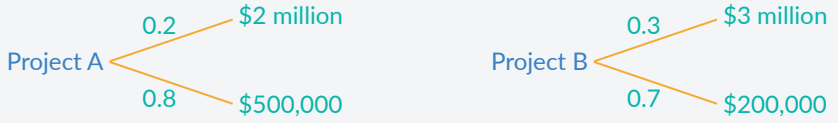
A mathematical function that gives the probabilities of occurrence of different possible outcomes for an experiment.



Expected value:

The value of X that we would 'expect' to get after performing an experiment an infinite number of times
Also called as expectation, mean or weighted average of probability

$$EV(X) = x_1 X P(x_1) + x_2 X P(x_2) + x_3 X P(x_3) + \ldots + x_n X P(x)$$
$$Where\ X = \{x_1, x_2, x_3, \ldots, x_n\}$$



EV (Project A) = [0.2 × $2,000,000] + [0.8 × $500,000] = $8,00,000
EV (Project B) = [0.3 × $3,000,000] + [0.7 × $200,000] = $1,040,000

Inferential Statistics

Probability Distribution

Discrete Probability Distribution	Continuous Probability Distribution
◦ Binomial distribution	◦ Normal distribution
◦ Bernoulli's distribution	◦ Standard normal distribution
◦ Poisson distribution	◦ Students T distribution
	◦ Chi-square distribution

Discrete Probability Distribution

Binomial Distribution

Determines the probability of observing a specified number of successful outcomes in a specified number of trials.

$$P(X = r) = {}^{n}C_{r}(p)^{r}(1\text{-}p)^{n-r}$$

Where **n** - total number of trials,
p - is the probability of success,
r - is the number of successes

Application: Tossing a coin 20 times to see how many tails occur
Not applicable: Tossing a coin until a head appears

Cumulative probability:

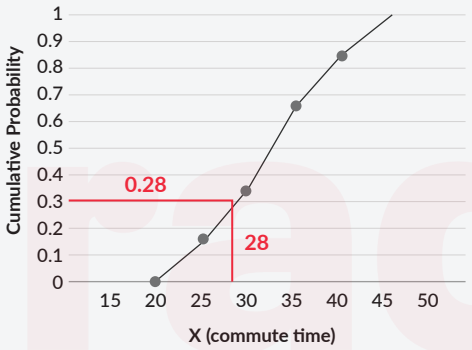
Cumulative probability of X, denoted by F(x), is defined as the probability of the variable being less than or equal to x.

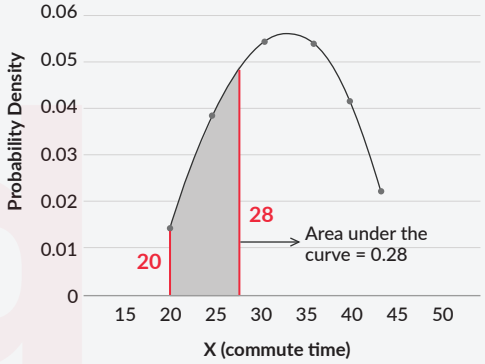
$$F(x) = P(X < x)$$

Continuous Probability Distribution

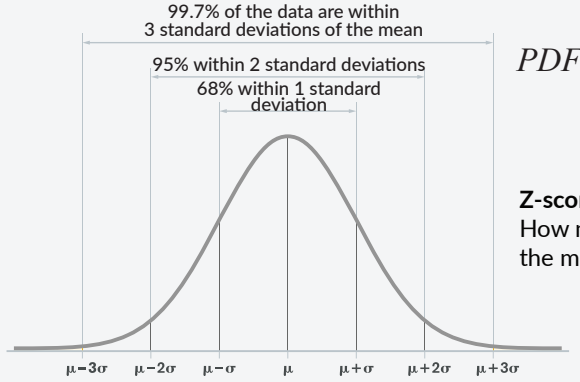
Frequently used concepts: Density Function

- Cumulative Distribution Function: A distribution that plots the cumulative probability of X against X
- Probability Density Function: A function in which the area under the curve gives the cumulative probability





Continuous Probability Distribution



$$PDF = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

μ = Mean of the distribution
σ = Standard deviation

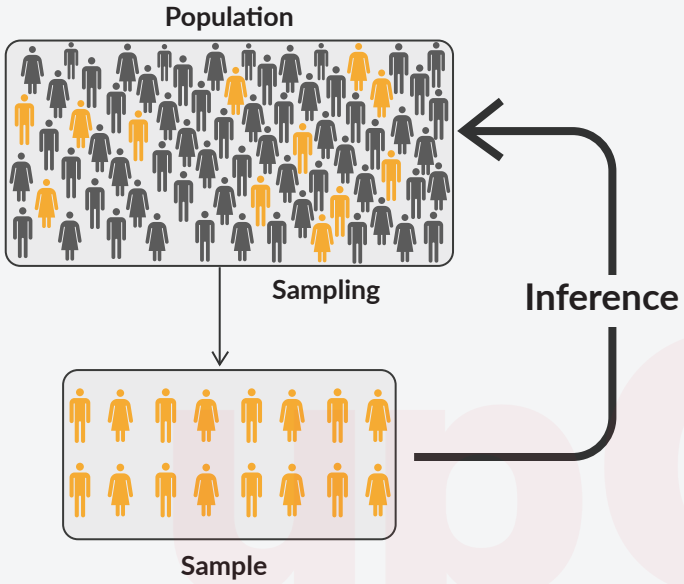
Z-score (Standardise normal variable):
How many standard deviations away from the mean is your random variable is given by

$$Z = \frac{x - \mu}{\sigma}$$

Inferential Statistics

Central Limit Theorem

If sufficiently large random samples are drawn from the population with replacement , then the distribution of the sample means will be approximately normally distributed.



Population/Sample	Term	Notation	Formula
Population $(X_1, X_2, X_3, ..., X_N)$	Population Size	N	Number of items/elements in the population
	Population Mean	μ	$\frac{\sum_{i=1}^{i=N} X_i}{N}$
	Population Variance	σ^2	$\frac{\sum_{i=1}^{i=N} (X_i - \mu)^2}{N}$
Sample $(X_1, X_2, X_3, ..., X_n)$ (Sample of Population)	Sample Size	n	Number of items/elements in the sample
	Sample Mean	$\bar{X}$	$\frac{\sum_{i=1}^{i=n} X_i}{n}$
	Sample Variance	S^2	$\frac{\sum_{i=1}^{i=n} (X_i - \bar{X})^2}{n-1}$

Sampling Distribution:

- Sampling Distribution's Mean $(\mu_{\bar{X}})$ = Population Mean (μ)
- Sampling Distribution's Standard Deviation (Standard error) = $\sigma/\sqrt{n}$
- For $n > 30$, the Sampling Distribution becomes a normal distribution.

Confidence Interval

Population Mean (μ) = Sample Mean (x) + Margin of error

A sample with sample size n, mean x and standard deviation S. Now, the y% confidence interval (i.e., the confidence interval corresponding to a y% confidence level) for μ would be given by the range:

$$CI = x \pm z \frac{S}{\sqrt{n}}$$

where, Z* is the Z-score associated with a y% confidence level

The probability associated with the claim is called the confidence level

1. The maximum error made in a sample mean is called the margin of error
2. The final interval of values is called the confidence interval. [Here, it is the range]