Module Probability Distribution

* **Sample space :**
The set of all possible outcomes of an experiment is called sample space.
It is denoted by 'S'

Note that the element of set S is called sample points.

<u>for example :</u> In a throw of two coins,
the sample space $S = \{(H,H), (H,T), (T,H), (T,T)\}$

* **Event :** A subset of sample space is called an event.

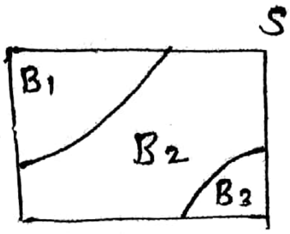
<u>for example :</u> ① If $S = \{(H,H), (H,T), (T,H), (T,T)\}$

then $A_1 = \{(H,H), (T,H)\}$
$A_2 = \{(H,T), (T,H), (T,T)\}$ are the events

② In a throw of a die,

The sample space $S = \{1, 2, 3, 4, 5, 6\}$

and $B_1 = \{1, 2\}$
$B_2 = \{3, 4, 5\}$
$B_3 = \{6\}$ are events



* probability:

If S is the sample space with n points which are mutually exclusive and A is the event (subset of S) with m points then the ratio $\frac{m}{n}$ is called probability of A and is denoted by 'P(A)'

that is $P(A) = \frac{m}{n} = \frac{\text{number of points in } A}{\text{number of points in } S}$

Important Note: i) $P(A) \geq 0$, for any A

ii) $P(S) = 1$

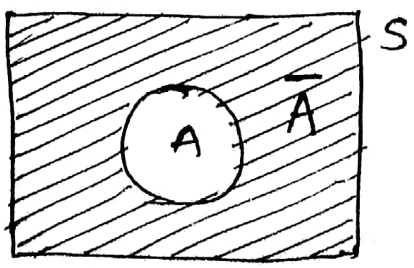
iii) $P(A \cup B) = P(A) + P(B)$,

for any exclusive events A, B

* Complement of the event:

If A be the event of a sample space S then the complement of A is denoted by $\bar{A}$ and is given by

$\bar{A} = S - A$



Note: $A \cup \bar{A} = S$

* Random Variables:

The variable which is associated with the outcomes of the sample space of the random experiment is called random variable

for example:

In a throw of three coins

Sample space 'S' = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

$\therefore n(S) = 8$

Random experiment : Tossing of three coins

Random variable (r.v) X = number of heads

$X = 0, 1, 2, 3$

Probability distribution

X	0	1	2	3
$P(X=x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

* Probability Distribution:-

let X be the random variable and x_i be the values of X then the set of pairs $\{x_i, P(x_i)\}$ is called as probability distribution, where $P(x_i)$ is probability of x_i

* Poisson Distribution :-

let n be the number of trials

p be the probability of success in each trials

and np be the avarrage success say 'm'

i.e. $\boxed{m = np}$

Then a random variable X is said to follow possion distribution if the probability of x is given by

$$\boxed{P(X = x) = \frac{e^{-m} m^x}{x!}} \quad , \quad x = 0, 1, 2, \ldots$$

<u>Note that</u> : * n is infinitely large i.e $n \to \infty$

* p is always constant and infinitely small i.e $p \to 0$

* m is finite and $\boxed{m = np}$

* Expected value :-

The sum of product of values and their probability is called as Expected value & it is denoted as E(X)

i.e. $E(X) = P_1 x_1 + P_2 x_2 + P_3 x_3 + \ldots$

<u>Note that</u> : $\boxed{E(X) = m}$

* Note that

— If X and Y are two variates.

then 1) $E(X+Y) = E(X)+E(Y)$

2) $E(X-Y) = E(X) - E(Y)$

— If X and Y are two independent variates

then $E(XY) = E(X) \cdot E(Y)$

— If X is a variates and 'a', 'b' are any constants

then $E(aX+b) = a\,E(X) + b$

* Moment Generating function :. (m.g.f)

The moment genrating function (m.g.f) of a random variate X is denoted by $M_0(t)$ and is defined by

$$M_0(t) = E(e^{tx}) \quad \text{(about origin)}$$

Ex. ① Derive the moments of the posson's Distributions:

