

1 K N

1. What is statistical distributions?

→ A statistical distribution describes how the values of a random variable are spread (or distributed) across possible outcomes.

Types:-

1) Discrete distributions

- Binomial
- Poisson
- Geometric

2) Continuous distributions

- Normal
- Exponential
- Uniform

2. What is a Q-Q plot & why is it used?

→ A Q-Q plot is a graphical tool to check whether a dataset follows a particular distribution

- x-axis - Theoretical quantiles
- y-axis - Sample quantiles

→ The points will approximately lie on straight diagonal line ($y=x$)

3 Diff b/w discrete & continuous	
→ Discrete	Continuous
Possible values Countable ($0, 1, 2, \dots$)	Uncountable (real numbers)
Probability assigned to each value	Defined over an interval
Function PMF (probability mass function)	PDF (probability density function)
Ex: No. of heads in 10 tosses	Height of a person

→ distribution
The binomial distribution is an extension of bernoulli. It models the probability of getting k successes in n independent bernoulli trials.

$$\text{formula: } P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

→ n = number of trials

K = number of successes

p = probability of success in one trial

$$\rightarrow P(K) = \frac{n!}{K!(n-K)!}$$

4 what is bernoulli distribution

- A bernoulli distribution is simplest probability distribution.
- It describes a random experiment with only 2 possible outcome.
- Success - 1 with probability p
Failure - 0 with probability $1-p$

$$\text{Formula: } P(X=x) = p^x (1-p)^{1-x}, x \in \{0, 1\}$$

X = random variable

p = Probability of success

5 Exp log normal distribution

- A log normal distribution describes a positive, right-skewed positive dataset where the logarithm of values follows normal distribution. If a variable x is lognormally distributed then its natural logarithm has a normal distribution.

- The variable x must be greater than zero
- It has a peak near zero & a long tail extending to right

Exp power law distribution?

A power law distribution describes a relationship where the frequency of events follows a power-law function, meaning small occurrences are common, while large ones are rare, creating a long tail.

$$\text{function: } y = Kx^{-\alpha}$$

$\rightarrow K$ & α are constants

\rightarrow No single "typical" size or value, the pattern holds across many scales

What is boxcox transformation?

A boxcox transformation is a powerful statistical technique used to convert non-normally distributed data (like skewed data or data with unequal variance) into a more normal (bell-shaped), making it suitable for methods assuming normality, such as linear regression or ANOVA.

$$\text{for } \lambda \neq 0: y(\lambda) = \frac{y^\lambda - 1}{\lambda}$$

$$\text{for } \lambda = 0: y(\lambda) = \log(y)$$

Exp Poisson distribution

11. The Poisson distribution models the probability of a given number of events happening in a fixed interval of time, space or area if:
- \rightarrow Events occur independently.
 - \rightarrow Events occur at a constant rate.

Formula:

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

$\rightarrow \lambda = \text{no. of occurrence}$

$\lambda = \text{avg rate of occurrence}$

$$e \approx 2.718$$

What is z score probability?

10. The Z score (also called standard score) tells us how many standard deviations a data point is away from mean.

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

$x = \text{data point}$

$\mu = \text{mean}$

$\sigma = \text{standard deviation}$

II Diff. PDF & CDF

PDF:- PDF describes the likelihood of a random variable taking on a specific value

→ The area under curve gives probability

Property:- Always ≥ 0

Area under curve = 1

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

CDF- CDF gives probability that a random variable x is less than or equal to a certain value x .

→ It is obtained by integrating PDF from $-\infty$ to x .

$$\text{Eg: } F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

Property:- Non-decreasing
Smooth for continuous data.