SEM-4

SAMPLING THEORY - MODULE-5

Sample : A finite subset of a population or universe is called a sample.
(or) A sample is a small portion of the population randomly selected.

Sample size : The number of individuals in a sample is sample size.

Sampling distribution :- Grouping different means according to their frequencies, the frequency distribution so obtained is known as "sampling distribution".

Sampling distribution of Means :-
The sampling distribution of sample means for 2 possible
types : 1) Random sampling with replacement.
2) Random sampling without replacement.
associated with "finite population."

Finite population :-
Consider a finite population of size "N" with mean "μ" & SD, "σ".
All possible samples of size without replacement from this population

where is Mean is ; S.D. $\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$, suppose ; $N > n$, then ; $\mu_{\bar{x}} = \mu = \text{Mean}$

where ; $\frac{N-n}{N-1}$ is known as "Finite population correction factor" "C".

* **Mean :** ; $\boxed{\mu_{\bar{x}} = \frac{\Sigma fx}{\Sigma f}}$ Std. Deviation $\boxed{SD : \sigma_{\bar{x}}^2 = \frac{\Sigma f(x-\mu_{\bar{x}})^2}{\Sigma f}}$

(c)

Infinite population :-

Suppose the samples are drawn from an infinite population or sampling is done with replacement then;

$$\mu_{\bar{x}} = \mu \quad \& \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$\therefore$ Standard Error of mean, $\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$, measures the reliability of the mean as an estimate of population mean μ.

$\therefore$ Standard sample mean = $\boxed{Z = \left[\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right]}$

* Problems & Solutions :-

1) A population consists of 4 numbers :- 2, 3, 4, 5, Consider all possible distinct samples of size two with replacement.

Find :- (a) The population mean (μ)

(b) The population standard deviation (σ)

(c) The sampling distribution mean (SDM) ($\mu_{\bar{x}}$)

(d) The mean of the standard deviation of means.

(e) Standard deviation of sampling distribution ($\sigma_{\bar{x}}$) of means.

Soln :- Given that; The population consists of 4 numbers.

$\Rightarrow$ WKT; "N" is the population size.

$\therefore$ $\boxed{N = 4}$

WKT; "n" is "sample size" listing all possible samples of size 2 from population : 2, 3, 4, 5 with replacement.

(a) Mean of the population = $\frac{\Sigma \text{ population}}{\text{Total no. of population size.}}$ $\Rightarrow \boxed{\mu_{\bar{x}} = \frac{\Sigma fx}{\Sigma f}}$

$\therefore$ Mean $= \frac{2+3+4+5}{4} = \frac{14}{4}$

$\therefore \boxed{\text{Mean} = 3.5}$, Mean

(b) Standard Deviation of population :- $\sigma_{\bar{x}}^2 = \frac{\Sigma f(x-\mu_{\bar{x}})^2}{\Sigma f}$

$\sigma_{\bar{x}}^2 = \frac{[(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2]}{4}$ // $x_0 = 2, x_1 = 3, x_2 = 4, x_3 = 5$

$\sigma_{\bar{x}}^2 = \frac{5}{4}$

$\sigma_{\bar{x}} = \sqrt{5/4} = \sqrt{1.25} \Rightarrow \boxed{\sigma_{\bar{x}} = 1.118032}$, Standard deviation

(c) Sampling with replacement :-

The total no of samples with replacement is ; // N=4

N^n where ; n is sample size listing all possible samples of size 2 from population 2, 3, 4, 5 with replacement.

ie; (2,2) (2,3) (2,4) (2,5) $\Rightarrow \boxed{n = 2}$ // (2,4) & (4,2) (diff is 2)

(3,2) (3,3) (3,4) (3,5)

(4,2) (4,3) (4,4) (4,5)

(5,2) (5,3) (5,4) (5,5)

$\Rightarrow N^n = 4^2 = 16 = N^n$.

Now, we compute statistic the arithmetic mean for each of these "16 samples"

The set of 16 means, $\bar{X}$ of these 16 samples gives rise to distribution of means of samples known as sampling distribn of mean.

[We get : Sampling distribn of mean :- 1^{st} (2,2) $\rightarrow \frac{2+2}{2} = 2$

2^{nd}: (3,2) $\rightarrow$ 2.5, 3^{rd}: (4,2) $\rightarrow \frac{4+2}{2} = 3$, (5,2) $\rightarrow \frac{5+2}{2} = 3.5$] $\longrightarrow$

We write the values from $2 \to 5$ in the form;

Since $\boxed{n=4}$

1st →	2	2.5	3	3.5
2nd →	2.5	3	3.5	4
3rd →	3	3.5	4	4.5
4th →	3.5	4	4.5	5.

frequency : It is no of repetition of val

2 → repeated : once (1) time

2.5 → repeated : 2 time.

3 → 3

3.5 → 4

4 → 3

4.5 → 2

5 → 1

This sampling distribution "means" can be arranged in form :

Sample mean ($\bar{x}$)	2	2.5	3	3.5	4	4.5	5
frequency (f)	1	2	3	4	3	2	1.

(d) The mean of these 16 means is known as " Mean of sampling distribution of means.

ie, $$\mu_{\bar{x}} = \frac{\sum x . f_x}{\sum f} = \frac{[2(1)+2(2.5)+3(3)+4(3)+4.5(3)(5)(1)]}{16.}$$

$$\mu_{\bar{x}} = \frac{56}{16} \Rightarrow \boxed{\mu_{\bar{x}} = 3.5}$$

(e) $$\sigma_{\bar{x}}^2 = \text{Variance} = \frac{\sum x f_x (x-\mu_{\bar{x}})^2}{\sum f}$$

$$= \frac{1}{16}\{1(2-3.5)^2 + 2(2.5-3.5)^2 + 3(3-3.5)^2 + 4(3.5-3.5)^2 + 3(4-3.5)^2 + 2(4.5-3.5)^2 + 1(5-3.5)^2\}.$$

$$= \frac{1}{16}\{2.25 + 2 + 0.75 + 0.75 + 2 + 2.25\}.$$

$$\sigma_{\bar{x}}^2 = 0.625$$

∴ standard deviation of sampling distribution of means is :- $\sigma_{\bar{x}} = \sqrt{0.625}$

$$\therefore \boxed{\sigma_{\bar{x}} = 0.7905694}$$

— * —

Q) Construct SD of means for the population; (3).
3, 7, 11, 15 by drawing samples of size two with replacement
Determine : (a) μ (c) SDM (e) $\sigma_{\bar{x}}$
(b) σ (d) $\mu_{\bar{x}}$

Soln :- (a) Population Mean $= \mu = \dfrac{3+7+11+15}{4} = \dfrac{36}{4}$

$$\boxed{\mu = 9}$$

(b) Population Variance $= \sigma^2 = \frac{1}{4}\{(3-9)^2+(7-9)^2+(11-9)^2 + (15-9)^2\}$.

$= \dfrac{80}{4}$

$\sigma^2 = 20$

$\therefore$ Population standard deviation $\Rightarrow \boxed{\sigma = \sqrt{20}}$

(c) Sampling distribution means (SDM) :-
Total number of samples with replacement is ;
$N^n = 4^2 = \underline{\underline{16}}$ samples.
N is the population size & n is the sample size.
Listing all possible samples of size "$\underline{\underline{2}}$" from population : 3, 7, 11, 15
with replacement, we get 16 samples.

1st (3,3) (3,7) (3,11), (3,15)
2nd (7,3) (7,7) (7,11) (7,15)
3rd (11,3) (11,7) (11,11) (11,15)
4th (15,3) (15,7) (15,11) (15,15).

16 possibilities.

[Sampling distribn of means :- 1st (3,3) $\rightarrow \frac{3+3}{2} = \underline{\underline{3}}$, 2nd, (7,3) $\rightarrow \frac{7+3}{2} = \underline{\underline{5}}$
3rd: (11,3) $\rightarrow \frac{11+3}{2} = \underline{\underline{7}}$, 4th: (15,3) $\rightarrow \frac{15+3}{2} = \underline{\underline{9}}$]

Sampling distribution means are :-

3	5	7	9
5	7	9	11
7	9	11	13
9	11	13	15

This sampling distribution means can be arranged in the form of frequency distribution.

