REVISION

* The set of all possible outcomes of statistical experiment is called sample space (represented by symbol S)

* An event is subset of sample space.

* The complement of event A w.r.t S is the subset of all elements of S that are not in A. We denote the complement of A by A'

* The intersection of two events A and B, denoted by symbol $A \cap B$, is the event containing all elements that are common to A and B.

* Two events A and B are mutually exclusive, or disjoint if $A \cap B = \emptyset$, what is, if A and B have no elements in common.

* The union of two events A and B, denoted by the symbol $A \cup B$, is the event containing all the elements that belong to A or B or both

* Important observations:

$A \cap \emptyset = \emptyset$	$A \cup A' = S$	$(A')' = A$
$A \cup \emptyset = A$	$S' = \emptyset$	$(A \cap B)' = A' \cup B'$
$A \cap A' = \emptyset$	$\emptyset' = S$	$(A \cup B)' = A' \cap B'$

* If an operation can be performed in $\rightarrow n_1$ ways
for which of this a second operation in $\rightarrow n_1 \times n_2$ ways
for which first & second, a third operation in $\rightarrow n_1 \times n_2 \times n_3$ wa
so forth the sequence, a K operation in $\rightarrow$
$n_1 \times n_2 \times n_3 \times \ldots\ldots \times n_{K-1} \times n_K$ ways.

* A permutation is arrangement of all or part of a set of objects.

* The no. of permutations of n objects = $n!$

$$\boxed{{}^{n}P_{r} = \frac{n!}{(n-r)!}}$$ → n distinct objects taken will be r

* The no. of permutations of n objects arranged in circle is $(n-1)!$

* The no. of distinct permutations of n things of which n_1 are of one kind, n_2 of second kind ... n_k of k-th kind is ..

$$= \frac{n!}{n_1!\, n_2! \ldots n_k!}$$

* The no. of combinations of n distinct objects taken r at a time is

$$\binom{n}{r} = \frac{n!}{r! \times (n-r)!}$$

IMP * Probability of event A is the sum of the weights of all sample points in A.

$$0 \le P(A) \le 1, \quad P(\phi) = 0, \quad P(S) = 1$$

If $A_1, A_2, A_3 \ldots A_n$ are mutually exclusive events,

$$P(A_1) + P(A_2) + P(A_3) + \ldots = P(A_1 \cup A_2 \cup A_3 \cup A_4 \ldots)$$

IMP * If an experiment can result in any one of N different equally likely outcomes and if exactly n of these outcomes correspond to event A, then the probability of event A is,

$$P(A) = \frac{n}{N}$$

*. If A and B are two events, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

*. If A and B are mutually exclusive then

$$P(A \cup B) = P(A) + P(B)$$

*. For 3 events A, B, C :

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

*. If A & A' are complementary events,

$$P(A) + P(A') = 1$$

$$P(S) = P(A \cup A') = P(A) + P(A') = 1$$

IMP * Conditional probability :

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$ → If A occurs then B occurs

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$ → If B occurs then A occurs

*. If A & B are independent events :

$P(B|A) = P(B)$ or $P(A|B) = P(A)$

* $P(A \cap B) = P(B \cap A) = P(B) \cdot P(A|B)$

$$= P(A) \cdot P(B|A)$$

* If A & B are independent events,

$$P(A \cap B) = P(A) \cdot P(B) = P(B \cap A)$$

* A collection of events, $A = \{A_1, A_2, A_3 \dots A_n\}$ are mutually independent if for any subset of

$$P(A_{i_1} \cap \dots A_{i_k}) = P(A_{i_1}) \dots \dots P(A_{i_k})$$

IMP

* $$P(A) = \sum_{i=1}^{K} P(B_i \cap A) = \sum_{i=1}^{K} P(B_i) \, P(A|B_i)$$

*. RANDOM VARIABLE :

A random variable is function that associates a real number with each element in sample space.

Ex: Two balls are drawn in succession without replacement from urn of 4 red balls, 3 black balls. The possible outcomes and the values of y of the random variable Y, where Y is no. of reds balls.

Sol: urn contains 4 red, 3 black balls. Y is no. of red balls. $y \in Y$

we have to draw two balls.

Sample space	RR	RB	BR	BB
y	2	1	1	0

i.e. each element of sample space, we are assign a real number.

*. Discrete Random Variable :

If its sample space contains finite no. of elements then variable is called Discrete Random variable.

Ex : No. of VITAP students of weight less than 0 kg.

No. of outcomes by tossing a dice, coins.

* . Continuous random variable :

When a random variable can take on values on a continuous scale, it is called Continuous random variable.

Ex : No. of real numbers between 9 to 10 (so... many)

* . Probability Distribution function : (PDF)

PDF is list of probabilities of values of random variable.

* . Probability Mass function :

The set of ordered pairs $(x, P(X=x))$ is a probability mass function of random variable X if for each possible outcome x,

1. $P(x) \leq 0$ 2. $\Sigma P(X) = 1$

Ex : Find the probability distribution function, if the distribution function is the no. of head by tossing two coins ?

Sample space $= \{HH, HT, TH, TT\}$

X = no. of heads	2	1	0
probability	1/4	2/4	1/4
	{H, H}	{H,T , T,H}	{TT}

one head $\Rightarrow$ $\dfrac{HT,\ TH}{HH,\ TH,\ HT,\ TT} = \dfrac{2}{4} = \dfrac{1}{2}$

zero head $\Rightarrow$ $\dfrac{TT}{HH,\ HT,\ TH,\ TT} = \dfrac{1}{4}$

$\Sigma\ P(X) = \frac{1}{4} + \frac{2}{4} + \frac{1}{4} = 1$

---

*. Example : A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school make a random purchase of 2 of these computers. Find PDF for no. of defectives.

Sol : Let x be a random variable whose values x are the possible numbers of defective computer purchased by the school. Then x can only take the no. of 0, 1, 2. Sample space $= \binom{20}{2}$

Sample space $= \binom{20}{2}$ i.e. $^{20}C_2$

$$f(0) = P(x=0) = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{68}{95}$$

$$f(1) = P(x=1) = \frac{\binom{3}{1} \times \binom{17}{2}}{\binom{20}{2}} = \frac{51}{190}$$

$$f(2) = P(x=2) = \frac{\binom{3}{2} \times \binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}$$

So, the probability distribution of x is,

x	0	1	2
$f(x)$	$\frac{68}{95}$	$\frac{51}{190}$	$\frac{3}{190}$

*. BERNOULLI PROCESS : (All these properties must satisfy)

1. The experiment consists of repeated trials.
2. Each trial results in outcome that may be classified as success or failure.
3. The probability of success, denoted by p, remains constant from trial to trial.
4. The repeated trials are independent. That is, the outcome of any trial doesnot affect the outcomes of other.

Example : Drawing 4 cards from deck without replacement and recording whether they are black or red cards.

event - 1 : $\frac{26}{52}$ = P(picking black card)

event - 2 : No. of black cards left = 26 - 1 = 25

No. of total cards left = 52 - 1 = 51

P(choosing black again) = $\frac{25}{51}$

event - 3 : P(choosing black again) = $\frac{25-1}{51-1}$ = $\frac{24}{50}$ = $\frac{12}{25}$

$\Rightarrow$ event - 1 $\neq$ event - 2 $\neq$ event - 3 (point ③ not satisfied)

$\therefore$ Not a Bernoulli process.

---

*. BINOMIAL DISTRIBUTION :

- The X of success in n Bernoulli trials is called "binomial random variable." The probability of this discrete random variable is Binomial Distribution $b(x; n, p)$

- A Binomial Trial can result in success with probability p and a failure with probability $\boxed{q = 1 - p}$. The p-distribution of binomial random variable X, the no. of success in n independent trials is

$$b(x; n, p) = P(x; n, p) = P(X = x) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, 2 \ldots n$$

IMP *

$$\boxed{P(x) = {}^nC_x \, p^x q^{n-x}}$$

p, q are probabilities

Ex : Find the probability that two of three kids are girls, if a family has 3 kids.
[Solve by general probability the verify the result by Binomial Distribution.] (Consider order)

$$\text{Sample} = \left\{ \begin{matrix} B\ B\ B \\ B\ B\ G \\ B\ G\ B \\ G\ B\ B \\ G\ G\ G \\ G\ G\ B \\ G\ B\ G \\ B\ G\ G \end{matrix} \right\} = 8$$

$$\xrightarrow{\text{From sample space}}$$

$$G = \text{Two girls} = \left\{ \begin{matrix} G\ G\ B \\ G\ B\ G \\ B\ G\ G \end{matrix} \right\}$$

$$\text{probability} = \frac{\text{Two girls}}{\text{total}} = \frac{3}{8} = 0.375 \quad \text{(Success)}$$

By Binomial distribution :

$$b(x; n, p) = {}^{n}C_{x}\, p^{x} q^{n-x}$$

Here, $x = 2$ (because 2 girls is success for us)

$n = 3$ (total kids)

$$p = \frac{1 \text{ boy}}{2 \text{ child}} = \frac{1}{2}$$

$$q = \frac{1 \text{ girl}}{2 \text{ child}} = \frac{1}{2}$$

$$b = {}^{n}C_{x}\, p^{x} q^{n-x}$$

$$= {}^{3}C_{2} \left(\tfrac{1}{2}\right)^{2} \left(\tfrac{1}{2}\right)^{3-2}$$

$$= \frac{3!}{2!\,1!} \left(\tfrac{1}{4}\right)\left(\tfrac{1}{2}\right)$$

$$= 3 \times \frac{1}{4} \times \frac{1}{2} = \frac{3}{8}$$

*. BASIC TERMINOLOGIES :

→ In each trial of experiment conducted under the identical conditions, the outcome is not unique.

e.g., Throwing a coin.

