

PR - 4 - Spread locator.

PART - A Theoretical Foundation

Q.1 what is statistical distribution?

→ A statistical distribution shows how data values are spread over possible outcomes. It defines the probability of each value or range.

Ex - Heights of people follow a Normal distribution

Q.2 What is Q-Q Plot and why is it used?

→ A Q-Q (Quantile-Quantile) plot compares two distributions - usually sample data vs. a theoretical distribution - to check normality. If the points form a straight line, the data is approximately normal.

Ex - comparing dataset "spread" values to normal distribution using a Q-Q plot.

Q.3 Difference between Discrete and Continuous distributions.

→ Discrete

Defi → Takes specific values

Continuous

Take any value in a range

Ex → Number of emails per day

Height, weight

Diagram → Bar Graph

Smooth curve

Q.4 What is Bernoulli distribution?

- Describe random experiments with two outcomes: success (1) or failure (0).

Parameters: p = probability of success

Ex - coin toss $\rightarrow P(\text{Head}) = 0.5, P(\text{Tail}) = 0.5$

Diagram formula = $P_x(x) = \begin{cases} p^x(1-p)^{n-x} & n=0,1 \\ 0 & \text{otherwise} \end{cases}$

Q.5 What is Binomial Distribution?

- Represents the number of successes in n independent Bernoulli trials.

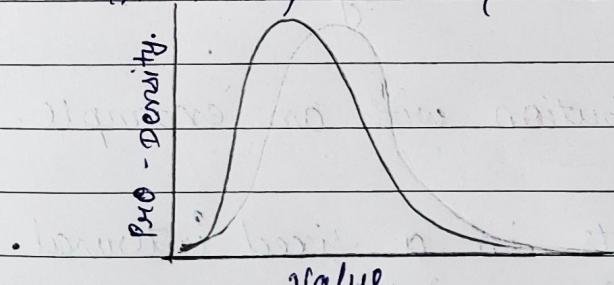
formula $\rightarrow P_x(x) = \begin{cases} \frac{(n)}{x} \cdot p^x \cdot (1-p)^{n-x} & \text{for } n=0,1,\dots,n \\ 0 & \text{otherwise} \end{cases}$

Ex → Getting 3 heads in 5 tosses

Q.6 Explain log-Normal Distribution

- If a variable's logarithm is normally distributed, the variable follows a log-Normal distribution.

Ex - Income, stock price

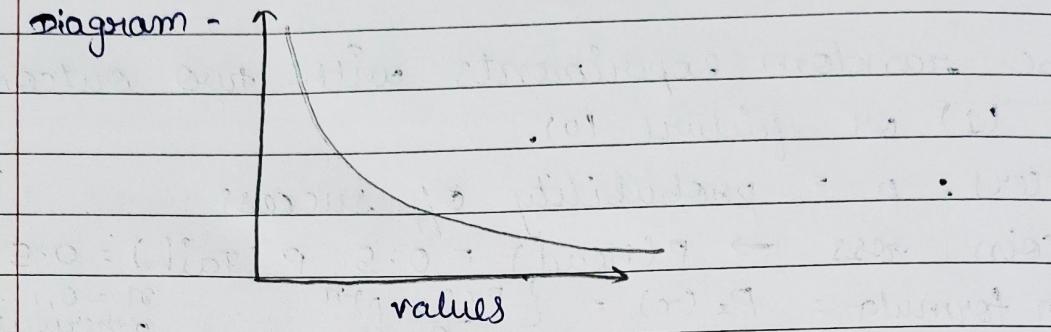


Q.7 Explain Power law Distribution.

- Follows $P(x) \propto x^{-\alpha}$, where few values are large, and most are small. like 20-80 rule.

Ex - Internet traffic, earthquake magnitude

Diagram - ↑



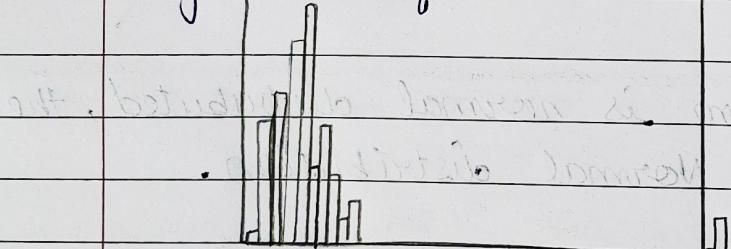
Q.8. What is Box-Cox transform?

→ Box-Cox transformation is used to make data more normally distributed and to stabilize variance.

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

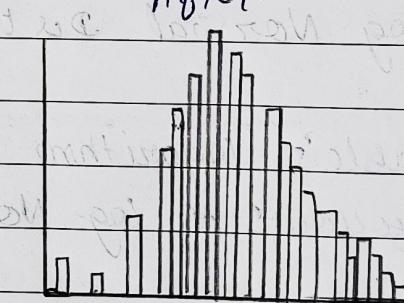
Ex - Applying it to "spread" column to reduce skewness

Diagram - Before



Accuracy = 60%

After.



Accuracy = 80%.

Q.9. Explain Poisson distribution with an example.

→ Models count of events in a fixed interval.

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

Ex - Number of emails received per hour.

Diagram -



Q.10 What is z-score probability?

→ z-score measures how far a value is from the mean in std. deviations.

$$\text{formula} - z = \frac{(x - \mu)}{\sigma}$$

Ex - A z of +2 means values is 20 above the mean.

- finding probability of a stu. scoring above Avg marks.

Q.11 Differentiate PDF and CDF

PDF

CDF

- Probability Density Function

- cumulative distribution funcⁿ

Def. - Probability of a specific value

- cumulative prob. up to a value

Range -

$0 - \infty$

$0 - 1$

Ex

PDF shows weight density at 56 kg

- CDF shows probability of weight ≤ 56 kg

Diagram - Smooth prob. curve.

- Increasing S-shaped curve.