Solution: here we find below the first two moments about the origin

① $\mu_1' = E(X) = \sum p_i x_i$

$$= \sum_{x=0} \frac{e^{-m} m^x}{x!} \cdot x = \sum_{x=1} \frac{e^{-m} m^x}{(x-1)!}$$

$$= \sum_{x=1} \frac{e^{-m} \cdot m \cdot m^{x-1}}{(x-1)!} = e^{-m} \cdot m \sum_{x=1} \frac{m^{x-1}}{(x-1)!}$$

$$= m e^{-m} \left[1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + \cdots \right]$$

$$= m e^{-m} \cdot e^{m}$$

$$= m$$

Hence, $\boxed{\text{mean} = m}$

② $$\mu_2' = E(x^2) = \sum P_i x_i^2 = \sum_{x=0} \frac{e^{-m} \cdot m^x}{x!} \cdot x^2$$

$$= \sum_{x=0} \frac{e^{-m} m^x}{x!} (x + x^2 - x)$$

$$= \sum_{x=0} \frac{e^{-m} \cdot m^x}{x!} [x + x(x-1)]$$

$$= \sum_{x=0} \frac{e^{-m} \cdot m^x \cdot x}{x!} + \sum_{x=0} \frac{e^{-m} \cdot m^x}{x!} x(x-1)$$

$$= e^{-m} \cdot m \sum_{x=1} \frac{m^{x-1}}{(x-1)!} + e^{-m} m^2 \sum_{x=2} \frac{m^{x-2}}{(x-2)!}$$

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

$$= e^{-m} \cdot m \cdot e^{m} + e^{-m} \cdot m^2 \cdot e^{m}$$

$$= m + m^2$$

$$\therefore \mu_2 = \mu_2' - \mu_1'^2 = m + m^2 - m^2 = m$$

$\therefore$ $\boxed{\text{variance} = m}$

Therefore, the mean and varience of the posson's distribution are both equal to 'm'

Ex (2) obtain the moment Generating function of poisson's distribution.

Solution: Note that the moment generating function about origin is

$$M_0(t) = E(e^{tx})$$

$$= \sum p(x)\, e^{tx}$$

$$= \sum_{x=0} \frac{e^{-m} \cdot m^x}{x!} \cdot e^{tx}$$

$$= e^{-m} \sum_{x=0} \frac{m^x \cdot (e^t)^x}{x!}$$

$$= e^{-m} \cdot \sum_{x=0} \frac{(me^t)^x}{x!}$$

$$= e^{-m}\left[1 + me^t + \frac{(me^t)^2}{2!} + \frac{(me^t)^3}{3!} + \cdots\right]$$

$$= e^{-m} \cdot e^{me^t}$$

$$= e^{-m+me^t}$$

$$= e^{m(e^t-1)}$$

$$\therefore \boxed{M_0(t) = e^{m(e^t-1)}}$$

Ex ③ If the variance of a poisson distribution is 2, find the probabilities of $r = 1, 2, 3, 4$ from the recurrence relation of poisson distribution.

$\longrightarrow$ Note that $P(X=x) = \dfrac{e^{-m} \cdot m^x}{x!}$

given that Varience $= m = 2$

$\therefore$ for $x=0$, $P(X=0) = \dfrac{e^{-2} \cdot 2^0}{0!} = e^{-2}$

i.e. $P(0) = e^{-2}$

Now, the recurrence relation is $P(x+1) = \dfrac{m}{x+1} P(x)$

$\therefore$ put $x=0$, $P(0+1) = P(1) = \dfrac{m}{0+1} P(0) = \dfrac{2}{1} e^{-2} = 2e^{-2}$

put $x=1$, $P(1+1) = P(2) = \dfrac{2}{1+1} P(1) = \dfrac{2}{2} 2e^{-2} = 2e^{-2}$

put $x=2$, $P(2+1) = P(3) = \dfrac{2}{2+1} P(2) = \dfrac{2}{3}(2e^{-2}) = \dfrac{4}{3} e^{-2}$

put $x=3$, $P(3+1) = P(4) = \dfrac{2}{3+1} P(3) = \dfrac{2}{4} \dfrac{4}{3} e^{-2} = \dfrac{2}{3} e^{-2}$

Therefore, $P(1) = 2e^{-2}$, $P(2) = 2e^{-2}$, $P(3) = \dfrac{4}{3} e^{-2}$, $P(4) = \dfrac{2}{3} e^{-2}$

Ex ④ If a random variable X follows poisson distribution such that $P(X=1) = 2P(X=2)$, find the mean and the variance of the distribution. Also find $P(X=3)$

$\longrightarrow$ Note that $P(X=x) = \dfrac{e^{-m} \cdot m^x}{x!}$

given that $P(X=1) = 2P(X=2)$

$\Rightarrow \dfrac{e^{-m} \cdot m^1}{1!} = 2 \dfrac{e^{-m} m^2}{2!}$

$$\Rightarrow \quad m e^{-m} = m^2 e^{-m}$$

$$\Rightarrow \quad m = m^2$$

$$\Rightarrow \quad m = 1$$

$\therefore$ The mean = variance = $m = 1$

Now $P(X=3) = \dfrac{e^{-1} \cdot (1)^3}{3!} = \dfrac{e^{-1}}{1\times2\times3} = \underline{\underline{0.0613}}$