Sampling mean	3	5	7	9	11	13	15
Frequency	1	2	3	4	3	2	1

(d) $\mu_{\bar{x}}$ = Mean of the SDM.

$$\mu_{\bar{x}} = \frac{\sum x f_x}{\sum f} = \frac{1(3)+2(5)+3(7)+4(9)+3(11)+2(13)+1(15)}{16}$$

$$\boxed{\mu_{\bar{x}} = \frac{144}{16} = 9}$$

(e) Variance, $\sigma_{\bar{x}}^2 = \frac{\sum f(x-\mu_{\bar{x}})^2}{\sum f}$

$$= \frac{1}{16}\left\{1(3-9)^2+2(5-9)^2+3(7-9)^2+4(9-9)^2+3(11-9)^2+2(13-9)^2+1(15-9)^2\right\}$$

$$= \frac{1}{16}\{160\}$$

$$\therefore \sigma_{\bar{x}} = \sqrt{10} = \boxed{3.16227 = \sigma_{\bar{x}}}$$

Verification: $\mu_{\bar{x}} = \mu = 9$.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{20}}{\sqrt{2}} = \sqrt{10} = 3.16227$$

3) A population consists of 4 numbers : 3, 7, 11, 15.
find the mean & standard deviation of the sampling distribution of means by considering samplings of size without replacement & verify :- ① $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right]$

② $\mu_{\bar{x}} = \mu$, where $\mu_{\bar{x}}$ is mean of sampling distribution & μ is population mean.

Soln :- w.k.t, Population mean $= \mu = \frac{3+7+11+15}{4} = \frac{36}{4}$

$\therefore \boxed{\mu = 9}$

Standard deviation of population $= \boxed{\sigma = \sqrt{20} = 4.47}$

$\sigma^2 = \frac{1}{4}\left\{(3-9)^2+(7-9)^2+(11-9)^2+(15-9)^2\right\}$

* Sampling distribution Mean (SDM) :-

Sampling without replacement (finite population)

ie, the total number of samples without replacement

is :- $NC_n = 4C_2 = 6$ $\quad \underline{n=2}$ (sample size)

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

$4C_2 = \frac{4!}{2!(4-2)!} = \frac{4!}{2!2!} = \frac{4\times3\times2!}{2!\times2} = 6$

$\therefore$ The six sample spaces are :-

(3, 7, 11, 15) population ↓

(3, 7) (3, 11), (3, 15) (7, 11) (7, 15) (11, 15)

$\therefore$ The sampling means are :- $\xrightarrow{\text{by sample size } n=2}$ {5, 7, 9, 11, 13}.

5   9

7   11

9   13

6 possibilities.

[$\frac{3+7}{2} = \frac{10}{2} = 5$, $\frac{3+11}{2} = \frac{14}{2} = 7$ ...]

∴ Sampling distribution of means is :-

$\bar{x}$	5	7	9	11	13
f	1	1	2	1	1

Now; $\sigma_{\bar{x}}^2 = \frac{\sum f(x-\mu_{\bar{x}})^2}{\sum f} = \frac{1}{6}\{1(5-9)^2+1(7-9)^2+2(9-9)^2+1(11-9)^2+1(13-9)^2\}$.

$= \frac{40}{6} = \boxed{\frac{20}{3} = \sigma_{\bar{x}}^2}$ — ①

Now, Consider ; $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right]$

$RHS = \frac{20}{2}\left[\frac{4-2}{4-1}\right] = 10\left(\frac{2}{3}\right) = LHS = \frac{20}{3}$ from ①..

3) Assmt
A population consists of 4 numbers :- 2,3,4,5, Consider all possible distinct samples of size 2 "without Replacement".

Find : ⓐ μ ⓑ σ ⓒ ~~$\mu_{\bar{x}}$~~ SDM ⓓ $\mu_{\bar{x}}$ ⓔ $\sigma_{\bar{x}}$.

Soln :-

ⓐ Mean population ; $\mu = \frac{2+3+4+5}{4} = \frac{14}{4} = \boxed{3.5 = \mu}$ (found already)

ⓑ Standard deviation population ;

$\boxed{\sigma = 1.118033}$ (already found). $\sigma = \frac{\sum f(x-\mu_x)^2}{\sum f}$.

ⓒ Sampling distribution means :

Sampling without replacement (finite population).

Sampling without replacement

∴ The total number of samples without replacement

is :- $NC_n = \frac{N!}{(N-n)!\,n!} = 4C_2 = \frac{4!}{(4-2)!\,2!} = \frac{4\times3\times2\times1}{2\times2}$

∴ $NC_n = 6$ There are : 6 samples.

The s
(2,3,

The samples are :- $\{(2,3),(2,4),(2,5),(3,4),(3,5)(4,5)\}$

$(2,3,4,5)$

Now, we Compute static Arithmetic mean for 6 samples ....

ie; $\{2.5, 3, 3.5, 4, 4.5\} \longrightarrow \left.\begin{matrix} 2.5 & 3.5 \\ 3 & 4 \\ 3.5 & 4.5 \end{matrix}\right\}$ 6 sample spaces.

∴ Sampling distribution table is Given by;

x	2.5	3	3.5	4	4.5
f	1	1	2	1	1

(d) $\mu_{\bar{x}}$ = Mean of SDM (sampling distribution of means).

$$= \frac{(2.5)1 + 3(1) + 3.5(2) + 4(1) + 4.5(1)}{6} = \frac{21}{6} = 3.5$$

$$\therefore \boxed{\mu_{\bar{x}} = 3.5}$$

(e) $$\sigma_{\bar{x}}^2 = \frac{\Sigma f[x - \mu_{\bar{x}}]^2}{\Sigma f} = \frac{1}{6}\left\{1(2.5-3.5)^2 + 1(3-3.5)^2 + 2(3.5-3.5)^2 + 1(4-3.5)^2 + 1(4.5-3.5)^2\right\}.$$

$$\sigma_x = \sqrt{0.4166} \Rightarrow \boxed{0.645497 = \sigma_{\bar{x}}}$$

Verification :- $\mu_x = \mu = 3.5$

$$\sigma_{\bar{x}}^2 = \left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n}$$

$$= \left(\frac{4-2}{4-1}\right)\frac{4}{2} = \frac{5}{12} = 0.4166.$$

$$\therefore \boxed{\sigma_x = \sqrt{0.4166} = 0.645497}$$

—— * ——

6) Determine the expected number of random samples having their me(2)

ⓐ b/w 22.39 & 22.41 ⓒ less than 22.37

ⓑ Greater than 22.42. ⓓ less than 22.38 / more than 22.41

Given : $N=1500$, $n=36$ $\mu=22.4$, $\sigma=0.048$

Soln:- By data : N = Size of population = 1500.

n = sample size = 36.

No of samples = N_s = 300.

Population mean = $\mu = 22.4$

Standard deviation $\sigma = 0.048$.

Now, by using the standard Normal variate.

ie; $z = \frac{x-\mu}{\sigma}$

In the case of sampling distribution of means, this is given by;

$$z = \frac{\bar{X}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{X}-22.4}{0.008} \quad // \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.048}{\sqrt{36}}.$$

ⓐ For : $\bar{X} = 22.39$; $z = \frac{22.39-22.4}{0.008} = -1.25$

For $\bar{X} = 22.41$; $z = \frac{22.41-22.4}{0.008} = +1.25$

$$\therefore P(22.39 < \bar{X} < 22.41) = P(-1.26 < z < +1.26)$$

$$= 2.\phi(0 \leq z \leq 1.26) \quad // \quad A(1.26) \quad \phi = \frac{1}{\sqrt{2\pi}}\int_0^{1.26} e^{-z^2/2}\,dz$$

$$= 2(0.3962)$$

$$= 0.7924.$$

$\therefore$ Expected no of samples who have mean lying b/w 23.39 to 22.41 is : $300 \times 0.7924 = 237.72 \simeq \boxed{238}$

(b) $P(\bar{X} > 22.42)$

For : $\bar{X} = 22.42$; $Z = \frac{22.42 - 22.4}{0.008} = 2.5$

$\therefore P(\bar{X} > 22.42) = P(Z > 2.50) = 0.5 - \phi(2.50)$ // $\phi(2.50) = \frac{1}{\sqrt{2\pi}} \int_0^{2.50} e^{-z^2/2} dz$.

$\Rightarrow 0.5 - 0.4933$

$\therefore \boxed{P(\bar{X} > 22.42) = 0.0062.}$

$\therefore$ Expected no of samples $= 300(0.0062)$
$= 1.86 \cong \boxed{2}$

(c) $P(\bar{X} > 22.37) = P(Z < -3.8)$
$= 0.5 - \phi(3.8)$
$= 0.5 - 0.4999$
$\therefore \boxed{P(\bar{X} > 22.37) = 0.0001}$

no need.