→ Exhaustive events : The total no. of possible outcomes of random experiment.

→ Favourable events : No. of outcomes which involve the happening of event.

→ Sample Space : The set of all possible outcomes of statistical experiment.

→ Sample point : Each outcome of sample space.

→ Axiom of non-negativity : $P(A) \geq 0, \quad A \in B$

Axiom of certainity : $P(S) = 1$

Axiom of Additivity : $P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$

Example: A die is loaded such that even number is twice as likely to occur as odd number. If E is event that no. less than 4 occurs on a single toss of die, find P(E).

1	2	3	4	5	6	
p	2p	p	2p	p	2p	= 9p

$9p = 1 \Rightarrow \boxed{p = \frac{1}{9}}$

Asked $4 = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9}$.

→ Probability of Impossible event is Zero.

→ $P(\bar{A}) = 1 - P(A)$

→ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

→ $P(\bar{A} \cap B) = P(B) - P(A \cap B)$

→ Conditional probability: event A occuring when it is known that some event B occured.

$P(A|B)$ - Conditional probability of A, given B

→ $P(A|B) = \frac{P(A \cap B)}{P(B)}$, $P(B) > 0$

*. RANDOM VARIABLES:

→ no. of sum of sample points of Experiment.

→ Discrete Random Variables: If X assumes only finetly many or at most countably many values and for which the assigned value for variables is depended on chance.

→ probability Model: Collection of all possible values and the probabilities associated with them

→ $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

*. Given $f(x) = \begin{cases} x^2/3 & -1 < x < 2 \\ 0 & \text{elsewhere} \end{cases}$

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt \qquad -\infty \leq x \leq \infty$$

If $f(x) = \frac{d}{dx} F(x)$ exists then,

$$P(a < x < b) = F(b) - F(a)$$

So, For given $f(x)$, the $P(0 < X \leq 1) = ?$

$$F(x) = \int_{-\infty}^{1} f(t)\, dt = \int_{-\infty}^{1} \frac{x^2}{3}\, dx = \frac{x^3+1}{9}$$

*. Let X be random variable with P.D.F $- f(x)$

Mean $= \mu = E(X) = \sum_x x f(x)$; X - discrete

Mean $= \mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$; X - continuous

---

$$f(x) = \begin{cases} \frac{20,000}{x^3} & , x > 100 \\ 0 & , \text{elsewhere} \end{cases}$$

$1 - P(X \leq 99)$

$$P(X \leq 99) = \int_{-\infty}^{99} \frac{20,000}{x^3}$$

*. MEAN OF A RANDOM VARIABLE :

$$\mu = E(X) = \sum_x x \cdot f(x)$$

If X is discrete ↑ and

If X is continuous,

$$\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Example : A lot containing 7 components is sampled by a quality inspector the lot of contains 4 good components and 3 defective components. A sample of 3 is taken by inspector. Find the expected value of the no. of good components in this sample.

Solution : Given,

Let X represented the no. of good components in the sample. The probability distribution of X is

$$P(X = x) = f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, \quad x = 0, 1, 2, 3$$

So,

$$f(x) = \frac{4C_x \times 3C_{(3-x)}}{7C_3}$$

$f(0) = \frac{1}{35}$ $\quad$ $f(2) = \frac{18}{35}$

$f(1) = \frac{12}{35}$ $\quad$ $f(3) = \frac{4}{35}$

Expected value = Mean = μ = $E(X) = \sum_{x} x \cdot f(x)$

$$\mu = 0\left(\frac{1}{35}\right) + 1\left(\frac{12}{35}\right) + 2\left(\frac{18}{35}\right) + 3\left(\frac{4}{35}\right) = \frac{12}{7} = 1.7$$

Example – Continuous R.V :

Consider probability density function, $f(x) = \begin{cases} \frac{20,000}{x^3} & , x > 100 \\ 0 & \text{elsewhere} \end{cases}$

Answer : Expected Value = $\mu = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$

$$\text{Mean} = \mu = \int_{100}^{\infty} x \cdot \frac{20,000}{x^3}\, dx = 20,000 \left[\frac{x^{-2}}{-2}\right]_{100}^{\infty}$$

$$= 20,000 \left[\frac{\infty^{-2}}{-2} + \frac{(100)^{-2}}{2}\right] = 200$$

---

*. Continuous random function $g(x)$:

Mean = $\mu = E(g(x)) = \sum_{x} g(x) \cdot f(x)$ → DISCRETE

Mean = $\mu = E(g(x)) = \int_{-\infty}^{\infty} g(x) \cdot f(x) \cdot dx$ → CONTINUOUS

Example :

Suppose the no. of cars X that pass through a car wash between 4.00 p.m and 5:00 p.m on any sunny Friday has the following PDF,

x	4	5	6	7	8	9
$P(X = x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let $g(x) = 2x - 1$ represents the amount of money, in dollars, paid to the attendant by manager, Find the attendant's expected earnings for this particular time periods.

Answer: The attendant can expect to receive,

$$E[g(x)] = E(2x-1) = \sum_{x=4}^{9} (2x-1)\, f(x)$$

$$E(2x-1) = \sum_{x=4}^{9} (2x-1)\, f(x)$$

$$E(2x-1) = 7\left(\frac{1}{12}\right) + 9\left(\frac{1}{12}\right) + 11\left(\frac{1}{4}\right) + 13\left(\frac{1}{4}\right) + 15\left(\frac{1}{6}\right)$$

$$+ 17\left(\frac{1}{6}\right) = 12.67$$

Consider, $f(x) = \begin{cases} x^2/3 & -1 < x < 2 \\ 0 & \text{elsewhere} \end{cases}$

Find the expected value of $g(x) = 4x+3$

$$E(4x+3) = \int_{-\infty}^{\infty} g(x) \cdot f(x)\, dx = \int_{-1}^{2} (4x+3)\left(\frac{x^2}{3}\right) dx$$

$$= \int_{-1}^{2} \frac{4x^3 + 3x^2}{3}\, dx = \underline{\underline{8}}$$

• If X is random variable with PDF $f(x)$ & mean μ,

Varience of X is $\boxed{\sigma^2 = E[(X-\mu)^2]}$

DISCRETE $\rightarrow$ $\sigma^2 = \sum_x (x-\mu)^2 f(x)$

CONTINUOUS $\rightarrow$ $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$

Standard deviation of X $\longrightarrow$ +ve square root of variance σ

Example : Let the random variable X represent the no. of automobiles that are official business purposes on any given workday. The probability distribution Company A is,

Find variance of probability for Company A.

x	1	2	3
$f(x)$	0.3	0.4	0.3

For DISCRETE :

$$\mu = \sum_x x \cdot f(x)$$

$$= 1(0.3) + 2(0.4) + 3(0.3) = \underline{\underline{2}}$$

$$\sigma^2 = \sum_{x} (x-\mu)^2 \cdot f(x)$$

$$= (1-2)^2 (0.3) + (2-2)^2 (0.4) + (3-2)^2 (0.3)$$

$$= 0.3 + 0 + 0.3 = 0.6$$

$\sigma^2 = 0.6 \rightarrow$ variance

$\sigma = \sqrt{0.6} = 0.77 \rightarrow$ Standard deviation

$\mu = 2 \rightarrow$ Mean

So, Lets calculate for Continuous R.V :

$$f(x) = \begin{cases} 2(x-1) & 1 < x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

$$\mu = \int_{-\infty}^{\infty} x \cdot f(x) = \cancel{\int_{-\infty}^{1}} + \int_{1}^{2} + \cancel{\int_{2}^{\infty}}$$

$$= \int_{1}^{2} x \cdot 2(x-1)\, dx = 2\int_{1}^{2} x^2 - 2\int_{1}^{2} x$$

$$= 2\left[\frac{1}{3}(x^3)_1^2 - \frac{1}{2}(x^2)_1^2\right] = 2\left[\frac{1}{3}(7) - \frac{1}{2}(3)\right]$$

$$= 2\left[\frac{7}{3} - \frac{3}{2}\right] = \cancel{2}\left[\frac{14-9}{\cancel{6}_{3}}\right] = \frac{5}{3} = \mu$$

$$\sigma = \int_{1}^{2} \left(x-\frac{5}{3}\right)^2 \cdot (2x-2) = 2\int_{1}^{2}\left(x-\frac{5}{3}\right)^2 (x-1)\,dx$$

$=$ solve

*. FOR BINOMIAL DISTRIBUTION :

→ Mean, $\mu = np$

→ Variance, $\sigma^2 = npq$

n - no. of trials

p - probability of success

q - probability of failure

Example :

Suppose you flip a fair coin 100 times and let X be the no. of heads. Find mean & variance.