Ex. ⑤ A hospital switch board receives an avarage of 4 emergency calls in a 10 minutes interval. What is the probability that

i) there are atmost 2 emergency calls

ii) there are exactly 3 emergency call in an interval of 10 minutes

iii) more than 2 emergency calls

$\longrightarrow$ Note that $P(X=x) = \dfrac{e^{-m} \cdot m^x}{x!}$

here, $m = 4$

i) $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$

$$= \frac{e^{-4} \cdot 4^0}{0!} + \frac{e^{-4} \cdot 4^1}{1!} + \frac{e^{-4} \cdot 4^2}{2!}$$

$$= e^{-4}(1 + 4 + 8) = \underline{\underline{0.238}}$$

ii) $P(X=3) = \dfrac{e^{-4}\, 4^3}{3!} = \underline{\underline{0.195}}$

iii) $P(X>2) = P(X=3) + P(X=4) + P(X=5) + \cdots$

$$= 1 - [P(X=0) + P(X=1) + P(X=2)]$$

$$= 1 - \left[\frac{e^{-4} \cdot 4^0}{0!} + \frac{e^{-4} \cdot 4^1}{1!} + \frac{e^{-4} \cdot 4^2}{2!}\right]$$

$$= 1 - 0.238$$

$$= 0.762$$

* Normal Distribution :-

let X be the random variable

m be the mean and

σ be the standard deviation

then the contineous random variable X is said to follow normal distribution if its probability density function is

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

Important Note : * If X is normal variate with parameter m, σ then $\boxed{Z = \frac{X-m}{\sigma}}$ is called Standard Normal variates

* Mean and varience of Normal Distribution :

$\boxed{\text{mean} = m}$ $\boxed{\text{Variance} = Var.(X) = \sigma^2}$

* $\boxed{\text{Mean} = \text{Median} = \text{Mode} = m}$

* Moment Genrating function of Normal distribution

$$\boxed{M_0(t) = e^{\left(mt + \frac{t^2\sigma^2}{2}\right)}}$$

Note that In standard normal variates,

mean $= m = 0$ and $\sigma = 1$

Hence, generating function of standard normal variates is $\boxed{M_0(t) = e^{\frac{t^2}{2}}}$

* Area property :

let X = Random variates

Z = standard Normal variates (S.N.V)

then the area under the normal curve of X between $X = m$ and $X = x_1$ is equal to area under the standard normal curve between $Z = 0$ to $Z = z_1$

$\Rightarrow$ clearly if $X = m$ then $Z = \frac{X-m}{\sigma} = \frac{m-m}{\sigma} = 0$

and if $X = x_1$ then $Z = \frac{x_1 - m}{\sigma} = z_1$ (say)

therefore,

$$\boxed{P(m \leq X \leq x_1) = P(0 \leq Z \leq z_1) = \frac{1}{\sqrt{2\pi}} \int_0^{z_1} e^{-\frac{1}{2}x^2} dx}$$

Note that:

* $P(-z_2 \leq Z \leq z_1) = P(0 \leq Z \leq z_2) + P(0 \leq Z \leq z_1)$

Ex ① If mean of normal variate is 2.5 and standard deviation is 3.5 then find the probability that $2 \le X \le 4.5$

→ given that $m = 2.5$

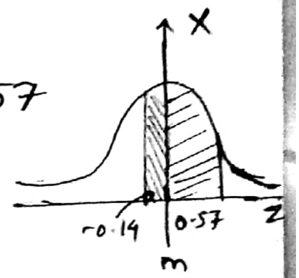
$\sigma = 3.5$ and

$2 \le X \le 4.5$

Note that Standard Normal variates $Z = \frac{X - m}{\sigma}$

$\therefore$ for $X = 2$, $Z = \frac{2 - 2.5}{3.5} = -0.14$

for $X = 4.5$, $Z = \frac{4.5 - 2.5}{3.5} = 0.57$



Therefore, the probability is

$P(2 \le X \le 4.5) = P(-0.14 \le Z \le 0.57)$

$= P(0 \le Z \le 0.14) + P(0 \le Z \le 0.57)$

$= \frac{1}{\sqrt{2\pi}} \int_0^{0.14} e^{-\frac{1}{2}x^2} dx + \frac{1}{\sqrt{2\pi}} \int_0^{0.57} e^{-\frac{1}{2}x^2} dx$

$= 0.0557 + 0.2157$

$= 0.2714$

Ex ② If X is a normal variates with mean 10 and the standard deviation 4, find

i) $P(|X-14| < 1)$ ii) $P(5 \le X \le 18)$ iii) $P(X \le 12)$

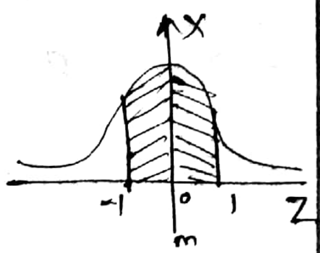
Solution: Note that the standard Normal variates is $Z = \frac{X - m}{\sigma}$

