PART-A : THEORETICAL FOUNDATION

---

1) What is Statistical Distribution?

A statistical distribution describes how the values of a dataset are spread across different ranges.
It shows the probability of occurrence of different outcomes.

In simple words:

It tells how data behaves.

Example:
Heights of people, income, product prices, number of orders per day.

---

2) What is a Q-Q Plot and why is it used?

A Quantile-Quantile (Q-Q) Plot compares the quantiles of a dataset with a theoretical distribution (usually normal distribution).

Purpose:
To check whether data follows normal distribution.

Interpretation:

- Straight line → Normal distribution

- Curved line → Not normal (skewed data)

---

3) Difference Between Discrete and Continuous Distribution

Feature	Discrete Distribution	Continuous Distribution
Values	Countable integers	Infinite decimal values
Example	Number of orders	Transaction amount
Graph	Bar chart	Smooth curve
Example Distribution	Binomial, Poisson	Normal, Log-Normal

---

4) What is Bernoulli Distribution?

Bernoulli distribution models experiments with only **two possible outcomes**:

Success (1) or Failure (0)

Formula:
[
$P(X=x) = p^x (1-p)^{1-x}$
]

Example:
Transaction success or fail in payment gateway.

---

5) What is Binomial Distribution?

Binomial distribution measures number of successes in fixed number of trials.

Conditions:

- Fixed number of attempts

- Two outcomes each time

- Same probability

Formula:
[
$P(X=k)=\binom{n}{k} p^k (1-p)^{n-k}$
]

Example:
Number of purchases a customer makes in a week.

---

6) Log-Normal Distribution

If logarithm of a variable is normally distributed → the variable follows log-normal distribution.

Used when:

- Data cannot be negative

- Data is right skewed

Example:
Income, product prices, transaction amount

---

7) Power Law Distribution

In power law, small events are common but large events are rare.

$$P(x) \propto x^{-\alpha}$$

Meaning:
Few customers spend huge money, most spend small money.

Also known as **80-20 Rule (Pareto Principle)**

---

8) What is Box-Cox Transformation?

A mathematical transformation used to convert skewed data into normal distribution.

$$y(\lambda)=\frac{x^\lambda -1}{\lambda}$$

Purpose:
Improve performance of statistical models and machine learning.

---

9) Poisson Distribution

Used to count number of events happening in fixed time interval.

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

Example:
Number of orders per day

---

10) What is Z-Score Probability?

Z-score measures how far a value is from the mean in terms of standard deviation.

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

Use:
Find probability of extreme values

Example:
Probability transaction exceeds ₹5000

---

11) Difference Between PDF and CDF

Function	Meaning
PDF (Probability Density Function)	Probability at exact value
CDF (Cumulative Distribution Function)	Probability up to that value

[
CDF(x)=P(X \le x)
]

Example:
PDF → probability of spending exactly ₹1000
CDF → probability of spending up to ₹1000