
Part A – Conceptual Questions

1 What is Bernoulli Distribution?

A **Bernoulli distribution** models a single experiment with only **two possible outcomes**: Success (1) or Failure (0).

It is defined by a single parameter **p**, which represents the probability of success.

✚ Example: A transaction being either *Successful* or *Failed*.

2 What is Binomial Distribution?

A **Binomial distribution** models the number of successes in a fixed number of independent Bernoulli trials.

It is defined by:

- **n** → Number of trials
- **p** → Probability of success

✚ Example: Number of successful transactions in a week.

3 What is Poisson Distribution?

A **Poisson distribution** models the number of events occurring in a fixed time interval when events happen randomly and independently.

It is defined by:

- **λ (lambda)** → Average number of events per interval

✚ Example: Number of transactions per day.

4 What is Log-Normal Distribution?

A **Log-Normal distribution** describes data where the logarithm of the variable follows a normal distribution.

It is commonly used for **positively skewed data**.

✚ Example: Transaction amounts, income, stock prices.

5 What is Power Law (Pareto) Distribution?

A **Power Law distribution** describes situations where a small number of observations contribute to a large portion of the total.

It shows **heavy-tailed behavior**.

✦ Example: Few customers generating most of the revenue.

6 What Does “Heavy-Tailed” Mean?

Heavy-tailed data means extreme values (very large observations) occur more frequently than expected in a normal distribution.

✦ This indicates higher risk or high-value outliers.

7 What is a Q-Q Plot?

A **Q-Q (Quantile-Quantile) plot** compares sample data with a theoretical distribution to check normality.

✦ If points lie on a straight line → Data is approximately normal.

8 What is Box-Cox Transformation?

The **Box-Cox transformation** is used to make skewed data more normally distributed.

It helps:

- Stabilize variance
- Improve regression performance

✦ Requirement: Data must be strictly positive.

9 What is Z-Score?

A **Z-score** measures how many standard deviations a value is from the mean.

Formula:

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

✚ Used for detecting outliers and probability calculation.

10 What is PDF and CDF?

- **PDF (Probability Density Function)** shows the likelihood of a variable taking a specific value.
- **CDF (Cumulative Distribution Function)** shows the probability that a variable is less than or equal to a value.

✚ PDF → Single value probability

✚ CDF → Cumulative probability