Here, $n = 100$

X = no. of heads of flip of fair coin

$p = \frac{1}{2}$ $\quad q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}$

$\boxed{\mu = np} = 100 \times \frac{1}{2} = 50$ (mean)

$\boxed{\sigma^2 = npq} \Rightarrow \sigma^2 = 100 \times \frac{1}{2} \times \frac{1}{2} = 25$ (variance)

$\sigma = \sqrt{npq} = \sqrt{25} = 5$ (standard deviation).

*. POISSON DISTRIBUTION :

1. No. of trials > 1 (finite n) } This is for Binomial Distribution
2. probability of success must be same for every trial.
3. probability of success must be independent.

$\underbrace{\hspace{6cm}}$

BINOMIAL DISTRIBUTION

Rules for applying Poisson's Distribution :

1. The no. of trials n is very large $(n \to \infty)$

2. The probability of success p for each trials is very small $(p \to 0)$

3. $np = \lambda$ (mean) is (finite).

*. $$\boxed{P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} = P(x; \lambda)}$$ * Formula

$$\Rightarrow P(a < x \leq b) = \sum_{x=0}^{b} \frac{e^{-\lambda}\lambda^x}{x!} - \sum_{x=0}^{a} \frac{e^{-\lambda}\lambda^x}{x!}$$

1. Mean, $\mu = np = \lambda$

2. Variance, $\sigma^2 = \lambda$

3. Standard Deviation, $\sigma = \sqrt{\lambda}$

Example : If probability of bad reaction from certain injection 0.01. Find the change that out of 200 individuals more than two will get bad reaction.

Given, $p = 0.01$

$n = 200$ trials

$P(X > 2) = ?$

$= 1 - P(X \leq 2) = 1 - \sum_{x=0}^{2} P(x; \lambda) = 1 - \sum_{x=0}^{2} P(x; 2)$

$$= 1 - \sum_{x=0}^{2} \frac{e^{-2} 2^x}{x!}$$

$$= 1 - \left[\frac{e^{-2}}{2!} \times 2^0 + \frac{e^{-2}}{2!} 2^1 + \frac{e^{-2}}{2!} \cdot 2^2 \right]$$

$$= 1 - [0.6767] = \underline{0.3233}$$

Example : During a laboratory experiment, the average no. of radio active particles passing through 1 millisecond is 4. What is the probability of 6 particles enter the counter in given millisecond ?

Given, $\lambda = 4$; $x = 6$

$P = ?$

$$\boxed{P(X=6) = P(X;6) - P(X;5)}$$ exactly 6

exactly 6 = upto 6 − upto 5.

$$= 0.8893 - 0.7851 = \underline{0.1042}$$

Example : Given that 2% of fuses manufactured by firm are defective. Find probability that box containing 200 fuses has,

i. Atleast 1 defective fuse.

ii. 3 or more defective fuses.

iii. No defective fuses.

Answer :

$p = 0.02$

$n = 200$

$\lambda = np = 200 \times 0.02 = 4$ (mean, average)

i. Atleast one defective $(X \geq 1)$

$\Rightarrow P(x \geq 1) = 1 - P(x < 1) = 1 - P(x = 0)$

$= 1 - \sum_{x=0}^{0} \frac{e^{-\lambda} \lambda^{x}}{x!} = 1 - \frac{e^{-4} \lambda^{0}}{0!} = 1 - e^{-4}$

$= 1 - 0.0183, = 0.9817$

ii. $P(x \geq 3) = 1 - P(x < 3) = 1 - \sum_{x=0}^{2} P(x; 4)$

$= 1 - 0.2381 = 0.7619$

iii. $P(x = 0) = \sum_{x=0}^{0} P(x; 4) = 0.0183$

---

*. Theoretical :

Very IMP

~~to~~ poisson distribution is limiting case Binomial distribution under following conditions :

i. $n \to \infty$

ii. $p \to 0$

iii. $np = \lambda$

restriction :

$\lambda \to$ finite

(mean is finite)

*. NORMAL DISTRIBUTION : (important)

→ If it is Continuous Random variable, then we can use Normal Distribution.

(also called Gaussian Distribution :)

$$\boxed{P(X = x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

* Normal distribution

where $-\infty < x < \infty$ $\quad \sigma > 0$

$-\infty < \mu < \infty$ $\quad X \sim N(\mu, \sigma)$ → Denoted by

Note : Mean of normal distribution :

$$\text{Mean} = \mu = \int_{-\infty}^{\infty} x \cdot P(x)\, dx$$

Z-Value (Standard normal value), $Z = \frac{\text{Value} - \text{mean}}{\text{Standard deviation}}$

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

x - random variable $\quad \mu$ - average / mean

σ - standard deviation.

*. properties of Normal Distribution :

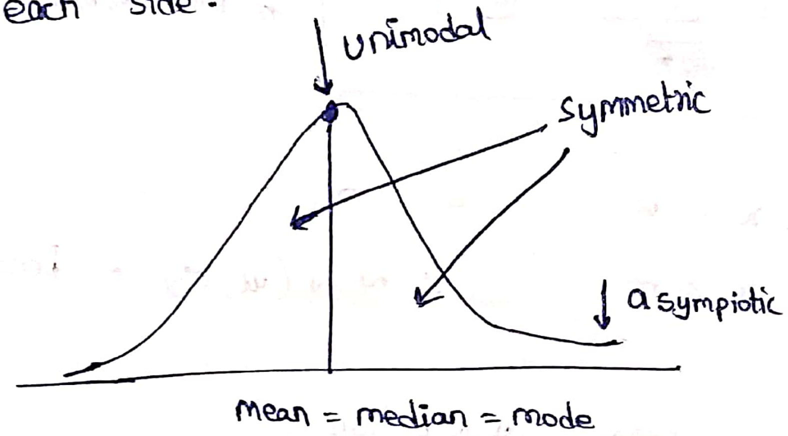
1. The normal distribution curve is Bell-shaped, unimodal (has only one mode).

2. The mean, mode, median are equal & located at center of Distribution.

3. The area under the normal curve is $= 1$.

4. The distribution is symmetric. It extends indefinitely in both directions, approaching but never touching the horizontal axis.

5. The mean divides the area in half, 0.5 on each side.



- The probability that the normal variate x with mean μ and standard deviation σ lies between x_1 & x_2 is given by,

$$P(\mu \leq x \leq x_1) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{\mu}^{x_1} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx$$

If $z = \frac{x-\mu}{\sigma}$ we get $P(0 \leq z \leq z_1) =$

$$= \frac{1}{\sqrt{2\pi}} \int_{z=0}^{z_1} e^{-\frac{z^2}{2}} \, dz .$$

*. Example : (Doubt)

Given a standard normal distribution, find the area under the curve that lies,

a. to the right of $Z = 1.84$ and

b. between $Z = -1.97$ & $Z = 0.86$.

Solution :

$$P(X = x) = \frac{1}{\sigma . \sqrt{2\pi}} \int_{\mu}^{x_1} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

very IMP * $\boxed{Z = \frac{x - \mu}{\sigma}}$ at $x = \mu \Rightarrow Z = 0$

- The point of inflexion of curve are

$\boxed{Z = \mu \pm \sigma}$ *

a. The area to the right of $Z = 1.84$

$= P(Z \geq 1.84) = 1 - P(Z < 1.84)$

$= 1 - 0.9671 = 0.0329$

b. $P(-1.97 \leq Z \leq 0.86)$

$= P(Z < 0.86) - P(Z < -1.97)$

$= 0.8051 - 0.0244 = 0.7807$

__Example :__ Given a standard normal distribution, find the value of K such that $P(z > k) = 0.3015$

a. $P(z > K) = 0.3015$

$1 - P(z < K) = 0.3015$

$P(z < K) = 0.6985$
↓
Search in table

$\Rightarrow \boxed{K = 0.52}$

__Example :__ Given a random variable X having a normal distribution with $\mu = 50$ and $\sigma = 10$, Find the probability that X assumes a value between 45 and 62

$\mu = 50$, $\sigma = 10$ $x_1 = 45$ $x_2 = 62$

$\cancel{z = \mu \pm \sigma}$, $\cancel{50 \pm 10}$

$$Z = \frac{x - \mu}{\sigma} \Rightarrow Z_1 = \frac{45 - 50}{10} = -0.5$$

$$\Rightarrow Z_2 = \frac{62 - 50}{10} = 1.2$$

$$P(45 < x < 62) = P(-0.5 < Z < 1.2)$$

$$= P(z < 1.2) - P(z < -0.5)$$

$$= 0.8849 - 0.3085$$

$$= 0.5764$$

$$\boxed{P(a < x < b) = P(x < b) - P(x < a)}$$

* Normal Distribution

Example : Given a normal distribution with $\mu = 40$ and $\sigma = 6$, find the value of x that has 45% of area to the left.