Scanned by CamScanner

here, $m = 10$ and $\sigma = 4$

i) To find $P(|X-14|<1)$

if $X = 14$, $Z = \frac{14-10}{4} = 1$



$\therefore P(|X-14|<1) = P(|Z| \leq 1)$

$= P(-1 \leq Z < 1)$

$= P(0 \leq Z \leq 1) + P(0 \leq Z \leq 1)$

$= 2P(0 \leq Z \leq 1)$

$= 2 \frac{1}{\sqrt{2\pi}} \int_0^1 e^{-\frac{1}{2}x^2} dx$

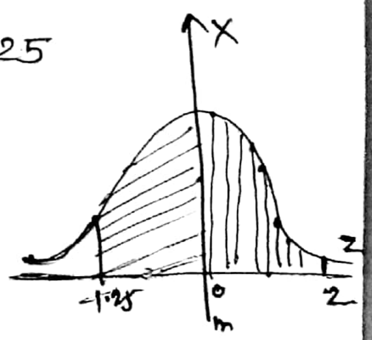
$= 2 (0.3413)$

$= 0.6826$

ii) To find $P(5 \leq X < 18)$

if $X = 5$, $Z = \frac{X-m}{\sigma} = \frac{5-10}{4} = -1.25$

if $X = 18$, $Z = \frac{X-m}{\sigma} = \frac{18-10}{4} = 2$



$\therefore P(5 < X < 18) = P(-1.25 \leq Z \leq 2)$

$= P(0 \leq Z \leq 1.25) + P(0 \leq Z \leq 2)$

$= \frac{1}{\sqrt{2\pi}} \int_0^{1.25} e^{-\frac{1}{2}x^2} dx + \frac{1}{\sqrt{2\pi}} \int_0^{2} e^{-\frac{1}{2}x^2} dx$

$= 0.3944 + 0.4772$

$= 0.8716$

iii) To find $P(X \leq 12)$

if $X = 12$, $Z = \frac{X-m}{\sigma} = \frac{12-10}{4} = 0.5$

$\therefore\ P(X \le 12) = P(Z \le 0.5)$

$= P(-\infty \le Z \le 0.5)$

$= P(0 \le Z \le \infty) + P(0 \le Z \le 0.5)$

$= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{1}{2}x^2}\, dx + \frac{1}{\sqrt{2\pi}} \int_0^{0.5} e^{-\frac{1}{2}x^2}\, dx$

$= 0.5 + 0.1915$

$= 0.6915$

Ex. ③ Monthly salary X in a big organization is normally distributed with mean ₹ 3000 and standard deviation of ₹ 250. what shoud be the minimum salary of a worker in this organisation, so that the probability that he belongs to top 5% workers ?

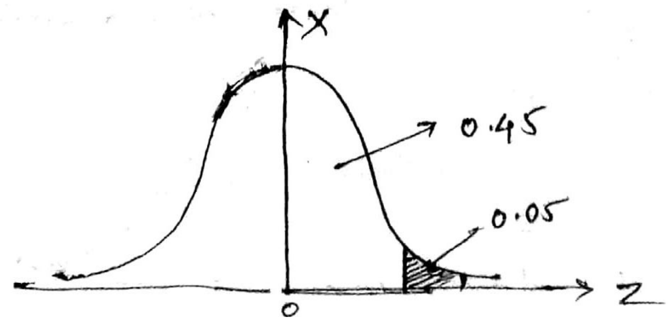
Solution: Note that the standard Normal variates is

$$Z = \frac{X - m}{\sigma}$$

here, mean $= m = 3000$

and standard deviation $= \sigma = 250$

To find z_1 such that $P(Z = z_1) = \frac{5}{100} = 0.05$



Scanned by CamScanner

$\therefore \quad 0.5 - 0.05 = 0.45$

The corresponding to 0.45 the entry in the area table is 1.64

$\therefore \quad z_1 = 1.64$

Now $\quad Z = z_1 = \dfrac{X - m}{\sigma}$

$\Rightarrow \quad 1.64 = \dfrac{X - 3000}{250}$

$\Rightarrow \quad X = 300 + (250 \times 1.64)$

$\Rightarrow \quad \underline{X = ₹\ 3410}$

∴ The minimum salary of a worker in this organization is ₹ 3410

Ex. 4 The marks obtained by 1000 students in an examination are found to be normally distributed with 70 and stadard deviation 5 Estimate the number of student whose marks will be i) between 60 and 75