(d) $P(\bar{X} < 22.38\ \&\ \bar{X} > 22.41) -$
$= P(Z < -2.53\ \&\ Z > 1.26)$
$= 0.5 - \phi(2.53) + 0.5 - \phi(1.26)$ $(0.5 - 1.2377) + (0.5 - 0.9884)$ +
$= 0.0057 + 0.1038 = 0.1095$

$\therefore$ Expected no of samples $= 300(0.1095) \rightarrow 32.85 \cong \boxed{33}$

Asmt Calculate prob that random sample of 16 computers will have an average life of :
(a) less than 775 hrs (c) more than 820 hrs assuming that
(b) b/w 790 & 810 hrs (d) length of life of Comp is approx normally distributed with $\mu = 800$ hrs & SD = 40hrs

[$\sigma_{\bar{x}} = 10$. , $Z = \frac{\bar{X} - 800}{10}$.

(a) $P(\bar{X} < 775) = P(Z < -2.5) = P(Z > 2.5)$
$P(0.5 - \phi(2.5) - P(0 < Z < 2.5)$
(a) $= 0.006$

(b) $P(790 < \bar{X} < 810)$
$P(-1 < Z < 1) = 2P(0 < Z < 1)$
(b) $= 0.6826$

(c) $P(\bar{X} > 820)$
$P(Z > 0) - P(0 < Z < 2)$
$0.5 - \phi(2) = 0.0228$]

Test of hypothesis :-

In order to arrive at a decision regarding the population through a sample of population, we make a certain assumption is reffered to as "hypothesis", which may be / may not be true..

Null hypothesis :- The hypothesis formulated for the purpose of its rejection under the assumption that it is true is called "Null hypothesis".

Alternate hypothesis :- Any hypothesis which is complimentary to the null hypothesis is called "Alternate hypothesis" denoted by H_1.

Test of hypothesis :-
The methods that are used to decide whether to accept or reject a null hypothesis or Alternate hypothesis are called "Test of hypothesis".

Alternate hypothesis :-
The "Alternate hypothesis" can be defined as any of following :-
The H_0 ...

i) $H_1 : \mu \neq \mu_0$ ie, $\mu > \mu_0 / \mu < \mu_0$ is 2 tailed Alternative.

ii) $H_1 : \mu > \mu_0$ is right tailed Alternative.

iii) $H_1 : \mu < \mu_0$ is left tailed Alternative.

The Alternatives (ii) & (iii) are also called "single tailed tests".
where; (i) → is "two-tailed test".

Null hypo
H

Null hypothesis :- Null hypothesis is defined as;

H_0 : The population has an assumed value of mean μ_0

ie; $\boxed{\mu = \mu_0}$

Critical Region :- A region corresponding to a statistic t, in the sample space S which amounts to rejection of the null hypothesis, H_0 is called as "Critical Region".

The sample space S, which amounts to the acceptance of H_0 is called "Acceptance Region".

Level of Significance :-

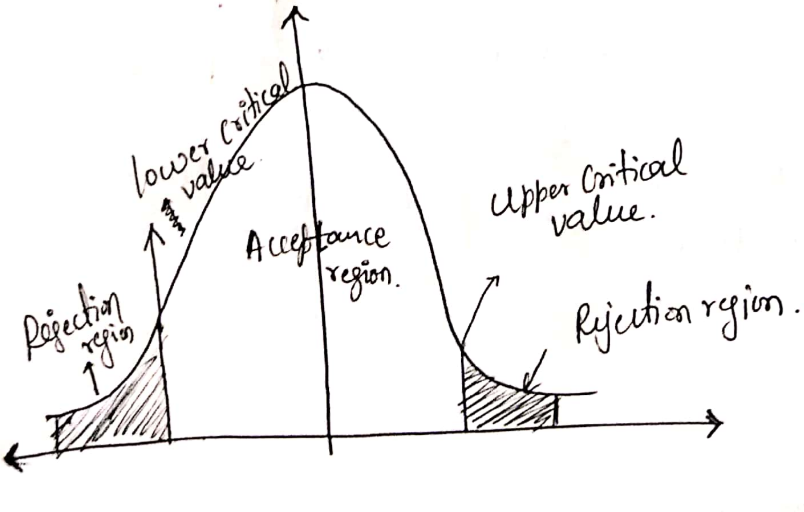
The probability of the value of the variate falling in the critical region is known as "Level of Significance."

Depending on the nature of problem we use a "Single-tail test" (or) "double-tail test" & estimate significance of a result.

For the 3 tests we have 3 results :-

① **Two tailed test :-**

Two tailed test with significance is;



$$\boxed{P(Z > Z_\alpha) = \frac{\alpha}{2}}$$

where; $Z_\alpha \rightarrow$ Critical value, $\alpha \rightarrow$ Total area of Critical region under probability Curve.

$\Rightarrow$ Thus, the Area under each tail is : $\alpha/2$

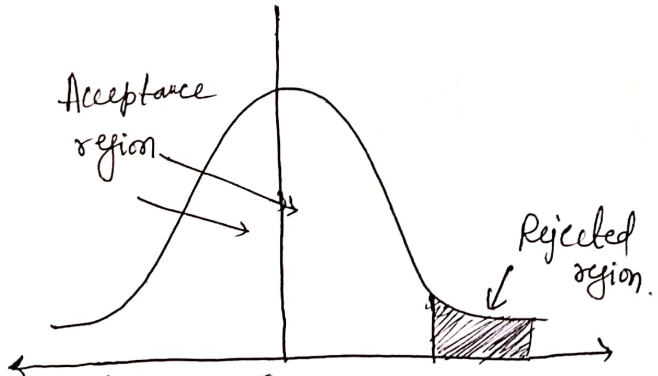
We call this value ; $Z = Z_\alpha$ as the "upper critical value".

$Z = -Z_\alpha$ as "lower Critical value".

$\therefore$ The acceptance region is given by ; $(-Z_\alpha, Z_\alpha)$ (or) $\left(-\frac{Z_\alpha}{2}, \frac{Z_\alpha}{2}\right)$.

2) Right-tailed test :-

Right tailed test curve is :-



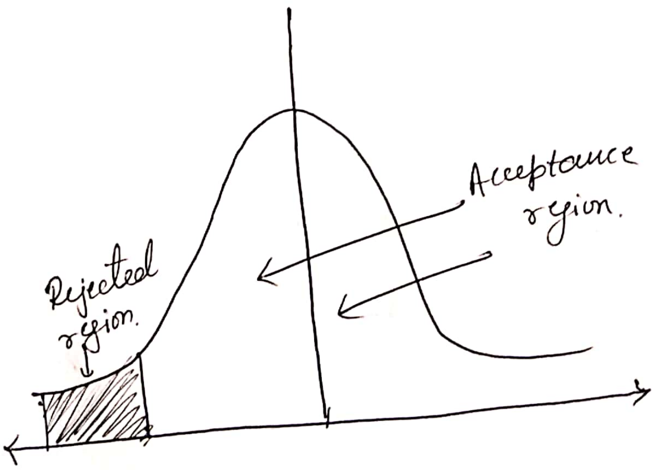
$$\boxed{(P(Z > Z_\alpha)) = \alpha}$$

where; $Z_\alpha \rightarrow$ Critical value.

$\alpha \rightarrow$ Total area of right tail ~~under~~ un probability Curve.

3 Left-tailed test :-

Left tailed test curve is ;



$$\boxed{P(Z < -Z_\alpha) = \alpha}$$

where; $Z_0 \rightarrow$ Critical value.

$\alpha \rightarrow$ Total area of critical region α, is area of left tail under probability Curve.

* The Critical value of Z at different level of Significance :-

$P(|Z| > Z_\alpha) = \alpha$; $P(Z > Z_\alpha) = \alpha$; $P(Z < -Z_\alpha) = \alpha$.

	Level of Significance.	
	1% (0.01)	5% (0.05).
Two tailed test	$\lvert Z_\alpha \rvert = 2.58$	$\lvert Z \rvert = 1.966$
Right tailed.	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$
Left tailed.	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$.

Errors :- If a hypothesis is rejected, while it should have been accepted is known as "Type I Errors." - Type -I- Error.

Type II Error :- If a hypothesis is accepted, while it should have been rejected is known as "Type II Errors".

→ Confidence lim....

* Test of significance for large samples :-

The sample size is taken as large if the sample size "$n > 30$", The sample size is taken as large if the sample size "$n>30$", For such sample we apply normal test as Binomial distribn tends to normal for large n (also for poisson's)

Under large sample test, the following are imp tests to test the significance :-

1. Testing of significance for Single proportion.
2. Testing of significance for diffrence of proportions.
3. Testing of significance for single mean.
4. Testing of significance for difference of means.

