

* Part-A Theoretical foundation

Q1 What is statistical distribution?

Ans A mathematical function that shows how data or random variables are spread. It tells us the probability of different outcomes.

eg: Normal, Binomial, Poisson)

Q2 What is a Q-Q plot and why it is used?

Ans A Scatter plot comparing the quantiles of sample data against a theoretical distribution (often normal). If points lie close to a straight line \rightarrow data follows that distribution.

Q3 Difference between discrete and continuous distribution?

Ans Discrete: Takes specific values (number of students, dice roll). Probability is described by PMF.

Continuous: Takes any value in a range (weight, time) probability is described by PDF.

Q4 What is Bernoulli distribution?

Ans Models a single trial with only 2 outcomes: success (1), failure (0) with probability $(1-p)$.

eg: Coin Toss.

Q5 What is Binomial distribution?

Ans Extension of Bernoulli \rightarrow probability of k success in n independent trials with probability p . example number of heads in 10 coin tosses.

Q6 What is log normal distribution?

Ans If $\log(x)$ is normally distributed, then x follows log-normal. Suitable for positive, skewed like incomes, stock prices or transaction amounts.

Q7 What is Power law Distribution?

Ans A heavy-tailed distribution where small values are common but very large values (outliers) still occur

→ Probability falls as a power of the value ($P(x) \propto x^{-a}$)
(Seen in wealth distribution, internet traffic)

Q8 what is Box-Cox transformation?

Ans A family of power transformation applied to data to reduce skewness and stabilize variance. Helps make data more normally distributed for statistical modeling.

Q9 what is poisson distribution?

Ans Models the number of events happening in a fixed interval (Time, area) given an average rate λ .

Q10 What is Z-score probability?

Ans Standardized Score: $Z = (X - \mu) / \sigma$
Tells us how far data point is from the mean in terms of standard deviations

→ Used to find probabilities and detect outliers.

Q11 Difference between PDF and CDF?

Ans PDF (Probability Density function): Shows likelihood of data near a value. The area under the curve = probability

CDF (Cumulative distribution function): Shows cumulative probability that $X \leq x$. Goes from 0 to 1.

— x — x —