ii) more than 75

solution: Note that the standard Normal variates

$$Z = \frac{X - m}{\sigma}$$

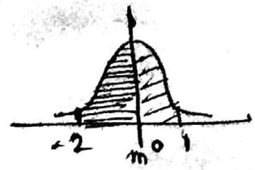
here, mean = m = 70

and standard deviation = σ = 5

i) if $x = 60$ then $z = \frac{x-m}{\sigma} = \frac{60-70}{5} = -2$

if $x = 75$ then $z = \frac{x-m}{\sigma} = \frac{75-70}{5} = 1$

$P(60 \le X \le 75) = P(-2 \le Z \le 1)$



$= P(0 \le Z \le 2) + P(0 \le Z \le 1)$

$= \frac{1}{\sqrt{2\pi}} \int_0^2 e^{-\frac{1}{2}x^2} dx + \frac{1}{\sqrt{2\pi}} \int_0^1 e^{-\frac{1}{2}x^2} dx$

$= 0.4772 + 0.3413$

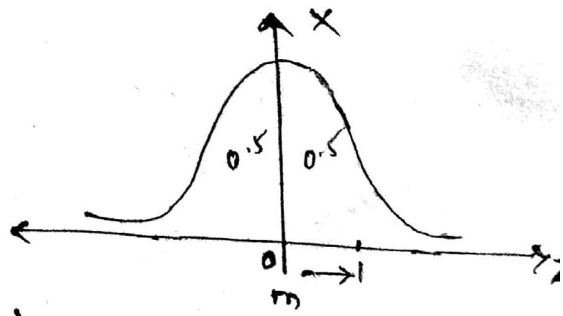
$= 0.8185$

$\therefore$ The number of students getting marks between 60 to 75 $= N \times P$

$= 1000 \times 0.8185$

$= 818$

ii) $P(X \ge 75) = P(Z \ge 1)$



$= 0.5 - P(0 \le Z \le 1)$

$= 0.5 - \frac{1}{\sqrt{2\pi}} \int_0^1 e^{-\frac{1}{2}x^2} dx$

$= 0.5 - 0.3413 = 0.1587$

$\therefore$ the number of students getting more than 75 marks $= NP = 1000 \times 0.1587 = \underline{159}$

Ex. 5: Marks obtained by students in an examination follow normal distribution. If 30 % of students got below 35 marks and 10 % got above 60 marks, find the mean and standard deviation.

Solution: Note that mean $= m$ and

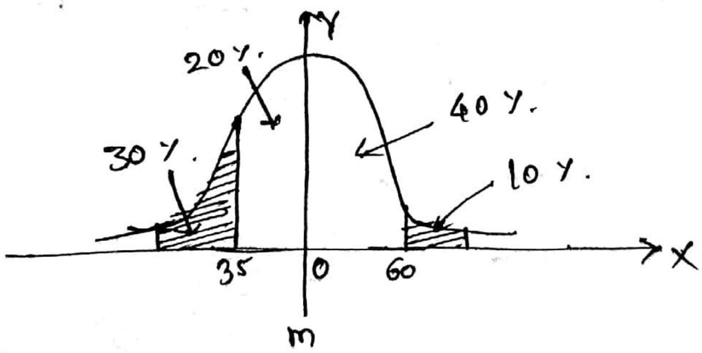
standard deviation $= \sigma$

Since, 30 % students are below 35

$\Rightarrow$ 20 % students are between 35 and m

Since, 10 % students are above 60

$\Rightarrow$ 40 % students are between m and 60



$\therefore$ from the given data (table) (will be provided in Exam)

0.2 area corresponds to $Z = 0.525$

and 0.4 area corresponds to $Z = 1.283$

(0.2 area corresponds) $Z = \dfrac{35 - m}{\sigma}$

$\Rightarrow \quad 0.525 = \dfrac{35 - m}{\sigma}$

$\Rightarrow \quad m + 0.525\,\sigma = 35 \quad \text{—— (1)}$

and (0.4 Area cump) $z = \frac{x-m}{\sigma} = \frac{60-m}{\sigma}$

$\Rightarrow \quad 1.283 = \frac{60-m}{\sigma}$

$\Rightarrow \quad m + 1.283\sigma = 60$ —— (2)

solving (1) and (2) we get

$m = 13.83$

$\sigma = 42.26$

(use calci fo system of equation)

$\therefore$ mean $= 13.83$

and standard deviation $= 42.26$.

Homework :

(1) In an Intelligenee test administered to 1000 students, the Avarage was 42 and standard deviation was 24 find the number of students

i> exceeding the score 50

ii> between 30 and 54

(2) In a distribution exactly normal 7% of items are under 35 and 89% are under 63 what is the mean and standard deviation.

* sampling distribution :.

* population: The group of individuals under study is called population or universe
It may be finite or infinite

* Sampling : A part selected from the population is called a sample and the process of selection of sample is called sampling.