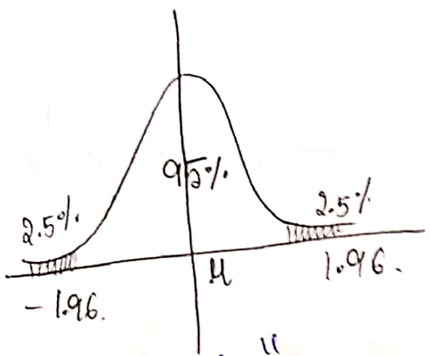
Confidence limits :- The end points of the interval in which population parameter is present is Confidence limits and the interval is called Confidence interval.

$$\therefore \bar{x} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + 1.96\left(\frac{\sigma}{\sqrt{n}}\right) \quad \sim (1)$$

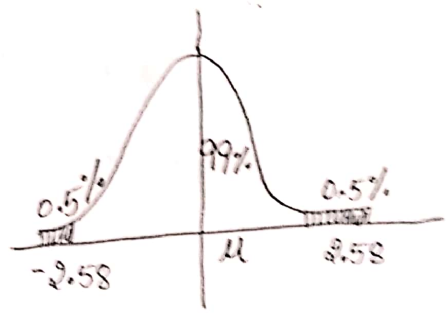
$$\therefore \bar{x} - 2.58\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + 2.58\left(\frac{\sigma}{\sqrt{n}}\right) \quad \sim (2)$$

where, (1) is 95% Confidence ~~limits~~ interval.
(2) is 99% Confidence interval.

Eqn (1) Graph.



Eqn (2) Graph.



The Constants "1.96, 2.58" are called Confidence coefficients denoted by Z_c

A coin is tossed 1000 times & head turns up 540 times, decide on the hypothesis that the coin is unbiased.

Soln:- Let us suppose that the coin is ~~tossed~~ unbiased.

The possibility of getting head in one toss $= \boxed{\frac{1}{2} = p}$

WKT; $p+q = 1$.

$\frac{1}{2} + q = 1$, $\boxed{q = \frac{1}{2}}$, $\boxed{n = 1000}$

$\therefore$ Expected number of heads in 1000 tosses $= np = 1000 \times \frac{1}{2}$ (500.)

$\boxed{np = 500}$

Actual number of heads = 540.

The difference $= x - np$

$= 540 - 500 = \underline{40}$

Consider; $z = \frac{x-np}{\sqrt{npq}} = \frac{40}{\sqrt{1000 \times \frac{1}{2} \times \frac{1}{2}}} = \underline{2.53} < 2.58$

$\therefore \boxed{z = 2.53}$

Since, $2.53 < 2.58$,

$\therefore$ Thus, we can say Coin is unbiased.

3) A random sample of 500 apples was taken from a large consignment & 65 were found to be bad, Estimate the proportion of bad apples in a Consignment as well as the standard error of the estimate, Also find the percentage of bad apples in the consignment.

Soln :- Proportion of bad apples in the sample : $p = \frac{65}{500} = \boxed{0.13 = p}$

NKT; $p+q = 1$, $0.13 + q = 1$, $\boxed{q = 0.87}$

$\therefore$ Standard error proportion of bad apples $= \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.13)(0.87)}{500}}$ Test of

Note

$\therefore$ probable limits of bad apples in consignment

$\Rightarrow \boxed{p \pm 2.58\sqrt{\frac{pq}{n}}}$ // formula

$\Rightarrow 0.13 \pm 2.58\,(0.015)$

$= 0.13 \pm 0.0387 \quad \Rightarrow 0.13 + 0.0387$ and $0.13 - 0.0387$

$\Rightarrow 0.0913$ and 0.1687

(in percentage). ie, $\Rightarrow 0.0913 \times 100\%$ & $0.1687 \times 100\%$

$\Rightarrow 9.13\%$ & 16.87%

Thus, the required percentage of bad apples in consignment lies b/w "9.13 & 16.87"

Assignmt

(3) A die is thrown 324 times & head turned up for odd number : 181 times, ~~Can~~ Is it reasonable to think that die is unbiased one?

Soln:- probability of head turning up of an odd number : $p = 3/6 = \boxed{1/2 = p}$

Since, $p + q = 1$, $1/2 + q = 1$, $\boxed{q = 1/2}$

Expected no of successes $= 1/2 \times 324 = 162$

observed no of successes $= 181$.

$\therefore$ Difference $= 181 - 162 = 19$

Consider; $Z = \frac{x - np}{\sqrt{npq}} = \frac{19}{\sqrt{324 \times 1/2 \times 1/2}} = \frac{19}{9} = 2.11 < 2.58$.

$\therefore \boxed{Z = 2.11}$

Since; $2.11 < 2.58$, then we conclude that die is unbiased.

Test of significance for single proportion:

Note:- (FORMULAS):

1) The mean of the distribution is : $\mu = np$

   Standard deviation is : $\sigma = \sqrt{npq}$

2) The proportion of success are given by:

   (a) Mean proportion of success $= \frac{np}{n} = p$

   (b) S.D proportion of success $= \frac{\sqrt{npq}}{n} = \sqrt{\frac{pq}{n}}$

3) The "standard normal variate" z is given by,

$$z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{npq}} \sim z$$

where; $x \rightarrow$ no of success in a sample size of n & then

$\mu = np \rightarrow$ be the expected no of success

Case (1) If $|z| > 2.58$, then
We conclude the difference is highly significant & reject hypothesis (It is biased)
(or) (Difference is significant at 1% level of significance)

Case (2) If $|z| < 1.96$, then
Difference b/w observed & expected no of success is not significant (Unbiased)

Case (3) If $|z| > 1.96$, then (Reject)
Difference b/w is significant at 5% level of significance

Case (4) If $|z| < 2.58$, then
Diff b/w observed & is not significant (Unbiased)

Problems & Solutions :-

1) A coin was tossed 400 times and the head turned up 216 times, test the hypotheses that the coin is unbiased at 5% level of significance.

Soln :- Let us suppose that the coin is unbiased.

$\therefore$ probability of getting head in a toss $= \boxed{p = 1/2}$ & , $\boxed{n = 400}$

Expected no of successes $= np = \frac{1}{2} \times 400 = \underline{200}$, Since ; $p+q=1$

$1/2 + q = 1$

$\boxed{q = 1/2}$

By data, observed no of success $= \underline{216 = x}$ (say).

The difference ; $x - np = 216 - 200 = \underline{16}$

Also ; SD of simple sampling $= \sqrt{npq} = \sqrt{400 \times \frac{1}{2} \times \frac{1}{2}} = \underline{10}$

Hence, $z = \frac{x-np}{\sqrt{npq}} = \frac{16}{10} = \boxed{1.6 = z}$, $\boxed{z = 1.6}$

$\therefore \boxed{z = 1.6}$, Since, $\boxed{z = 1.6 < 1.96}$

Since, $\boxed{z < 1.96}$, therefore the hypothesis is accepted at 5% level of significance (Not significant)

Hence, we conclude that coin is "Unbiased at 5% level of significance"

2) A Coin is tossed. $\rightarrow$

Prob & Soln :-

① In an Exit poll enquiry it was revealed that 600 voters in one locality & 400 voters from an other locality favoured 55% & 48% respectively, a particular party to come to power, test the hypothesis that there is a diff in locality in respect of opinion.

Soln :- By data : ⇒ P_1 be probability of persons from locality 1 favouring 1^{st} party : $p_1 = \frac{55}{100} = \boxed{0.55 = p_1}$

⇒ P_2 be probability of persons from locality 2 favouring another (2) party : $p_2 = \frac{48}{100} = \boxed{0.48 = p_2}$

⇒ H_0 is the null hypothesis, that there is no difference in locality.

∴ Population proportion $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ where, $n_1 = 600$, $n_2 = 400$

∴ $p = \frac{600(0.55) + 400(0.48)}{600 + 400}$, $\boxed{p = 0.522}$

Also, $p + q = 1$

∴ $\boxed{q = 0.478}$

Consider; $z = \frac{p_1 - p_2}{\sqrt{pq(1/n_1 + 1/n_2)}}$

$z = \frac{0.55 - 0.48}{\sqrt{(0.522)(0.478)(1/600 + 1/400)}}$ ⇒ $\boxed{z = 2.171}$

∴ $z = 2.171 \begin{cases} > z_{0.05} = 1.96 \text{ (Two tailed test)} \\ < z_{0.01} = 2.58 \text{ (Two tailed test)} \end{cases}$

Note :-

Test	Critical values of z.	
	(0.05) 5% level	1% level (0.01
One-tailed test	−1.645 or 1.645	−2.33 or 2.33
two-tailed test	−1.96 and 1.96	−2.58 and 2.58.

(4) A die is thrown 9000 times & a throw of 3 or 4 was observed 3240 times, S.T the die cannot be regarded as an unbiased one.