$\mu = 40$

$\sigma = 6$

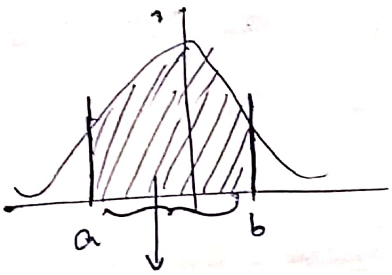
45% of area to the left $= \frac{45}{100}$ of area to left

$\Rightarrow P(z < x) = 0.45 \Rightarrow z = -0.13$

$$\boxed{x = \sigma z + \mu}$$

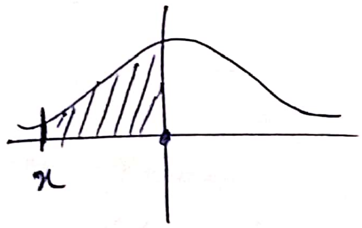
$x = 6(-0.13) + 40$

$= 39.22$



$P(a < x < b)$

Normally, $P(z < x) \Rightarrow$



Application : A certain type of storage battery lasts on average 3.0 years with a standard deviation of 0.5 year. Assuming that battery life is normally distributed, find the probability that a given battery will last less than 2.3 years.

$\mu = 3$, $\sigma = 0.5$

$P(x < 2.3)$

$$z = \frac{x-\mu}{\sigma} = \frac{2.3 - 3}{0.5} = -1.4$$

$$P(x < 2.3) = P(z < -1.4) = 0.0808$$

Application 2 : An electric firm manufactures light bulbs that have a life, before burn-out, that is normally distributed with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a bulb between 778 and 834 hours.

$\mu = 800$, $\sigma = 40$ $\quad 778 < x < 834$

$$z = \frac{x-\mu}{\sigma} \Rightarrow z_1 = \frac{778-800}{40} = -0.55$$

$$z_2 = \frac{834-800}{400} = 0.85$$

$$P(z_1 < z < z_2) = P(-0.55 < z < 0.85)$$

$$= P(z < 0.85) - P(z < -0.55)$$

$$= 0.8023 - 0.2912 = 0.5111$$

---

*. MODULE – 2 : CORRELATION & REGRESSION

Relation between two variables :- Correlation

e.g., Price of petrol with cost of goods.

Price of petrol with no. of vehicles used.

- The sample correlation coefficient r is,

$$\boxed{r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}}$$

, where r – sample correlation coefficient

$S_x =$

$S_y =$

$$\boxed{S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2}$$ * IMP

$$\boxed{S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2}$$ * IMP

$$\boxed{S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}$$ * IMP

→ If $0 < |r| < 0.5$ → correlation is weak.

→ If $0.5 < |r| < 0.8$ → correlation is moderate.

→ If $0.8 < |r| < 1$ → correlation is strong.

1. Karl Pearson's Coefficient of Corelation – Quantity test.
2. Rank Correlation – Quality test

① Karl Pearson's coefficient of Correlation :

The sample correlation coefficient γ is

$$\gamma = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{\text{Covariance of } (x,y)}{\sigma_x \cdot \sigma_y}$$

$\bar{x}$ = mean of x values $\qquad x_i$ = x data, values

$\bar{y}$ = mean of y values $\qquad y_i$ = y data, values

- Co-variance of (x,y) = $cov(x,y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$

- $\sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$ $\qquad$ • $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

- $\sigma_y = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2}$ $\qquad$ • $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$

- $cov(x,y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$

- Correlation Coefficient, (Karl Pearson's) $\boxed{\gamma = \frac{Cov(x,y)}{\sigma_x \cdot \sigma_y}}$ *

*• Based on this, Regression :

$$\begin{cases} 0 < |\gamma| < 0.5 \longrightarrow \text{Correlation is weak} \\ 0.5 < |\gamma| < 0.8 \longrightarrow \text{Corelation is moderate} \\ 0.8 < |\gamma| < 1 \longrightarrow \text{Correlation is strong.} \end{cases}$$

$$\begin{cases} \text{If } \gamma = 1 \longrightarrow \text{Perfect +ve correlation.} \\ \text{If } \gamma = 0 \longrightarrow \text{un correlation.} \end{cases}$$

If $0 < r < 1 \rightarrow$ positive correlation.
If $-1 < r < 0 \rightarrow$ negative correlation.
If $r = -1 \rightarrow$ perfect $-ve$ correlation.

*. Example : Find the correlation coefficient between x and y.

x	1	2	3	4	5	6	7	8	9
y	9	8	10	12	11	13	14	16	15

Solution:

x	y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	9	-4	-3	16	9
2	8	-3	-4	9	16
3	10	-2	-2	4	4
4	12	-1	0	1	0
5	11	0	-1	0	1
6	13	1	1	1	1
7	14	2	2	4	4
8	16	3	4	9	16
9	15	4	3	16	9
$\Sigma x_i = 45$	18	0	0	60	60

$n = 9$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{45}{9} = 5$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{108}{9} = 12$$

So, $$r = \frac{\sum_{i=1}^{9} (x_i - 5)(y_i - 12)}{\sqrt{\sum_{i=1}^{n} (x_i - 5)^2 \sum_{i=1}^{n} (y_i - 12)^2}} = \frac{57}{\sqrt{60 \times 60}} = 0.95$$

*. RANK CORRELATION :

The rank correlation coefficient, * $$\boxed{\rho = 1 - \frac{6 \sum_{i=1}^{n} (x_i - y_i)^2}{n(n^2 - 1)}}$$

Example : The ranks of 15 students in two subjects mathematics & Physics are given :

Math	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Physics	10	7	2	6	4	8	3	1	11	15	9	5	14	12	13

X	Y	(X − Y)	(X − Y)²
1	10	=	=
2	7		
.	2	Calculate	Calculate
.	.		
.	.		
.	.		
13	14		
14	12		
15	13		

$$\boxed{\rho = 0.51}$$

*. LINEAR REGRESSION :

The equation of line of regression

1. Regression equation y on x

$$Y - \bar{y} = b_{yx}(X - \bar{x}) \quad \text{where} \quad b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

2. Regression equation x on y

$$X - \bar{x} = b_{xy}(Y - \bar{y}) \quad \text{where} \quad b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

Here r - Correlation coefficient

$$r = \frac{Cov(x,y)}{\sigma_x \cdot \sigma_y}$$

$$b_{yx} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

$$b_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(y-\bar{y})^2}$$

Example : From the data, obtain the two regression equations

Sales	91	97	108	121	67	124	51	73	111	57
Purchases	71	75	69	97	70	91	39	61	80	47

So, Check the table in next page :

$$\bar{x} = \frac{\sum x}{n} = \frac{900}{10} = 90$$

$$\bar{y} = \frac{\sum y}{n} = \frac{700}{10} = 70$$

$$\boxed{b_{yx} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}}$$ * IMP

$$\boxed{b_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(y-\bar{y})^2}}$$ * IMP

x	y	$x-\bar{x}$	$y-\bar{y} = y-90$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
91	71	1	1	1	1	1
97	75	7	5	49	25	35
108	69	18	-1	324	1	18
121	97	31	27	961	729	837
67	70	-23	0	529	0	0
124	91	34	21	1156	441	714
51	39	-39	-31	1521	961	1209
73	61	-17	-9	289	81	153
111	80	21	10	441	100	210
57	47	-33	-23	1089	529	759
$\Sigma =$ 900	700	0	0	6360	2868	3900

$$\text{mean} = \bar{x} = \frac{\Sigma x}{n} = \frac{900}{10} = 90$$

$$\text{mean} = \bar{y} = \frac{\Sigma y}{n} = \frac{700}{10} = 70$$

$$b_{yx} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} = \frac{3900}{6360} = 0.6132$$

$$b_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2} = \frac{3900}{2868} = 1.361$$

↓

these are regression coefficient found using correlation coefficient

Equation of regression of y on x is,

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$y - 70 = 0.6132(x - 90)$$

$$\boxed{y = 0.6132x + 14.812}$$

Equation of regression of x on y is,

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$x - 90 = 1.361(y - 70)$$

$$\boxed{x = 1.361y - 5.27}$$

Example 2:

Sales	25	28	35	32	31	36	29	38	34	32
Purchases	43	46	49	41	36	32	31	30	33	39

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
25	43	-7	5	49	25	-35
28	46	-4	8	16	64	-32
35	49	3	11	9	121	33
32	41	0	3	0	9	0
31	36	-1	-2	1	4	2
36	32	4	-6	16	36	-24
29	31	-3	-7	9	49	21
38	30	6	-8	36	64	-48
34	33	2	-5	4	25	-10
32	39	0	1	0	1	0
$\Sigma x = 320$	$\Sigma y = 380$	0	0	$\Sigma = 140$	$\Sigma = 398$	$\Sigma = -93$

$$\frac{\Sigma x}{n} = \bar{x} = \frac{320}{10} = 32$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{380}{10} = 38$$

$$\Sigma(x-\bar{x})(y-\bar{y}) = -93$$

$$\Sigma(x-\bar{x})^2 = 140$$

$$\Sigma(y-\bar{y})^2 = 398$$

$$b_{yx} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} = \frac{-93}{140} = -0.6643$$

$$b_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2} = \frac{-93}{398} = -0.2337$$

Equation x on y

$$x - \bar{x} = b_{xy}(y-\bar{y})$$

$$x - 32 = -0.2337(y-38)$$

$$\boxed{x = -0.2337y + 40.8806} \quad *$$

Equation y on x

$$y - \bar{y} = b_{yx}(x-\bar{x})$$

$$y - 38 = -0.6643(x-32)$$

$$\boxed{y = -0.6643x + 59.2576} \quad *$$

MODULE - 3 : INFERENCE OF DECISION MAKING

- The formation of data-based decision procedure can be produced a conclusion about some scientific system.

