Skewness:

PROBABILITY:

Random variable: A random variable X is a function that assign a real number to each outcome in sample space of random exp.

Discreate: countable no. of distinct values

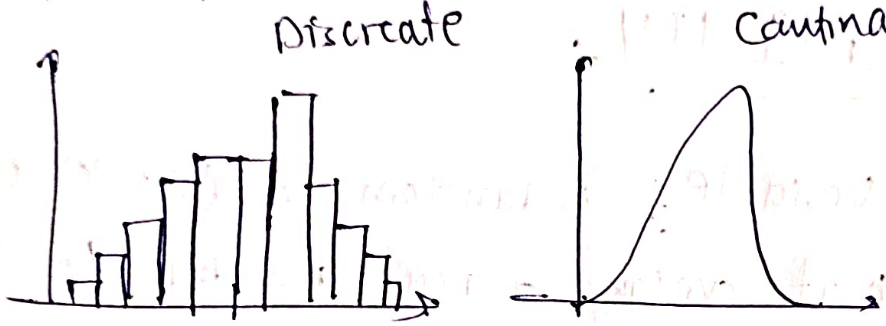
Continuous: take at any value ~~within~~ within a given range or interval.

Probability measured the likelihood of a particular is typically expressed as a no. b/w 0 to 1

$$P(A) = \frac{\text{No. of time A occur}}{\text{Total no. of possible outcome.}}$$

Probability Distribution :-

Probability distributions describe how the probabilities are distributed over the sample space of a random variable



Probabity Distribution function :-

A probability distribution is the mathematical function that gives the probabity of occurrence of diff possible outcome

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

Probability Mass function :-

uses discreat random variables

The sum of all probability $= 1$

PMFs tells the probability of each individual

$P(X = x_i) = p(x_i)$

Cumulative Distribution Function (CDF):

gives the probabilities that a random variable is less than or equal to certain value.

→ CDF is non-decreasing function

X = head → <= →

$$F(X \le head) = P(X = heads) = 0.5$$

X = Tail → last possible outcome:

$$F(X \le Tail) = P(X = head) + P(X = tail) = 1$$

NORMAL DISTRIBUTION:

↳ It is also called Gaussian distribution

↳ Probability distribution is symmetric about mean

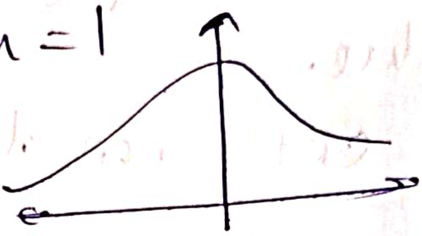
↳ more frequent in occurance than data far from mean

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

Standard Normal distribution:

The standard Normal distribution also called z-distribution or z-score is a special case of normal distribution.

mean=0 standard deviation = 1



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

$\boxed{\sigma = 1}$ $\mu = 0$ $\boxed{f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}}$

Covariance:

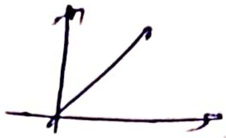
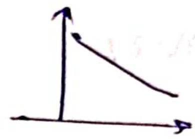
covariance signifies the direction of linear relationship b/w two variables

$$Cov(x,y) = \frac{\Sigma(x_i - \bar{x}) * (y - \bar{y})}{N}$$

Corelation:

correlation analysis is a method of statistical evaluation.

$$\boxed{\text{Correlation} = \frac{Cov(x, y)}{\sigma_x * \sigma_y}}$$

Pearson correlation coeff ÷ as saam as value changes from 1 to -1  , to 

the graph b/w them will be spreading away.

Central limit theorem ÷

The central limit theorem states that when plotting a sample distribution of mean, the mean will be equal to papulation mean.

HYPOTHETICAL TESTING ÷

Medicine A → Medicine B

3hrs

2hr

Result

Result ⟷ Result

campanision → hypothesis testing.

part of statistic analysis, where we test the assumption made regarding a population parameter. It is generally used when we were to compare a single group with an external standard and two or more groups with each other.