Soln :- Probability of getting 3 or 4 in a single throw

i.e ; $p = \frac{2}{6} = \boxed{\frac{1}{3} = p}$ $\Rightarrow$ $p+q=1$ $\boxed{q = 2/3}$

∴ Expected no of successes

$= \frac{1}{3} \times 9000 = \underline{3000}$

Observed no of successes = 3240.

The difference = 3240 − 3000 = $\underline{240}$

Consider ; $Z = \frac{x-np}{\sqrt{npq}} = \frac{240}{\sqrt{9000 \times 1/3 \times 2/3}} = \frac{240}{\sqrt{2000}} = \underline{5.37}$

Since; $\boxed{Z = 5.37 > 2.58}$

We conclude, the die is "$\underline{\text{Biased}}$."

Type 2: Testing of Significance for difference of proportions :-

To test the significance of the difference b/w the sample proportions, the test statistic under null hypothesis H_0 that there is no significant diff b/w 2 samples proportions :-

$$\boxed{Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}}$$

, where, $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

& $Q = 1 - P$.

p_1 & p_2 are sample proportions in respect of an attribute corresponding to large samples of size n_1 & n_2.

2) Random sample of 1000 engineering students from a City A and 800 from City B were taken, It was found that 400 students in each of sample were from payment quota, Does the data reveal a significant diffrence b/w 2 Cities in respect of payment quota student.

Soln:- Let Given data :- $n_1 = 1000$, $n_2 = 800$

WKT; $p_1 = \frac{400}{1000} = 0.4$ & $p_2 = \frac{400}{800} = 0.5$

$\therefore p = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} = \frac{1000(0.4) + 800(0.5)}{800 + 1000} = \boxed{\frac{4}{9} = p}$

$\therefore p + q = 1$

$\boxed{q = 5/9}$

Let Ho be the null hypothesis that there is no significant difference b/w 2 Cities;

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(1/n_1 + 1/n_2\right)}}$$

$$z = \frac{0.1}{\sqrt{4/9 \times 5/9 \left(1/1000 + 1/800\right)}} = 4.243$$

$\therefore \boxed{z = 4.243}$

$\therefore z = 4.243 > \begin{cases} z_{0.05} = 1.96 \\ z_{0.01} = 2.58 \end{cases}$

Asmt

③ One type of aircraft is found to develop engine ~~trouble~~ trouble in 5 flights out of 100 & another type in 7 flights out of total of 200 flights, Is there a significant diff in 2 types of aircraft so far as engine defects are Concerned?

$p_1 = 0.05$ $\boxed{p = 0.04}$ $\boxed{z = 0.625}$

$p_2 = 0.035$ $\boxed{q = 0.96}$

$z = 0.625 \begin{cases} < z_{0.05} = 1.96 \ (2\text{-}t\text{-}t) \\ < z_{0.01} = 2.58 \ (2\text{-}t\text{-}t) \end{cases}$

Type:-4

* Test of significance of sample mean :- (Type-3)

1) It has been found from experience that the mean breaking strength of particular brand of thread is 275.6 gms, with SD = 39.7 gms, Recently a sample of 36 pieces of thread showed a mean breaking strength of 253.2 gms, Can one conclude at a significance level of (a) 0.05 or 5% (b) 0.01 (1%), that thread become inferior?

Soln :- We have to decide b/w 2 hypothesis;

$H_0 : \mu = 275.6$ gms, mean breaking strength.

$H_1 : \mu < 275.6$ gms, inferior in breaking strength.

We choose : the "One tailed test".

Mean breaking strength of sample of 36 pieces = 253.2.

$\therefore$ Difference $= 275.6 - 253.2 = \underline{22.4}$; $\boxed{n = 36}$

$$\therefore z = \frac{\text{Difference}}{(\sigma/\sqrt{n})} = \frac{22.4}{(39.7/6)} = \boxed{3.38 = z}$$

$\because$ The value of z is Greater than the critical value of $\boxed{z = 1.645}$ at 5% level & 2.33 at 1% level of significance.

Under the hypothesis H_1, that thread has become inferior is accepted at both 0.05 & 0.01 levels in accordance with one tailed test.

* **Type:4** Testing of Significance for difference of Means :-

1) Intelligent tests were given to 2 groups of boys & girls and data collected is :-

	Mean	S.D	Size
Girls	75	8	60
Boys	73	10	100.

Find out if the 2 mean significantly differ ~~by~~ at 5% level of significance.

Soln :- Null Hypothesis : Ho : There is no significant difference b/w the mean scores ie ; $\bar{x}_1 = \bar{x}_2$

Alternate Hypothesis : $H_1 : \bar{x}_1 \neq \bar{x}_2$

Under the null hypothesis : $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

$$Z = \frac{75-73}{\sqrt{\frac{64}{60} + \frac{100}{100}}} = \frac{2}{1.439}$$

$\therefore \boxed{Z = 1.3898.}$

So, $|Z| = 1.3898 < 1.96$, the significant value of Z at 5% level of significance, Ho is accepted.

ie, There is no significant difference b/w mean scores.

2) In an examination given to students at a large no of different schools the mean grade was : 74.5 and SD is : 8. At one particular school, where 200 students took the examination, the mean grade was : 75.9, Discuss the significance of this result from the view point of (a) One tailed test (b) two tailed test at both 5% & 1% level of significance.

(13)

Soln :- Let H_0 & H_1 be 2 hypothesis.

$H_0 : \mu = 74.5$ & there is no change in mean grade.

$H_1 : \mu \neq 74.5$, ie, $\mu > 74.5$ & $\mu < 74.5$

$\therefore$ $\boxed{\mu = 74.5}$ in H_0 & mean of a sample of size 200(n) is : 75.9

$\therefore$ Difference ; $75.9 - 74.5 = 1.4$.

$\therefore Z = \dfrac{\text{diffence}}{(\sigma/\sqrt{n})} = \dfrac{1.4}{8/\sqrt{200}} = \boxed{2.475 = Z}$

We have the table for the critical values of Z in the case of one & 2 tailed tests..

Test	$Z_{0.05}$	$Z_{0.01}$
One tailed	± 1.645	± 2.33
Two tailed.	± 1.96.	± 2.58

$\therefore$ The calculated value of Z is more than $Z_{0.05}$, $Z_{0.01}$ in 1 tailed test as well as $Z_{0.05}$ in 2 tailed test

Thus, we conclude : The difference in mean grade is significant in these tests, but the same is not significant in 2 tailed test at 1% level of significance

3) A manufacturer claimed that atleast 95% of equipment which he supplied to factory conformed to specifications, An examination of sample of 200 pieces of equipment revealed that 18 of them were faulty, test his claim at significance level of 1% & 5%

$\left[\begin{array}{l} p = 0.95 \\ q = 0.05 \end{array} \right.$ $H_0 : p = 0.95 \rightarrow$ Claim correct, $\boxed{\mu = np = 190}$

$H_1 : p < 0.95 \rightarrow$ claim false $\boxed{\sigma = \sqrt{npq} = 3.082}$

diff = 8, $Z = \dfrac{x - np}{\sqrt{npq}} = \boxed{2.6 = Z}$, the value of z is greater than critical value, claim is rejected.

(14)

2) The average income of persons was Rs: 210 with a S.D: 10 Rs in sample of 100 people of a city. For another sample of 150 persons; the average income was 220 with standard deviation of Rs: 12. The SD of incomes of the people of the city was Rs: 11, Test whether there is any significant difference b/w average incomes of the localities.

Soln :- by data ; $n_1 = 100$, $n_2 = 150$

$\bar{X}_1 = 210$, $\bar{X}_2 = 220$

$\sigma_1 = 10$, $\sigma_2 = 12$

Null hypothesis : There is no difference b/w incomes of the localities, ie, the diff is not significant.

Null hypothesis : $H_0 : \bar{X}_1 = \bar{X}_2$, $H_1 : \bar{X}_1 \neq \bar{X}_2$ (Alternate hypothesis)

Under H_0, we have test statistic ; $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{210-220}{\sqrt{\dfrac{100}{100} + \dfrac{144}{150}}}$

$\therefore \boxed{Z = -7.1428}$

$|Z| = 7.1428 > 1.96$, the significant diff b/w average incomes of the localities

H_0 is rejected, there is significant diff value of z at 5% level of significance.