- Population - The complete collection of all elements to be studied.

- Sample - A sub collection of elements drawn from a population set.

- Hypothesis - A statistical hypothesis is an assertion or conjecter about the parameters of one or more populations. ~~The~~

- Hypothesis Testing - The decision making procedure that about the hypothesis is hypothesis testing.

* Null Hypothesis (H_0) - It is working model that we adopt temporarily or for sake of arguement

* Alternative Hypothisis (H_1) or (H_a) - It contains all the values of the parameter that we will consider plausible if we reject the null hypothesis.

Example : Suppose we are interested in deciding whether or not the mean burning rate of solid propellent used to power aircrew escape systems is 50 cm.

H_0 : $\mu = 50$ cm/s

H_1 : $\mu \neq 50$ cm/s

*· Alternative Hypothesis specifies values of μ that could be either greater or less than 50 cm/s, this is called "TWO-SIDED ALTERNATIVE HYPOTHESIS"

*· Some situations, we may wish to formulate a one-sided alternative hypothesis.

---

*· SIGNIFICANT LEVEL : How much error is there!

- Test the validity of H_0 against H_1 at certain level of significance, i.e. 5% or 1% etc...

Suppose we will not reject the null hypothesis,

$H_0 : \mu = 50$, if $48.5 \leq \bar{x} \leq 51.5$,

we will reject the alternative hypothesis

---

*· ERRORS OF SAMPLING :

① Type I error or α error :

If the null hypothesis H_0 is true but it is rejected by test procedure, then the error made is called Type I error.

② Type II error or β error :

If the null hypothesis, H_0 is false but it is accepted by test, the error commited is called Type II error.

Case	Ho is True	Ho is False
Do not reject Ho	Correct decision	Type II error
Reject Ho	Type I error	Correct decision

<u>Example</u> : A certain type of COVID -19 vaccine is known to be only 25% effective after a period of 2 years. To determine if a new and some what more expensive vaccine is superior in providing protection against the same virus for a longer period of time for 20 people are choosen from the sample.

<u>Solution</u> :

$$P = 25\% = \frac{25}{100} = \frac{1}{4} = 0.25$$

n = no. of people = 20

null Hypothesis, H_0 : $P = 0.25$ (we assume)

alternative Hypothesis, H_1 : $P > 0.25$ (to be conclude) yes or no

one - sided & $P < 0.25$ is not our requirement.

Given, $p = 25\% = \frac{25}{100} = \frac{1}{4} = 0.25$

$H_o : p = 0.25$

$H_a : p > 0.25$ (alternative)

Critical value $= 8$ (assumption)

Donot reject H_o	Reject H_o
$P = 0.25$	$P > 0.25$
0 1 2 3 4 5 6 7 8	9 10 11 12 13 --- 20

Type - I error :

$$\alpha = P(\text{type I error}) = P\left(x > 8,\ p = \tfrac{1}{4}\right)$$

$$= \sum_{x=9}^{20} b\left(x; 20, \tfrac{1}{4}\right) = 1 - \sum_{x=0}^{8} b\left(x; 20, \tfrac{1}{4}\right)$$

$$= 1 - 0.9591 = 0.0409$$

---

*. HYPOTHESIS TESTING :

Example : There are 30 students with results in Mathematics of ABC school with average score of 85. But to perform an Experiment, Randomly 5 students were choosen whose average score was found to be 95. What are the Conclusions ?

Conclusion - 1 : These 30 students are different from ABC school's other class students

i.e. These are two different populations.

Conclusion - 2 : There is no difference at all i.e.
The result is due to random chance only.

*. Which Conclusion is Correct Explaination. Here are some options ? :

1. Increase sample size
2. Test for other samples

$\}$ More time & Budget

3. Calculate random chance probability

↓

NULL AND ALTERNATIVE HYPOTHESES :

- Null Hypothesis (H_0) :

A population parameter (such as mean, the standard deviation, and so on) is equal to a hypothesized value.

In the above Example : The average marks in maths of a class students of ABC school are 85.

$\boxed{H_0 : \mu = 85}$ → Null

- Alternative Hypothesis (H_1) :

A population parameter is smaller, greater or different than the hypothesized value in the null hypothesis.

In the above Example : The average marks in maths of a class students of ABC school are not equal to 85.

$\boxed{H_1 : \mu \neq 85}$ → Alternative

- **Two - sided Alternative Hypothesis :**

- It is also called as " Non directional Hypothesis "
- To determine whether the population parameter is either greater than or less than the hypothesized value.
- Less power than one - sided Alternative hypothesis test.
- In the above Example :
  - The average marks in science of class 9th students of ABC school are not equal to 85.

$$\boxed{H1 : \mu \neq 85}$$

- **One - sided Alternative Hypothesis :**

- Also known as a " Directional hypothesis ".
- To determine whether the population parameter differs from the hypothesized value in specific direction.
- In the above Example, The average marks in ~~science~~ math of a class [illegible] students of ABC school are greater than 85.

$$Ho : \mu = 85 \quad vs \quad H1 : \mu > 85$$

- **What are type-1 error and type - II error :**

Type - I error
Type - II error

- Risks of these two errors are inversely related.
- Determined by level of significance and the power of the test.

Before you define their risks, you conclude

Decision based on sample	Truth about the population	
	H_0 is True	H_0 is False
Fail to reject H_0	Correct Decision (probability $= 1-\alpha$)	Type II error. (probability $= \beta$)
Reject H_0	Type I error (probability $= \alpha$)	Correct decision (probability $= 1-\beta$)

α = level of significance you set for your Hypothesis Test.

Suppose $\alpha = 0.05$, indicates that you are willing to accept a 5% chance that you are wrong when you reject the null hypothesis.

β = depends on the power of the test.

$1-\beta$ = power of test value.

EXAMPLE :

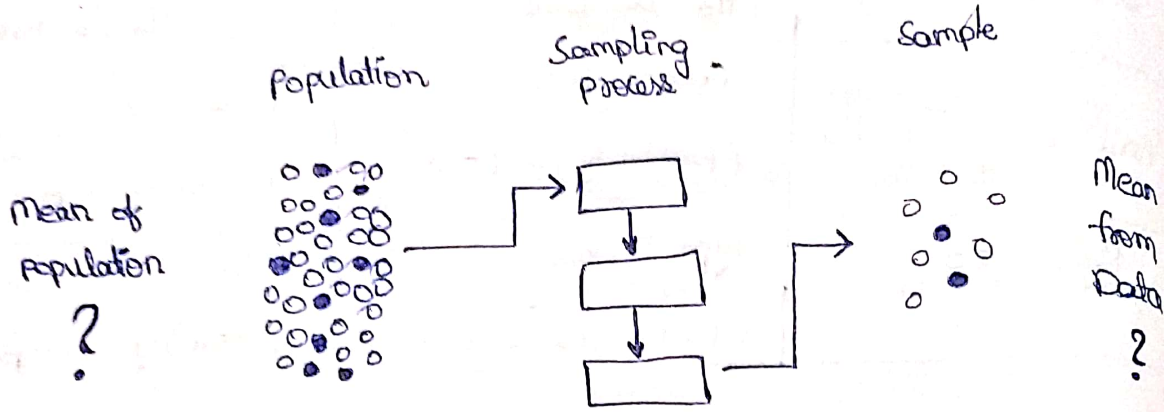
A medical researcher wants to compare effectiveness of two medications,

1. Null Hypothesis (H_0) : $\mu_1 = \mu_2$ (both are equally effective)

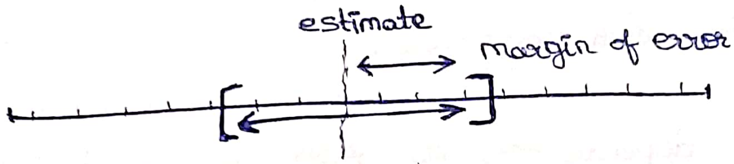
2. Alternative Hypothesis (H_1) : $\mu_1 \neq \mu_2$ (both are not equally effective)

✓ Type I error	✗ Type II error
Rejects H_0, when it is True	Fails to reject H_0, when it is to be,
Considers, $\boxed{\mu_1 \neq \mu_2}$	Considers $\boxed{\mu_1 = \mu_2}$
This is not too Severe, The patient still benefits from the same level of effectiveness	This error is life-threatening If less Medication is sold of more effectiveness.

• Confidence Interval and Confidence Level :



- A confidence interval addresses this issue because it provides a range of values which is likely to contain the population parameter of interest.