Null hypothesis :- Null hypothesis is a statistical theory that suggest there is no statistical significant → H_0 and H-naught (A=B)

Alternative hypothesis :- A<B A≠B there is some statistical difference. H_a, H_1

- Z-test
- T-test
- chi-square test

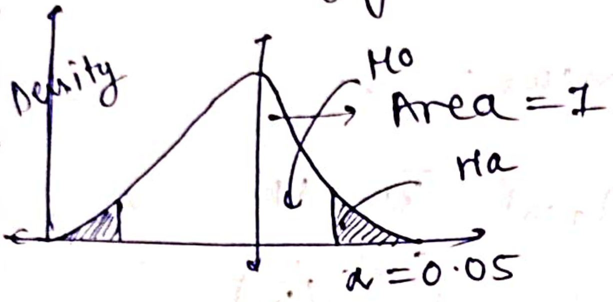
Steps :-

1. Start null (H_0) and alternate (H_1) hypothesis
2. choose level of significance (α)
3. find critical values
4. find test statistic
5. draw your conclusion

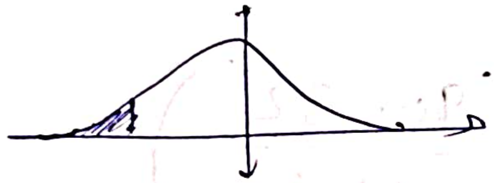
① - States null (Ho) and alternative (H1)

② - choose level of significance

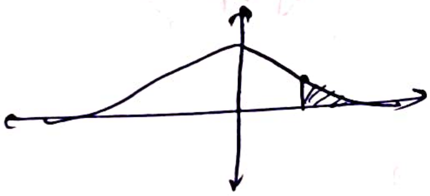
↳ Denoted by alpha a. It has fixed probability of wrongly rejecting a true Null hypothesis.



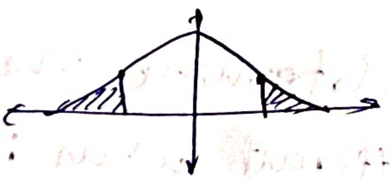
α → provides interval called confidence Interval.



left tailed ÷ new new is worse



Right tailed ÷ new is better



two-tailed ÷ not good not bad.

Find critical values

z test }
T test } Any value.

z test → pop mean stand
sample ≥ 30

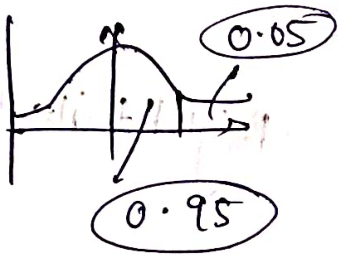
t-test → Pop Mean
Sample → std
sample < 30

Q → Mean score of student = 82
Stand deviation = 20
Sample → 81 student mean = 90.

① → pop std
② → pop Mean
③ → no of sample > 30

$$z_{test} = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})}$$

$H_0 = \mu \neq 82$
$H_a = \mu > 82$



$$\left(\frac{90 - 82}{20/\sqrt{81}}\right)$$

z-test :-

A z-test is used to determine whether two population means are different when the variance are known $(n > 30)$

$$\left(z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right)$$

compare two sample :-

$$\boxed{z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}$$

T-test :-

A t-test used when the population variance is known and the sample size is small ($n < 30$)

formula :- $\boxed{T = \frac{\bar{x} - \mu}{s/\sqrt{n}}}$

↳ $\boxed{T = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}}$

chi - square test :-

A chi - square test is used to test relationship b/w categorical variables. It checks weather the observation freq in each categorical diff significant expected freq

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$O_i \rightarrow$ observed freq

$E_i \rightarrow$ Expected freq.

use case:-

↳ testing the independence b/w two cat

↳ checking the goodness of fit for an observed distributed to a theoritical one.