3) The no of accident per day were studied for 144 days, in town A & 100 days in town B, & follo information was obtained,

	Mean no of accident	SD.
Town A	4.5	1.2
Town B	5.4	1.5

Is the difference b/w mean Accident of 2 towns statistically significant?

$$\left[\; Z = \frac{4.5 - 5.4}{\sqrt{\dfrac{(1.2)^2}{144} + \dfrac{(1.5)^2}{100}}} = \frac{-0.9}{\sqrt{0.01 + 0.0225}} = \frac{-0.9}{\sqrt{0.0325}} = \frac{-0.9}{0.18027} \right.$$

$\therefore \boxed{Z = -4.992} > 2.5$ (or) 1.96 $\therefore$ It is "highly significant"]

STUDENT'S t distribution (Test) :-

If $x_1, x_2, \ldots x_n$ is a random sample of size "n" form a normal population with mean μ and variance σ^2, the student t statistic is defined as :

$$\boxed{t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{\bar{x}-\mu}{S}\cdot\sqrt{n} \quad (\text{or}) \quad \frac{\bar{x}-\mu}{\sqrt{S^2/n}}}$$

where ; $\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$ is the sample mean.

& $S^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i-\bar{x})^2$ is the sample variance.

$V(\text{new}) = (n-1)$, denotes the no of degree of freedom of t.

Note:- ① Under Ho, the test statistic is :-

$t = \frac{\bar{x}-\mu}{S}\sqrt{n}$, where, $\bar{x} = \frac{1}{n}\sum x_i$, $S^2 = \frac{1}{n}\sum (x_i-\bar{x})^2$

* Level of Significance :-

① If $\boxed{|t| > t_{0.05}}$, the diff b/w $\bar{x}$ & μ is said to be significant at 5% level of significance., Ho is Rejected.

② If $\boxed{|t| > t_{0.01}}$, the diff is said to be significant at 1% level of significance, Ho is rejected.

* ③ If $|t| <$ the tabulated t, the data is said to be consistent with hypothesis that μ is mean of population, Ho is accepted, at level of significance adopted.

Formula: Confidence limit : 95% Conf lim (lev of sig 5%) = $\boxed{\bar{x} \pm t_{0.05}\frac{S}{\sqrt{n}}}$

99% Conf lim = $\boxed{\bar{x} \pm \frac{S}{\sqrt{n}} t_{0.01}}$

Problems & Solns :-

1. ~~A machine is a~~ Ten individuals are choosen at random from a population and their heights in inches are :- found to be : 63, 63, 66, 67, 68, 69, 70, 70, 71, 71, Test the hypothesis that the mean height of the universe is 66 inches. [$t_{0.05} = 2.262$ for 9 : d.f]

Soln :- We have ; $\mu = 66$, $n = 10$

$$\therefore \bar{x} = \frac{\sum x_i}{n} = \frac{63+63+66+67+68+69+70+70+71+71}{10} = \frac{678}{10}$$

$$\boxed{67.8 = \bar{x}}$$

WKT ; $S^2 = \frac{1}{n-1} \sum (x - \bar{x})^2$, | $x_0 = 63$, $x_1 = 63$, $x_2 = 66$, $x_3 = 67$, $x_4 = 68$, $x_5 = 69$, $x_6 = 70$, $x_7 = 70$, $x_8 = 71$, $x_9 = 71$.

$$= \frac{1}{10-1} \{ (63-67.8)^2 + (63-67.8)^2 + (66-67.8)^2 + (67-67.8)^2 + (68-67.8)^2 + (69-67.8)^2 + (70-67.8)^2 + (70-67.8)^2 + (71-67.8)^2 + (71-67.8)^2 \}.$$

$$\therefore \boxed{S = 3.011}$$

We have, $t = \frac{\bar{x} - \mu}{S} . \sqrt{n} = \frac{(67.8 - 66)}{3.011} . \sqrt{10}.$

$\therefore t = 1.89 < 2.262.$

Thus, the hypothesis is accepted at 5% level of significance

2. A machine is expected to produce nails of length 3 inches, A random sample of 25 nails gave an average length of 3.1 inch with SD : 0.3, , Can it be said that the machine is producing nails as per specification, ($t_{0.05}$ for 24 df = 2.064).

Soln :- By data :- $\mu = 3$, $\bar{x} = 3.1$, $n = 25$, $S = 0.3$

$$\therefore t = \frac{\bar{x} - \mu}{S} . \sqrt{n} = \frac{0.1}{0.3} \sqrt{25} = 1.67 < 2.064$$

Thus, the hypothesis that the machine is producing nails as per specification is accepted at 5% level of significance.

③ A sample of 10 measurements of the diameter of a sph. gave a mean of 12 cm & a s.d. = 0.15 cm, Find 95% confid. limits for actual diameter.

Test

Soln :- By data : $n = 10$, $\bar{x} = 12$, $S = 0.15$

Also : $t_{0.05}$ for 9 d.f = 2.262

∴ Confidence limits for the actual diameter is given by;

$$\Rightarrow \bar{x} \pm \left[\frac{S}{\sqrt{n}}\right] t_{0.05} = 12 \pm \frac{0.15}{\sqrt{10}}(2.262) = 12 \pm 0.1073.$$

$$\Rightarrow 11.892 \text{ \& } 12.107$$

Thus, 11.892 cm to 12.107 cm is "Confidence limits" for actual diameter

④ A certain stimulus administered to each other of the 12 patients resulted in follo change in blood pressure : 5, 2, 8, −1, 3, 0, 6, −2, 1, 5, 0, Can it be concluded that the stimulus will increase blood pressure? ($t_{0.05}$ for 11 df − 2.201)

Soln :- $\bar{x} = \frac{\Sigma x}{n} = \frac{31}{12} = 2.5833$

$$S^2 = \frac{1}{n-1}\Sigma(x-\bar{x})^2 = \frac{1}{n-1}\left\{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2\right\}$$

$$S^2 = \frac{1}{11}\left\{185 - \frac{1}{12}(31)^2\right\} = 9.538 \qquad \therefore \boxed{S = 3.088}$$

We have, $t = \frac{\bar{x} - \mu}{S} \cdot \sqrt{n}$

Let us suppose that, the stimulus administration is not accomplished with increase in B.P, we can take; $\mu = 0$

$$\therefore t = \frac{2.5833 - 0}{3.088}\sqrt{12}$$

$$\therefore t = 2.8979 \cong 2.9 > 2.201$$

Hence, the hypothesis is Rejected at 5% level of significance, We conclude with 95% confidence that stimulus in general is accompained with increase in B.P.

(16)

(5) A group of boys & girls were given intelligent test, the mean, SD score & numbers in each group are as follows:–

	Boys	Girls
Mean	74	70
SD	8	10
n	12	10.

Is there diff b/w the means of 2 groups significant at 5% level of significance ($t_{0.05} = 2.086$ for 20 d.f).

Soln :– $\bar{x} = 74$, $S_1 = 8$, $n_1 = 12$ (Boys)

$\bar{y} = 70$, $S_2 = 10$, $n_2 = 10$ (Girls)

Also; $$t = \frac{\bar{x} - \bar{y}}{S\sqrt{1/n_1 + 1/n_2}}$$

where, $$S = \frac{1}{n_1+n_2-2}\left\{\sum_{i=1}^{n_1}(x_i-\bar{x})^2 + \sum_{j=1}^{n_2}(y_j-\bar{y})^2\right\}.$$

(or) $$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1+n_2-2}, \quad S^2 = \frac{12(64)+10(100)}{20}.$$

$S^2 = 88.4$, $\boxed{S = 9.402 \approx 9.4}$

Hence, $$t = \frac{74-70}{9.4\sqrt{1/12 + 1/10}} = 0.994$$

$\therefore \boxed{t = 0.994 < t_{0.05} = 2.086}$

Thus, the hypothesis that there is a diff b/w the means of 2 groups is accepted at 5% level of significance.

(6) A group of 10 boys fed on diet A & another group of 8 boys fed on diet B for a period of 6 months recorded the following increase in weights (lbs).

Diet (A)	5	6	8	1	12	4	3	9	6	10.
Diet (B)	2	3	6	8	10	1	2	8		

Test, whether diets A & B differ significantly regarding their effect on increase in weight.