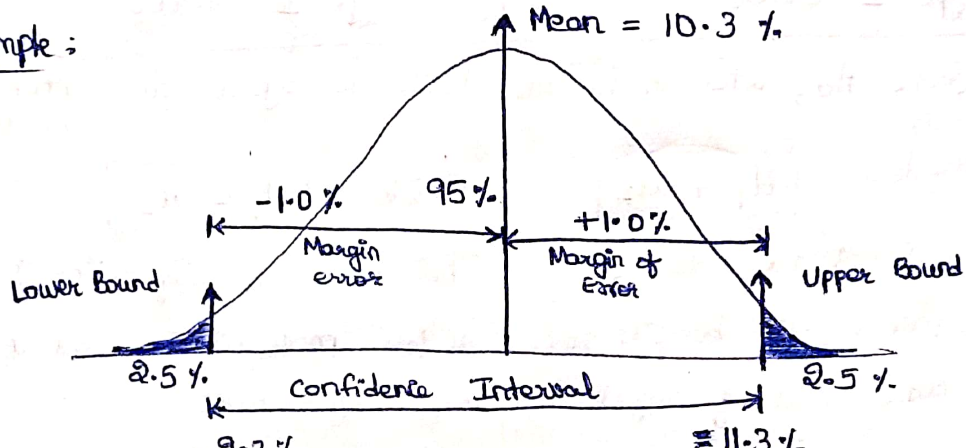


90% confidence interval for parameter

- These confidence intervals are constructed at a confidence level, such as 95%.
- It means that it is probability that confidence interval actually contains the population parameter in approximately 95% of cases.

(So, there is $\frac{1}{20}$ chance (5%) that our confidence interval doesnot include true mean)

<u>Example</u> :



- This means that Sample conversion rate would fall between 9.3 % and 11.3 % in at least 19 of these samples. (Total 5% 20 samples)

- 10.3 % ± 1.0 % at 95 % confidence is our actual conversion rate for this case.

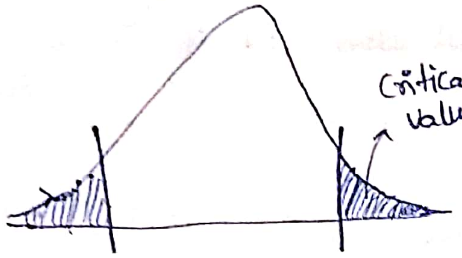
*. What is TEST STATISTIC ?

- A test statistic is random variable that is calculated from sample data and used in hypothesis test.
- To determine whether to reject the null hypothesis.
- To calculate the p-value.
- A test statistic measures the degree of agreement between a sample of data and null hypothesis.

Hypothesis test	Test statistic
Z-test	Z-statistic
t-test	t-statistic
ANOVA	F-statistic
chi-square test	chi-square statistic

*. What is Critical value and p-value ?

- A critical value is a point on the distribution of the test statistic under the null hypothesis that defines a set of values that call for rejecting the null hypothesis.



The probability that the test statistics has value in the rejection region of the test when the null hypothesis is True and equals the significance level of α.

- The p-value is measure for the strength of the evidence in your data

$\downarrow$ p-value $\Longleftrightarrow$ Ho rejection $\uparrow$

- The p-value is the smallest value of α that results in the rejection of Ho.

$\alpha >$ p-value $\longrightarrow$ Fail to reject Ho

$\alpha \leq$ p-value $\longrightarrow$ reject Ho

*. Hypothesis Test :

- The rule that specifies whether to accept or reject a claim about a population depending on the evidence provided by sample of data.

Null hypothesis (H_0) Alternative Hypothesis (H_1)

$\alpha >$ p-Value $\longrightarrow$ Fail to reject Null Hypothesis

$\alpha \leq$ p-Value $\longrightarrow$ reject Null Hypothesis

- When we reject Null Hypothesis $\longrightarrow$ alternative is True
- If we failed to reject null hypothesis $\longrightarrow$ Dont have statistical proof that the null hypothesis is true.

Algorithm :

1. Specify the Hypothesis ($H_0 : \mu = ?$)

Condition to Test	Alternative hypothesis
The population mean is less than the target	one sided $\mu < ?$
The population mean is greater than the target	one sided $\mu > ?$
The population mean is differs from the target	Two sided $\mu \neq ?$

2. Select significance level (also called α)

3. Determine the power and sample size for test.

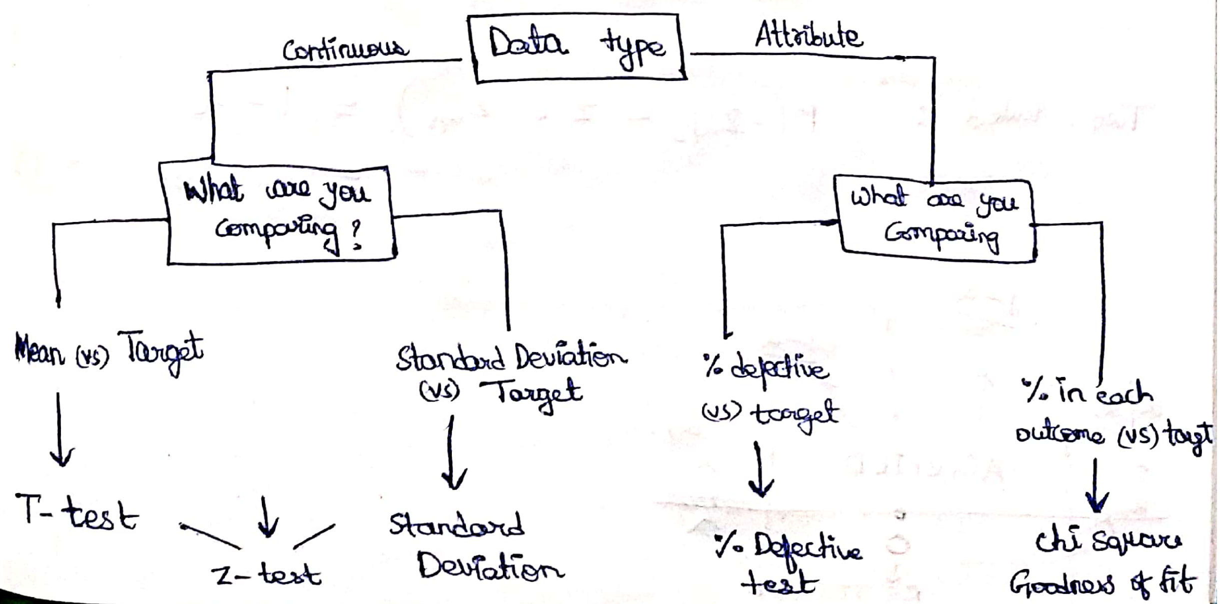
4. Collect the data

5. Compare the p-value from the test to the significance level

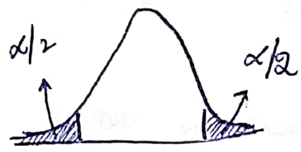
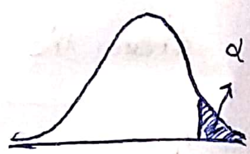
$\alpha >$ p-value or $\alpha \leq$ p-value

6. Decide whether → reject Ho
→ reject the fail to Ho

***. Which type of Hypothesis Test ?**



*. ONE - SAMPLE Z TEST :

One-Tail Test (left tail)	Two-tail Test	One-Tail Test (right Tail)
$H_0 : \mu = \mu_0$	$H_0 : \mu = \mu_0$	$H_0 : \mu = \mu_0$
$H_1 : \mu < \mu_0$	$H_1 : \mu \neq \mu_0$	$H_1 : \mu > \mu_0$
		

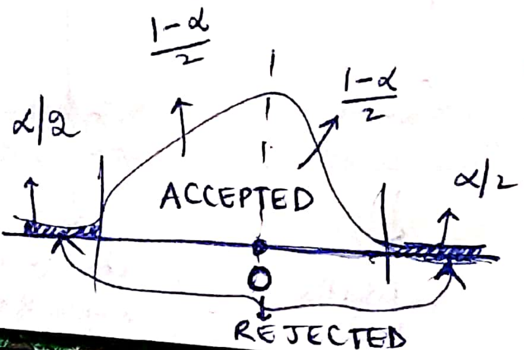
Example : A manufacturer of certain brand of rice cereal claims that the average saturated fat content doesnot exceed 1.5 g per serving. State the null and alternative hypothesis to be used in testing this claim and determine where the critical region is located.

Sol : $H_0 : \mu = 1.5$

$H_1 : \mu < 1.5$

Two - tailed : $P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1 - \alpha$

$= 1 - 0.01 = 0.99$



- For one - tailed :

(1) right tailed :

$P(Z \leq Z_\alpha) = \alpha = 0.01$

$P(Z > Z_\alpha) = 1 - P(Z < Z_\alpha)$

$P(Z > Z_\alpha) = 1 - P(Z > Z_\alpha) = 1 - 0.01 = 0.99$

(2) left - tailed :

$P(Z < Z_\alpha) = \alpha = 0.01$

$P(Z < Z_\alpha) = \alpha = 0.01$ if $Z_\alpha = -2.33$

$|Z_\alpha| = 2.33$

The area under the normal curve with $0.01 = -2.33$
Critical value at level of significance, $(\alpha = 0.01) = -2.33$

- If $\alpha = 0.05$,
the critical values for one sided alternatives is -1.645 and $+1.645$.
for two sided alternatives is -1.96 and $+1.96$.