<u>Note that</u> :

* mean of population is denoted by 'μ'
* standard deviation of population is denoted by 'σ'
* Mean of Sample is denoted by '$\bar{x}$'
* standard deviation of sample is denoted by 's'
* Size of population is denoted by 'N'
* size of Sample is denoted by 'n'

* Testing of hypothesis :

on the basis of sample information, we make certain decisions about population. In taking such decision we make assumptions these assumption are known as statical hypothesis.

* Null hypothesis: (H_0)

null hypothesis is no difference, thus we shall persume that there is no significant difference between the observed value and the expected value.

* Alternative hypothesis: (H_a)

It specifies a range of values rather than one value.

* levels of significance:- (α)

It is expressed in the percentage as 5 % level of significance or 1 % level of significance.

* Critical region:

The levels marked by probabilities 0.05 or 0.01 which decide the significance of an event are called level of significance and are expressed in percentage as 5 % level of significance or 1 % level of significance. The corresponding region are called critical region.

* Two Tailed and one Tailed test :

The probability distribution of a sample statistic is normal distribution.

The z - curve is symmetrical as we know and the parts of the curve at the two ends are called the two tails of the curve.

If the rejection area lies on two sides i.e on the two tailes the test is called the two tailed test.

If on the other hand the rejection area lies on one side only the test is called one tailed test

Note that :

1) $\mu > \mu_0$ or $\mu < \mu_0$ is two tailed test

2) $\mu > \mu_0$ is Right Tailed Test (one tailed test)

3) $\mu < \mu_0$ is left tailed test (one tailed test

	level of significance	
	1 %	5 %
Two tailed Test	$Z_\alpha = 2.576$	1.96
one tailed test	$Z_\alpha = 2.326$	1.64

Ex 1. A random sample of 50 items gives the mean 6.2 and variance 10.24. Can it be regarded as drawn from a normal population with mean 5.4 at 5% level of significance?

solution: Given: $n = 50$

$\bar{x} = 6.2$

$\mu = 5.4$

$s = \sqrt{10.24}$

i) Null Hypothesis (H_0): $\mu = 5.4$

Alternative Hypothesis H_a: $\mu \neq 5.4$

ii) Test statistic:

Since the population S.D. is unknown

But sample S.D. 's' is known

$$\therefore \quad Z = \left| \frac{\bar{x} - \mu}{s/\sqrt{n}} \right|$$

$$= \left| \frac{6.2 - 5.4}{\sqrt{10.24}/\sqrt{50}} \right| = \left| \frac{0.8}{3.2/7.07} \right|$$

$$= 1.77$$

iii) Level of significance: $\alpha = 0.05$ $\left(5\% = \frac{5}{100}\right)$

iv) Critical value: The value of Z_α at 5% level of significance from the table $= 1.96$

v) Decision: since; $|z| = 1.77$ i.e calculated value which is less than the critical value $z_\alpha = 1.96$

hence, the null hypothesis is accepted.

$\therefore$ The sample is drawn from the population with mean 5.4.

Ex. 2. A random sample of 400 members is found to have a mean of 4.45 cms can it be reasonably regarded as a sample from a large population whose mean is 5 cms and variance is 4 cms.

Solution: Given:

$n = 400$

$\bar{x} = 4.45$

$\mu = 5$

$\sigma = \sqrt{4} = 2$

i) Null Hypothesis $H_0 : \mu = 5$

Alternative hypothesis $H_a : \mu \neq 5$

ii) Test statistic :

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

$\therefore \quad Z = \left| \frac{4.45 - 5}{2/\sqrt{400}} \right| = \left| \frac{0.55}{2/20} \right|$

$\Rightarrow \quad Z = 5.5$

iii) Level of significance : $\alpha = 0.05$ ($\because$ large sample)

iv) Critical value :

The value of Z_α at 5% level of significance from the table $= 1.96$

v) Decision : Since the computed value of $Z = 5.5$ is greater that the critical value $Z_\alpha = 1.96$

$\therefore$ The null hypothesis is rejected and the alternative hypothesis is accepted.

$\therefore$ The sample is not drawn from the above population.

* Distribution of the difference between means

– procedure to test the hypothesis :

Step 1. Given: sizes of two samples n_1, n_2 with mean $\bar{X}_1$, $\bar{X}_2$ respectively and means of populations μ_1, μ_2 and standard deviation of population σ_1, σ_2

Step 2. calculate $\bar{X}_1 - \bar{X}_2$

Step 3. find standard error (S. E)

$$S = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Step 4. find $Z = \left| \frac{\bar{X}_1 - \bar{X}_2}{S} \right|$

& take the decision.