Soln:- Let the variable x correspond to diet A & y to B.

$$\therefore \bar{x} = \frac{\sum x}{n_1} = \frac{64}{10} = 6.4 \;; \quad \bar{y} = \frac{\sum y}{n_2} = \frac{40}{8} = 5$$

$$\Rightarrow \sum_1^{n_1} (x-\bar{x})^2 = 102.4 \;; \quad \sum_1^{n_2} (y-\bar{y})^2 = 82.$$

WKT; $$S^2 = \frac{1}{n_1+n_2-2} \left\{ \sum_1^{n_1} (x-\bar{x})^2 + \sum_1^{n_2} (y-\bar{y})^2 \right\}.$$

$$S^2 = \frac{1}{16} \left\{ (102.4 + 82) \right\} = \frac{184.4}{16} = 11.525$$

$$\therefore \boxed{S = 3.3395}$$

Concider, $$t = \frac{\bar{x}-\bar{y}}{S\sqrt{1/n_1 + 1/n_2}} = \frac{1.4}{3.395\sqrt{1/10 + 1/8}} = 0.86935 \cong 0.87$$

But, $t_{0.05}$ for 16 d.f $= 2.12$ from tables, $t = 0.87$ is less than the table value for 16 d.f at 5% level of significance

Thus, we Conclude that the 2 diets do not ~~diffient~~ differ significantly regarding their effect on increase in weight

* Chi-square (χ^2) test as a test of goodness of fit :-

[The quantity (χ^2) is a greek letter, pronounced as "Ki"]

If $O_1, O_2, \ldots O_n$ be a set of observed (experimental) frequencies and $E_1, E_2, \ldots, E_n$ be the corresponding set of expected (theoretical) frequencies, then χ^2 is defined by the relation.

$$\chi^2 = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2} + \cdots + \frac{(O_n-E_n)^2}{E_n}.$$

(or) $$\boxed{\chi^2 = \sum_{i=1}^{n} \frac{(O_i-E_i)^2}{E_i}}$$

with $(n-1)$ degrees of freedom ($\Sigma O_i = \Sigma E_i = n$ = total frequency)

* Degree of freedom :-

The number of degrees of freedom is the total no of observations less the number of independent constraints imposed on the observations.

It is denoted by; "ν" (nu)

$\boxed{\nu = n-k}$, where $K \rightarrow$ no of independent constraints in set of data of n observations.

* χ^2 : Test as a test of Goodness of Fit :-

Procedure :-

Step : 1 Set up the null hypothesis & Calculate;

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i-E_i)^2}{E_i}$$

Step: 2 Find the degree of freedom (d.f) and read the corresponding values of χ^2 at a prescribed significance level from table.

Step: 3 If calculated value of χ^2 is less than corresponding tabulated value then it is said to be "non-significant" at required level of significance.
ie, data do not provide any evidence against null hypothesis & may be concluded that; there is good correspondence b/w theory & experiment.

Step: 4 On the other hand, if calculated value of χ^2 is greater than the tabulated value it is to be significant & we reject null hypothesis, thus we conclude that exp doesn't support theory.

Note :- (1). If $\chi^2 = 0$, then O_i and E_i agree exactly.

(2). If $\chi^2 > 0$, then

(a) χ^2 small : O_i are close to E_i indicating "Good fit."

(b). χ^2 large : O_i differ considerably from E_i, indicating "Poor fit."

Problems & Soln

1. A die is thrown 264 times & the number appearing on the face (x) follows the following frequency distribution.

x	1	2	3	4	5	6
f	40	32	28	58	54	60.

, Calculate the value of χ^2 (Ki) square.

Soln:- The frequency in the given data are the observed frequencies. Assuming that the dice is unbiased, the expected number of frequencies for the numbers : $\{\underbrace{1,2,3,4,5,6}_{\text{dice-6.}}\}$ to appear on the face is ; $\frac{264}{6} = 44$ each

Now, the data is as follows;

No of on the dice.	1	2	3	4	5	6
Observed frequency (Oi)	40 (O_1)	32 (O_2)	28 (O_3)	58 (O_4)	54 (O_5)	60 (O_6)
Expected frequency (Ei)	44 (E_1)	44 (E_2)	44 (E_3)	44 (E_4)	44 (E_5)	44 (E_6)

WKT; $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$

$$= \frac{(40-44)^2}{44} + \frac{(32-44)^2}{44} + \frac{(28-44)^2}{44} + \frac{(58-44)^2}{44} + \frac{(58-44)^2}{44} + \frac{(60-44)^2}{44}$$

$$= \frac{1}{44}\{16+144+256+196+100+256\}.$$

$$= \frac{968}{44} \; 22$$

$$\therefore \boxed{\chi^2 = 22.}$$

2. In experiments on pea breeding the following frequencies of seeds were obtained.

Round & Yellow	Wrinkled & Yellow	Round & Green	Wrinkled & Green	Total
315	101	108	32	556

Theory predicts that the frequencies should be in proportions 9:3:3:1, Example the correspondance, between theory & Experiment.

Soln:- The corresponding frequencies are :-

Given: Ratios are : 9:3:3:1 ⇒ 9+3+3+1 = 16 total.

∴ frequencies are :- $\frac{9}{16} \times 556 = 313 = f_1^{(E_1)}$ $\quad 3/16 \times 556 = 104 = f_3^{(E_3)}$

$3/16 \times 556 = 104 = f_2^{(E_2)}$ $\quad 1/16 \times 556 = 35 = f_4^{(E_4)}$

Now, Consider; $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$

$= \frac{(315-313)^2}{313} + \frac{(101-104)^2}{104} + \frac{(108-104)^2}{104} + \frac{(32-35)^2}{35}$

$= \frac{4}{313} + \frac{9}{104} + \frac{16}{104} + \frac{9}{35} = 0.51$

∴ $\boxed{\chi^2 = 0.51}$

WKT; d.f ⇒ $\nu = n - k$

$= 4 - 1$

ν d.f $= 3$.

for; $\nu = 3$, we have.. $\chi^2_{0.05} = 7.815$ (tabulated value)

Since, Calculated value $\boxed{0.51 < 7.815}$ (tabulated value), there is a very high degree of agreement b/w theory & experiment.

(3) A set of 5 similar coins is tossed 320 times and the result is;

No. of heads	0	1	2	3	4	5
Frequency	6	27	72	112	71	32

Given; $\chi^2_{0.05} = 11.07$

Test the hypothesis that the data follow a binomial distribution.

Soln :- Null hypothesis, Ho: Data follows Binomial distribution.

d.f : $\nu = 6-1 = 5$

For, $\nu=5$, we have : $\chi^2_{0.05} = 11.07$ (Tabulated value)

Since, the distribution is binomial.

$\therefore$ Probability of getting a head $= \boxed{p=\frac{1}{2}}$ WKT; $p+q=1$. $\boxed{q=\frac{1}{2}}$

ie, the probability of getting tail is ; $\boxed{q=\frac{1}{2}}$

$\therefore$ Total frequency from the data $= 320$ // $[6+27+72+112+71+32]$

Theoretical frequency ~~from the~~ of getting $\{0,1,2,3,4,5\}$ heads are the successive terms of the binomial expansion

: $320(p+q)^n$: $320(p+q)^5$, $\boxed{n=5}$

$$\therefore (320)(p+q)^5 = 320\left[p^5 + {}^5C_1 p^4 q + {}^5C_2 p^3 q^2 + {}^5C_3 p^2 q^3 + {}^5C_4 p q^4 + {}^5C_5 q^5\right]$$

$$= 320\left[p^5 + 5p^4 q + 10p^3 q^2 + 10p^2 q^3 + 5pq^4 + q^5\right]$$

$$= 320\left[\frac{1}{32} + \frac{5}{32} + \frac{10}{32} + \frac{10}{32} + \frac{5}{32} + \frac{1}{32}\right]$$

$$= 10 + 50 + 100 + 100 + 50 + 10.$$

Thus, the theoretical frequencies are : $\{10, 50, 100, 100, 50, 10\}$

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

$$= \frac{(6-10)^2}{10} + \frac{(27-50)^2}{50} + \frac{(72-100)^2}{100} + \frac{(112-100)^2}{100} + \frac{(71-50)^2}{50} + \frac{(32-10)^2}{10}$$

$$= \frac{16}{10} + \frac{529}{50} + \frac{784}{100} + \frac{144}{100} + \frac{441}{50} + \frac{484}{10}$$

$$= \frac{160 + 1058 + 784 + 144 + 882 + 4840}{100} = \frac{7868}{100}.$$

$\therefore \boxed{\chi^2 = 78.68}$ & df $v = 6-1 = 5$

Since, the calculated value of $\chi^2 = 78.68 > \chi^2_{0.05} = 11.07$

$\therefore$ The hypothesis that the data follow the binomial distribution law is rejected.

(4) A die is thrown 60 times and the frequency distribution for the number appearing on face x is given by the following table;

x	1	2	3	4	5	6
observed frequency	15	6	4	7	11	17

Test the hypothesis that the die is unbiased given that;

$\chi^2_{0.05}(5) = 11.07$ & $\chi^2_{0.01}(5) = 15.09$

Soln :- Null hypothesis : H_0 : Die is unbiased.

Expected frequency for each face $= \frac{15+6+4+7+11+7}{6} = \frac{60}{6} = 10$

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

$$= \frac{(15-10)^2}{10} + \frac{(6-10)^2}{10} + \frac{(4-10)^2}{10} + \frac{(7-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(17-10)^2}{10}$$

$$\chi^2 = \frac{1}{10}[25+16+36+9+1+49] = \frac{136}{10} = \boxed{13.6 = \chi^2}$$

Given; $\chi^2_{0.05}(5) = 11.07$

So, Calculated value : 13.6 > the tabulated value 11.07

$\therefore$ Ho is rejected at 5% level of significance.