- If $\alpha = 0.01$, the critical values for
one sided alternatives are -2.33 and $+2.33$
two sided alternative are -2.575 and 2.575

* CRITICAL VALUES of (Z)

S.No.	Alternative Hypothesis	Level of Significance: 0.05 (5%)	Level of Significance: 0.01 (1%)
1	one-sided (right)	$Z_\alpha = Z_{0.05} = 1.645$	$Z_\alpha = Z_{0.01} = 2.33$
2	one-sided (left)	$-Z_\alpha = -Z_{0.05} = -1.645$	$-Z_\alpha = -Z_{0.01} = -2.33$
3	Two sided	$Z_{\alpha/2} = Z_{0.025} = 1.96$	$Z_{\frac{\alpha}{2}} = Z_{0.005} = 2.58$

*. Procedure for testing Hypothesis :

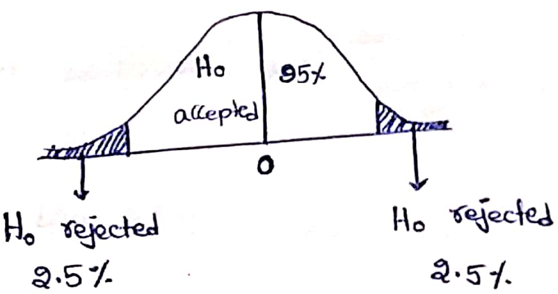
Step-1 : To set the Hypothesis

a) Null (H_0)

b) Alternative (H_a or H_1)

Step-2 : To set suitable significant level

Test the validity of H_0 against H_1 at Certain level of significance (5% or 1% or 10%.)



Step-3 : Test statistic under Ho :

Compute test statistic $Z = \dfrac{t - E(t)}{S.E \text{ of } t}$

$$Z = \frac{\text{observed value} - \text{Expected value}}{\text{Standard Error of } t}$$

under Null Hyp. (Ho)

Here, t - test Sample statistic

SE - Standard Error

Step-4 : Compute the values

Step-5 : Conclusion

Method-i : Compare the test statistic z with critical value z_α at given level of significance (α)

Follow this always

If $|z| \le z_\alpha$, $\rightarrow$ accept Null Hypothesis (H_0)

If $|z| > z_\alpha$, $\rightarrow$ reject Null Hypothesis (H_0)

Method-ii : If $p \le \alpha$ $\rightarrow$ reject Null Hypothesis

If $p > \alpha$ $\rightarrow$ fail to reject Null hypothesis

Write-up :

Reject H_0 in favour of H_1 because of sufficient evidence in data or Fail to reject H_0 because of insufficient evidence in the data.

---

*. p-Value :

- The p-Value is the probability of obtaining a value for test statistic that is as extreme or more extreme than the value actually observed.
- probability (p) is calculated under Null Hypothesis.
- p-Value is the probability of getting the sample data given that the null hypothesis is True, and fails to reject if the p-Value is quite large.

Examples :

If $\alpha = 0.05$ & $p = 0.0341$ $\longrightarrow$ reject H_0

If $\alpha = 0.05$ & $p = 0.0612$ $\longrightarrow$ fail to reject H_0

In simple words,

$p \le \alpha$ $\longrightarrow$ reject H_0, accept H_α

$p > \alpha$ $\longrightarrow$ accept H_0, reject H_α

- Calculation of p-Value :

① Two-tailed test $(\mu \neq \mu_0)$: $p = 2[1 - F(z)]$

② Upper-tailed (Right) test $(\mu > \mu_0)$: $p = 1 - F(z)$

③ lower-tailed (left) test $(\mu < \mu_0)$: $p = F(z)$

(② and ③: one-tailed test)

*• TESTING HYPOTHESIS on MEAN, VARIANCE KNOWN :

① Null Hypothesis H_0 : $\mu = \mu_0$

② Test statistic : $Z_0 = \dfrac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ / $Z_0 = \dfrac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}}$

Example : A random sample of 100 recorded in the united states during the past year showed an average life span of 71.8 years. Assuming a population standard deviation of 8.9 years, does this seem to indicate that the mean life span today is greater than 70 years? Use a 0.05 level of significance.

Null Hypothesis, H_0 : $\mu = 70$

Alternative Hypothesis, H_a : $\mu > 70$ (right tailed)

Total sample size, $n = 100$

mean, $\bar{x} = 71.8$

Standard deviation, $\sigma = 8.9$

$\alpha = 5\% =$ level of significance $= \frac{5}{100} = 0.05$

$Z_{0.05}$ in right-tailed test $= 1.645$

From test statistic, $Z = \dfrac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \dfrac{71.8 - 70}{\frac{8.9}{\sqrt{100}}} = 2.02$

Since right-tailed test,

$0 < Z < Z_{0.05} \rightarrow$ accept H_0 else reject H_0

Since $Z_{0.05} = 1.645$ & $Z = 2.02$

$Z > Z_{0.05} \rightarrow$ reject H_0 $\rightarrow$ accept H_a

$\therefore$ The mean life span today is greater than 70 years.

Example 2 : A company manufacturing electric bulbs claims that the average life of its blubs is 1600 hrs. The average life & standard deviation of random sample of 100 bulbs is 1570 hours & 120 hours respectively. Should we accept the claim of company ?

Mean of population, $\mu = 1600$, σ is not known

Total sample size, $n = 100$

mean $\bar{x} = 1570$, $s = 120$ (of random sample) let Null Hypothesis (H_0) be : $\mu = 1600$

then Alternative Hypothesis (H_a) : $\mu \neq 1600$ } Two tailed

level of significance, let $\alpha = 5\% = 0.05$

then $Z_{\frac{\alpha}{2}} = Z_{\frac{0.05}{2}} = 1.96$ (from table)

So, test statistic :

$$Z = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{1570 - 1600}{\left(\frac{120}{\sqrt{100}}\right)} = -2.5$$

Here, $Z_{\frac{\alpha}{2}} = 1.96$ & $Z = -2.5$

$|Z| = 2.5 > Z_{\frac{\alpha}{2}} \Rightarrow 2.5 > 1.96$

$\therefore$ Null Hypothesis is rejected at 5% Loss of Signifi-

Conclusion : We conclude that the claim of the Company claim that the average life of its bulbs is 1600 hrs shouldnot be accepted at 5% level of significance (LOS).

---

*(IMP) Example 3 : A sample of 450 items are taken from population whose standard deviation is 20. The mean of the sample is 30. Test whether the Sample has come from a population with mean 29. Also calculate the 95% confidence limits of population Mean.

Sol :

Mean of population, $\mu = 29$

Standard deviation of population, $\sigma = 20$ (known)

Total sample size, $n = 450$

Mean of sample, $\bar{x} = 30$

Let Null hypothesis, $H_0 : \mu = 29$

then Alternative hypothesis, $H_a : \mu \neq 29$

$\Big\}$ two tailed test

Level of Confidence limits $= 95\%$

$\Downarrow$

level of significance $= 100\% - 95\% = 5\%$

$Z_{0.05} = 1.96$ (two tailed)

Test statistic value,

$$Z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{30 - 29}{\left(\frac{20}{\sqrt{450}}\right)} = 1.061$$

$|z| = 1.061 < Z_\alpha \Rightarrow 1.0161 < 1.96$

$\therefore$ Null Hypothesis is Accepted, H_0 ✓

Conclusion: The sample has Come from population for mean.

95% confidence limits of population are,

$$\bar{x} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right) \Rightarrow \left[\bar{x} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right), \bar{x} + 1.96\left(\frac{\sigma}{\sqrt{n}}\right)\right]$$

$$= 30 \pm 1.96\left(\frac{20}{\sqrt{450}}\right) \Rightarrow [27.521 , 32.029]$$

*. TESTING THE SIGNIFICANCE OF DIFFERENCE OF MEAN:

Population I

size N_1

mean μ_1

variance σ_1^2

Population II

size N_2

Mean μ_2

Variance σ_2^2

Sample I

size N_1

Mean $\bar{x}_1$

Variance s_1^2

Sample II

size N_2

Mean $\bar{x}_2$

Variance s_2^2

*. TESTING THE SIGNIFICANCE OF DIFFERENCE OF MEAN

- Null Hypothesis : let $\mu_1 - \mu_2 = H_0$
- Alternative Hypothesis : may be $\mu_1 \neq \mu_2$
- Level of significance : either 1%, 5%, 10%, 0.1%

---

*. TEST STATISTIC :

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{\left(\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{N_2}}\right)}$$ when σ_1, σ_2 are known

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$ when σ's are unknown

Conclusion :

*IMP

If $|z| < z_\alpha \longrightarrow$ accept H_0, reject H_a

If $|z| > z_\alpha \longrightarrow$ reject H_0, accept H_a

Example - 1 :

The research investigator is interested in studying whether there is a significant difference in salaries of B.Tech grades in two metro cities. A random sample of size 100 from Kolkata yeilds an average income of Rs. 20,150. Another random sample of 60 from Delhi results in Average income of Rs. 20,250. If Variance of both population are given as $\sigma_1^2 = 40,000/-$ & $\sigma_2^2 = 32,400$ respectively. ($\alpha = 5\%$)

Solution : Given :

Kolkata	Delhi
Size, $N_1 = 100$	Size $n_2 = 60$
mean, $\bar{x}_1 = 20,150$	mean $\bar{x}_2 = 20\,250$
Variance, $\sigma_1^2 = 40000$	Variance $\sigma_2^2 = 32400$

Let Null Hypothesis, $H_0 : \mu_1 = \mu_2$

then Alternative Hypothesis, $H_a : \mu_1 \neq \mu_2$

} Two tailed Test

Given, level of significance, $\alpha = 5\% = 0.05$

$$Z_{\frac{0.05}{2}} = Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$$

Test statistic :

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{n_2}}}, \quad \sigma\text{'s are Known}$$

$$Z = \frac{20,150 - 20\,250}{\sqrt{\frac{40,000}{100} + \frac{32400}{60}}} = 3.26$$

$|z| = 3.26$

$|z| \notin (-1.96, 1.96) \rightarrow |z| > Z_{\frac{\alpha}{2}} \rightarrow$ reject H_0

Conclusion : Here there is no significant difference in Salaries of B.Tech grades in two cities.