Example: 1. The mean of two sample of sizes 1000 and 2000 respectively are 67.50 and 68.0 inches. Can the samples be regarded as drawn from the same population of standard deviation.

Solution: Given: $n_1 = 1000$, $n_2 = 2000$

$\bar{X}_1 = 67.50$, $\bar{X}_2 = 68.0$

$\sigma_1 = 2.5$ $\sigma_2 = 2.5$

i) Null Hypothesis H_0 : $\mu_1 = \mu_2$

Alternative hypothesis H_a : $\mu_1 \neq \mu_2$

ii) Calculation of statistic :

$$\overline{X}_1 - \overline{X}_2 = 67.5 - 68.0 = -0.5$$

Now, standard error (S.E.) is

$$S = \sqrt{\frac{\sigma_1^2}{n_1^2} + \frac{\sigma_2^2}{n_2^2}} = \sqrt{\frac{(2.5)^2}{1000} + \frac{(2.5)^2}{2000}}$$

$$= (2.5)\sqrt{\frac{1}{1000} + \frac{1}{2000}}$$

$$= 0.097$$

$$\therefore \quad Z = \left|\frac{\overline{X}_1 - \overline{X}_2}{S}\right| = \left|\frac{-0.5}{0.097}\right| = |-5.15| = 5.15$$

iii) Level of significance : $\alpha = 0.27\ \%$ (given)

iv) critical value :

The value of Z_α at $0.27\ \%$ level of significance from the table is 3

v) Decision :

Note that the computed value of $|Z| = 5.15$ is greater than the critical value $Z_\alpha = 3$

$\therefore$ The Hypothesis is Rejected.

$\therefore$ The sample cannot be regarded as drawn from the same population.

Ex. 2 The avarage of marks scored by 32 boys is 72 with standard deviation 8 while that of 36 girls is 70 with standard deviation 6. Test at 1 % level of significance whether the boys perform better than the girls.

Solution: Given:

$n_1 = 32$, $n_2 = 36$

$\overline{x}_1 = 72$, $\overline{x}_2 = 70$

$s_1 = 8$, $s_2 = 6$

i) Null Hypothesis $H_0 : \mu_1 = \mu_2$

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

ii) Calculation of statistic : $\overline{x}_1 - \overline{x}_2 = 72 - 70 = 2$

the standard error (S.E)

$$s = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(8)^2}{32} + \frac{(6)^2}{36}}$$

$$= \sqrt{3}$$

$$\therefore Z = \left| \frac{\overline{x}_1 - \overline{x}_2}{s} \right| = \left| \frac{2}{\sqrt{3}} \right| = 1.15$$

iii) Level of significance : $\alpha = 0.01$ $\left(1\% = \frac{1}{100}\right)$

iv) Critical value : The value of z_α at 1 % level of significance from the table is 2.58

v) Decision:

since the computed value of $Z = 1.15$ is less than the critical value $Z_\alpha = 2.58$

Hence, the Null Hypothesis is accepted.

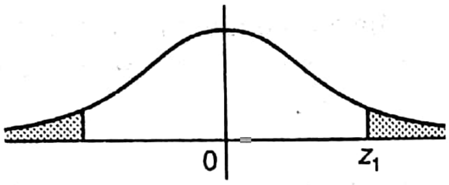
$\therefore$ Boys do not perform better than the Girls.

Homework:

Ex 3 Test the significance of the difference between the means of two normal population with the same standard deviation from the following data.

	Size	Mean	S.D
sample I	100	64	6
sample II	200	67	8

Percentage Points of *t* - distribution



Example

For Φ = 10 d. o. f.

P (| *t* | > 1.812) = 0.1

Φ \ *P*	0.20	0.10	0.05	0.02	0.01
1	3.078	6.314	12.706	31.812	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.287
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.325	2.576

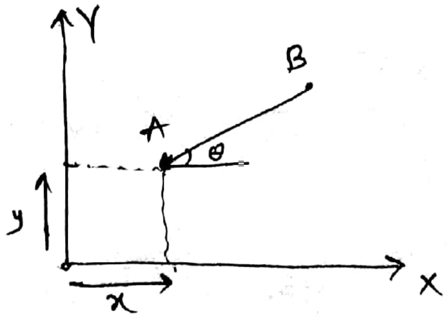
* Student's t - distribution (for small sample) :-
  (t - distribution)

— The t - distribution is used when
  i) The sample size is 30 or less and
  ii) population standard deviation is not known

— Uses of t - distribution :-
  i) To estimate the population mean μ from the sample mean $\bar{x}$
  ii) To Test the hypothesis that the population mean is μ with the help of the sample mean $\bar{x}$
  iii) To Test the hypothesis that two population have same mean with the help of the sample mean.