$\therefore$ The calculated value ; $\chi^2 = 13.6$ is less than tabulated value at 1% level of significance

ie; $\chi^2 = 13.6 < \chi^2_{0.01} = 15.09$

$\therefore$ It is not significant & Ho is accepted..

ie, The die is unbiased

(4) A Coins are tossed 100 times and the following results were obtained, Fit a binomial distribution for the data & test the goodness of fit ($\chi^2_{0.05} = 9.49$ for 4.d.f).

No of heads	0	1	2	3	4
frequency	5	29	36	25	5

Theoretical values of freq are Given : {7, 26, 37, 24, 6}

Soln: Theoretical values of freq are :- {7, 26, 37, 24, 6}.

We have the following table ;

Oi	5	29	36	25	5
Ei	7	26	37	24	6

We have ; $\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = \frac{4}{7} + \frac{9}{26} + \frac{1}{37} + \frac{1}{24} + \frac{1}{6} = 1.15$.

$\therefore \boxed{\chi^2 = 1.15} < \chi^2_{0.05} = 9.49$

Thus, the hypothesis that the fitness is good can be "accepted"

⑤ Fit a poisson distribution for the following data & test the goodness of fit given that : $\chi^2_{0.05} = 7.815$ for 3 d.f.

x	0	1	2	3	4
f	122	60	15	2	1

Given theoretical frequencies are :- {121, 61, 15, 3, 0}.

Soln :- Given theoretical frequencies ; {121, 61, 15, 3, 0}

We have the following table :-

O_i	122	60	15	2+1=3
E_i	121	61	15	3+0=3

$\therefore \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{1}{121} + \frac{1}{61} + 0 + 0 = \boxed{0.025 = \chi^2}$

$\therefore \chi^2 = 0.025 < \chi^2_{0.05} = 7.815$. The fitness is considered good.

Then, the hypothesis that fitness is good can be accepted.

⑥ The no. of accidents per day (x) as recorded in textile industry over period of 400 days is given below, Test the goodness of fit in respect of "Poisson's distribn" of fit to given data ; ($\chi^2_{0.05} = 9.49$ for 4 d.f).

x	0	1	2	3	4	5
f	173	168	37	18	3	1

Given theoretical freq. are :- {183, 143, 56, 15, 3, 0}.

Soln :- Given theoretical freq : {183, 143, 56, 15, 3, 0}.

$\therefore O_i$	173	168	37	18	3+1=4
E_i	183	143	56	15	3+0=3

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

$= \frac{100}{183} + \frac{625}{143} + \frac{361}{56} + \frac{9}{15} + \frac{1}{3} = 12.297 \approx 12.3$

$\therefore \chi^2 = 12.3 > \chi^2_{0.05} = 9.49$, the fitness is not good

Then, the hypothesis that the fitness is good is rejected

— * —

STOCHASTIC PROCESS :-

Stochastic process :- Stochastic process consists of a sequence of experiments in which each experiment has a finite number of outcomes with given probabilities.

The values assumed by the random variables $x(t)$ are called states. The set of possible values are called State space of process "l".

If the state space is discrete, the stochastic process is known as a Chain.

Markov :- Markov process is stochastic process whose entire past history is summarized in its current (present) state.

ie, The future is independent of its past.

Markov Chain :-

It is a markov process in which the state space I is discrete (finite | Countably infinite).

(or) Markov chain is a finite stochastic process consisting of sequence of trials whose outcomes say; $x_1, x_2 \dots$ satisfy 2 Conditions :-

1. Each out-come belongs to state space; $I = \{a_1, a_2 \dots a_m\}$ which is finite set of out comes.
2. The outcomes of any trial depends at most upon outcome of the immediately proceeding trial & not upon any other previous outcomes.

$$= P(X_n = i_n \mid X_{n-1} = i_{n-1})$$

where, P is known as transition probability.

* Transition matrix :-

P is a square matrix of transition probability P.

$$\text{ie, } P = \begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{array} \begin{array}{c} \begin{array}{cccc} a_1 & a_2 & \cdots & a_n \end{array} \\ \begin{bmatrix} P_{11} & P_{21} & \cdots & P_{1m} \\ P_{21} & P_{22} & \cdots & P_{2m} \\ \vdots & \vdots & & \vdots \\ P_{n1} & P_{n2} & & P_{nm} \end{bmatrix} \end{array}$$

The i^{th} row of P namely : $\{P_{i1}\ P_{i2} \ldots P_{in}\}$ represents probabilities of that system will change from a_i to $\{a_1, a_2 \ldots a_m\}$.

* Probability Vector :-

If a vector is $V = \{v_1, v_2 \ldots v_n)$ $((v_i \geq 0)$ for every $i)$

and $\boxed{\sum_{i=1}^{n} v_i = 1}$

* Stochastic matrix :- A square matrix P is called a stochastic matrix if all the entries of P are non-negative & the sum of the entries of any row is one.

(or) A square matrix P is called a stochastic matrix with each row being a probability vector is stochastic matrix

* Fixed Vector :- A vector v is said to be a fixed vector or fixed point of a matrix A if $vA = v$ & $v \neq 0$

Regular stochastic matrix :- A stochastic matrix P is said to be Regular, if all the entries of some of P^m are positive.

Higher Transition probabilities :-

1) One step - transition probabilities :-

The probability that a markov chain will move from state a_i to the state a_j in one step. is one step transition probability.

ie ; $P_0 = P(X_n = j \mid X_{n-1} = i)$.

2) n-step transition probabilities :-

The probability that marcov chain will move from one state i to state j in exactly n steps and its denoted by;

$$P_{ij}^{(n)} = P_{ij}(n) = P(X_{m+n} = j / X_m = i).$$

Problems :-

① which vectors are probability vectors :-

ⓐ $\left(\frac{1}{4}, \frac{3}{4}, -\frac{1}{4}, \frac{1}{2}\right)$

Soln:- It is not a probability vector,
Since ; It has negative entry

ⓑ $\left(\frac{5}{2}, 0, \frac{3}{8}, \frac{1}{6}\right)$

Soln:- It is not a probability vector because, their sum is not equal to 1.

ⓒ $\left(\frac{1}{12}, \frac{3}{2}, \frac{1}{6}, 0, \frac{1}{4}\right)$.

Soln:- It is a probability vector,
because all entries are non-negative & their sum is equal to 1.

② Which matrices are <u>Stochastic</u>.

ⓐ $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ It is a stochastic matrix, since sum of each row is equal to 1 & all entries are non negative.

ⓑ. $\begin{bmatrix} 0 & 1 \\ 1/2 & -1/3 \end{bmatrix}$ It is not a stochastic matrix, because of negative entry.

③ Which of the stochastic matrices are <u>regular</u> :-

ⓐ $A = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}$ <u>Soln</u>:- A is not regular, because 1 appears on the principal main diagonal.

ⓑ $B = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$ <u>Soln</u>:- $B^2 = B.B = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 1/2 \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 1/2 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 7/16 & 7/16 & 1/8 \end{bmatrix}$

Since the entries ; b_{13} & b_{23} are <u>zero's</u>

$\therefore$ B is not regular.

⑧ = [

④ Find the unique fixed probability vector of $A = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/3 & 2/3 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

<u>Soln</u>:- Let $t = (x, y, z)$ be the fixed probability vector.

By definition, WKT ; $x + y + z = 1$

So ; $t = (x, y, 1 - x - y)$

t is said to be fixed vector if ; <u>$tA = t$</u>.

$$\Rightarrow (x, y, 1-x-y)\begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/3 & 2/3 & 0 \\ 0 & 1 & 0 \end{bmatrix} = (x \quad y \quad 1-x-y)$$

On Multiplication; we get;

$$1/3\, y = x \Rightarrow y = 3x \quad \text{—}①$$

$$1/2\, x + 2/3\, y + 1 - x - y = y.$$

$$1/2\, x = 1 - x - 3x.$$

Using, ① $\Rightarrow$ $1/2\, x = 1 - x - 3x$

$$4x + 1/2\, x = 1 \quad (\text{or}) \quad \boxed{x = 2/9}$$

$$\Rightarrow y = 3x = \boxed{2/3 = y}$$

$$z = 1 - x - y$$

$$z = 1 - 2/9 - 2/3, \quad \boxed{z = 1/9}$$

$\therefore$ Required fixed probability vector;

$$t = (x, y, z) = (2/9, 2/3, 1/9)$$