---

*. TESTING THE SIGNIFICANCE OF SINGLE PROPORTION :

Suppose a large random sample of size N has a sample proportion p of members possessing a certain attribute (i.e. proportion of success).

- Aim : To test the hypothesis that the proportion P in the population has specified value P_0.

- Procedure :

1. Null Hypothesis : Let $H_0 : P = P_0$
2. Alternative Hypothesis : then $H_a : P \neq P_0$, $P > P_0$, $P < P_0$
3. Level of Significance : $\alpha = 1\%, 5\%, 10\%, 0.01\%$
4. Test Statistic : $$Z = \frac{p - P_0}{\left(\sqrt{\frac{PQ}{n}}\right)}$$
5. Conclussion :

$|z| < Z_\alpha \rightarrow$ accept H_0, reject H_a

$|z| > Z_\alpha \rightarrow$ reject H_0, accept H_a

Example :

In Hospital 480 females & 520 male babies were born in a week. Do these figures confirm the Hypothesis that males & females are born in equal number?

Solution :

Let probability of male happening $= 0.5 = 1/2 \rightarrow P$

Let probability of female happening $= 0.5 = 1/2 \rightarrow Q$

Assumption :

Let Null Hypothesis, H_0 : $P = P_0 = 0.5$

then Alternative Hypothesis, H_a: $P \neq 0.5$

$\}$ Two tailed test

Total births $= 480 + 520 = 1000$

$$\text{probability} = \frac{\text{actual}}{\text{total}}$$

$$\text{Actual probability of female} = \frac{480}{1000} = 0.48$$

$$\text{Actual probability of Male} = \frac{520}{1000} = 0.52$$

Level of significance, $\alpha = 5\% = \frac{5}{100} = 0.05$

$$Z_{\frac{\alpha}{2}} = Z_{0.025} \Rightarrow (-1.96, 1.96)$$

Test statistic :

$$Z = \frac{p - P}{\left(\sqrt{\frac{PQ}{N}}\right)} = \frac{0.48 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}}$$

$$Z = -1.265 \in (-1.96, 1.96)$$

Conclusion : We accept the null hypothesis that the males and females are born in equal ratio.

Example :

A survey was conducted among the citizens of city to study their preference towards consumption of tea and coffee. Among 1000 randomly selected persons, it is found that 560 are tea-drinkers & remaing are coffee drinkers. Can we conclude at 1% LOS from this info that both tea & coffee are equally preferred among citizens in city?

Solution :

Let P denote the proportion of people in city who preferred to consume tea. Then, the null and the alternative hypothesis are :

Null Hypothesis, $H_0 : P = 0.5$ } Two tailed test

Alternative Hypothesis, $H_1 : P \neq 0.5$

Data of the sample : $n = 1000$

No. of tea-drinkers, $= 560$

Sample proportion, $p = \frac{560}{1000} = 0.56$

level of significance, $\alpha = 1\%$

$$Z_{\frac{\alpha}{2}} = Z_{\frac{(\frac{1}{100})}{2}} = (-2.58 \text{ to } +2.58)$$

Here, Test statistic, $Z = \frac{p-P}{\left(\sqrt{\frac{PQ}{N}}\right)} = \frac{0.56 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}}$

$$Z = 3.79 \notin (-2.58, 2.58)$$

$|Z| > 2.58 \longrightarrow$ reject H_0.

Conclusion : $Z \notin (-2.58, 2.58)$

reject H_0 at 1% level of significance, Preference of Tea & Coffee are different.

Example 3 :

A new radar device is being considered for certain missile defense system. The system is checked by experimenting with air craft in which a kill & a no kill is simulated. If, in 300 trials, 250 kills occur, accept or reject, at 0.04 level of significance the claim that the probability of kill with the new System doesn't exceed the 0.8 probability of existing device.

Solution :

Total kills, $n = 300$

Kills happened, $x = 250$

$$\text{probability of kill} = \frac{250}{300} = \frac{x}{n} = 0.83$$

Let Null Hypothesis, $H_0 : P = 0.8$

then Alternative Hypothesis, $H_1 : P < 0.8$ } left tailed test

Level of significance, $\alpha = 0.04 = 4\%$

$$Z_\alpha = Z_{0.04} = -1.75$$

Test statistic,

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{N}}} = \frac{0.83 - 0.8}{\sqrt{\frac{0.8 * 0.2}{300}}} = 1.299$$

Conclusion : $z = 1.299$, $Z_\alpha = -1.75$

$|z| > Z_\alpha$ $1.299 > -1.75$ $\longrightarrow$ accept H_0

(In left tailed test, its opposite)

IN-SHORT

I. Z-TEST :

(i.) Left tailed Test :

$|z| \leq Z_\alpha$ $\longrightarrow$ reject Null Hypothesis (H_0)

$|z| > Z_\alpha$ $\longrightarrow$ accept Null Hypothesis (H_0)

(ii.) Right tailed Test :

$|z| \leq Z_\alpha$ $\longrightarrow$ accept Null Hypothesis (H_0)

$|z| > Z_\alpha$ $\longrightarrow$ reject Null Hypothesis (H_0)

(iii.) Two-tailed Test :

$|z| \notin (-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}})$ $\longrightarrow$ reject Null Hypothesis (H_0)

$|z| \in (-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}})$ $\longrightarrow$ accept Null Hypothesis (H_0)

[OR]

$|z| \leq Z_\alpha$ $\longrightarrow$ accept Null Hypothesis (H_0)

$|z| > Z_\alpha$ $\longrightarrow$ reject Null Hypothesis (H_0)

II. T-Test : The same ↑ is valid for T-Test.

*. Z-TEST ON DOUBLE PROPORTION :

- Let P_1 & P_2 be the proportion of sample sizes n_1 and n_2 and the number of people affected be x_1 & x_2 respectively. Then

$$\hat{P}_1 = \frac{x_1}{n_1} \quad \& \quad \hat{P}_2 = \frac{x_2}{n_2}$$

- The common proportion $P = \dfrac{\hat{P}_1 n_1 + \hat{P}_2 n_2}{n_1 + n_2} = \dfrac{x_1 + x_2}{n_1 + n_2}$

- Test statistic, $Z = \dfrac{\hat{P}_1 - \hat{P}_2 - (P_1 - P_2)}{\sqrt{P(1-P)\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$

Example :

Dell Company manufactures laptops. For quality control, two sets of laptops were asked. In first group, 32 out of 800 were found to contain some sort of defect. In the second group, 30 out of 500 were found to have a defect. Is the difference between the two groups significant with 5% level of significance.

$$\hat{P}_1 = \frac{x_1}{n_1} = \frac{32}{800} = 0.04$$

$$\hat{P}_2 = \frac{x_2}{n_2} = \frac{30}{500} = 0.06$$

$$\alpha = 5\% = 0.05$$

For sample 1, we have that the sample size, $N_1 = 800$

The number of favourable cases is $X_1 = 32$, so then

the sample proportion,

$$\hat{P}_1 = \frac{X_1}{n_1} = \frac{32}{800} = 0.04$$

For Sample 2,

we have that sample size, $N_2 = 500$, the number of favourable cases is $X_2 = 30$, So, the sample proportion

$$P_2 = \frac{X_2}{N_2} = \frac{30}{500} = 0.06$$

The value of the pooled proportion is computed by,

$$\bar{P} = \frac{X_1 + X_2}{N_1 + N_2} = \frac{32 + 30}{800 + 500} = 0.0477$$

Also, the given significance level is, $\alpha = 0.05$

i. Null & Alternative Hypothesis :

The following null and alternative hypothesis need to be

$H_0 : P_1 = P_2$ $\quad$ $H_a : P_1 \neq P_2$

This corresponds to two-tailed test, for which a Z-test for two population proportions needs to be conducted.

ii. Rejection Region :

Based on the information provided, the significance level is $\alpha = 0.05$. the critical value for two-tailed test is $Z_\alpha = 1.96$

The rejection region for this two tailed test is

$R = \{Z : |Z| > 1.96\}$

iii. Test Statistics :

The z-statistic is computed as follow :

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p(1-p)\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} = \frac{0.04 - 0.06}{\sqrt{0.0477(1-0.0477) \times \left[\frac{1}{800} + \frac{1}{500}\right]}}$$

$$= -1.646$$

iv. Decision about Null Hypothesis :

Since it is observed that $|z| = 1.646 \leq z_\alpha = 1.96$

It is concluded that "NULL HYPOTHESIS is ACCEPTED"

v. The 95 % confidence interval for $P_1 - P_2$ is :

$$-0.045 < P_1 - P_2 < 0.005$$

Conclusion : It is concluded that the null hypothesis H_0 is not rejected (H_0 accepted). Therefore, there is not enough evidence to claim that the population proportion P_1 is different than P_2, at 5% LOS.

{ Check more Examples in Text Book }