* Degree of freedom :-

It is defined as the number of independent parameters required to specified the location of every link within a mechanism.



degree of freedom = 3

Note that :

* The sample standard deviation is

$$\boxed{s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1}}$$

* The t-distribution formula is

$$\boxed{t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}}$$

here, the degree of freedom = n-1

Example 1. A Random sample of size 16 from a normal population showed a mean of 103.75 cm. and sum of squares of deviation from the mean 843.75 cm^2. Can we say that the population has a mean of 108.75 cm. ?

solution : Given: $n = 16$, $\mu = 108.75$, $\bar{x} = 103.75$

and $\Sigma (x_i - \bar{x})^2 = 843.75$

i) Null hypothesis (H_o) : $\mu = 108.75$

ii) Alternative hypothesis (H_a): $\mu \neq 108.75$

iii) Calculate of test statistic :

* $s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n} = \frac{843.75}{16} = 52.73$

* $t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{103.75 - 108.75}{\sqrt{52.73}/\sqrt{16-1}}$

$$\Rightarrow \quad t = -\frac{5}{1.875} = -2.67$$

$$\Rightarrow \quad |t| = 2.67$$

iv) Level of significance : $\alpha = 0.05$ (5 %)

v) critical value : The value of t_α for 5 % level of significance from the table is 2.131 currerponds to the degree of freedom = $16 - 1 = 15$

vi) Decision : Note that the computed value $|t| = 2.67$ is greater than the table value $t_\alpha = 2.131$

Henee, the null hypothesis is rejected.

Therefore, we cannot say that the population mean is 108.75

<u>Example 2</u> : Ten individuals are chosen at random from a population and their heights are found to be 63, 63, 64, 65, 66, 69, 69, 70, 70, 71 inches. Discuss the suggestion that the mean height of the universe is 65 inches.

<u>solution</u>: Given: $\mu = 65$ inches

the values of x_i's are

63, 63, 64, 65, 66, 69, 69, 70, 70, 71

$\therefore \quad \bar{x} = \frac{\sum x_i}{n} = \frac{63+63+64+65+66+69+69+70+70+71}{10}$

$= \frac{670}{10} = 67$

$\therefore \quad \bar{x} = 67$

x_i	63	63	64	65	66	69	69	70	70	71
$x_i - \bar{x}$	-4	-4	3	2	1	2	2	3	3	4
$(x_i - \bar{x})^2$	16	16	9	4	1	4	4	9	9	16

$\therefore \quad s^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{16+16+9+4+1+4+4+9+9+16}{10}$

$= \frac{88}{10} = 8.8$

i) Null hypothesis (H_0) : $\mu = 65$

ii) Alternate hypothesis (H_a) : $\mu \neq 65$

iii) Calculation of test statistic :

$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{67-65}{\sqrt{8.8}/\sqrt{10-1}} = \frac{6}{2.97} = 2.02$

$\therefore \quad |t| = 2.02$

iv) Level of significance : $\alpha = 0.05$ (5 %)

v) critical value : The value of t_α at 5 % level of significance corresponding to the degree of freedom $10-1 = 9$ is 2.6

vi) Decision: Note that the computed value of t is 2.02 is less than the table value

$t_\alpha = 2.6$

Hence, the null hypothesis is accepted.

$\therefore$ The mean height of the universe may be 65 inches

Example 3. Test made on breaking strength of 10 pieces of a metal wire gave the following results

578, 572, 570, 568, 572, 570, 570, 572, 596 and 584 kgs.

Test if the breaking strength of the metal wire can be assumed to be 577 kg?

Hint.

$\bar{x} = 575.2$, $\mu = 577$

$s^2 = 68.16$

$|t| = 0.65$

Critical value : 2.25 (DoF : 10-1=9)

Decision: Accepted.

* Testing the Difference between means :-

Case 1 If samples are independent

* formulae for standard deviation :

- Genral :

$$S_p = \sqrt{\frac{\Sigma (x_{i1} - \bar{x}_1)^2 + \Sigma (x_{i2} - \bar{x}_2)^2}{n_1 + n_2 - 2}}$$

- for unbiased standard deviation :

$$S_p = \sqrt{\frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}}$$

- for standard deviation :

$$S_p = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$$

* Standard error (S.E) :

$$\boxed{S.E = S_p \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

if standard deviation of populations σ_1, σ_2 are given then

$$S.E = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

* t - distribution formula:

$$\boxed{t = \frac{\bar{x}_1 - \bar{x}_2}{S.E.}}$$

Example 1. If two independent random samples of sizes 15 and 8 have respectively the following mean and population standard deviation,

$\bar{x}_1 = 980$ $\bar{x}_2 = 1012$

$\sigma_1 = 75$ $\sigma_2 = 80$

Test the hypothesis that $\mu_1 = \mu_2$ at 5% level of significance

solution:- Given: $\bar{x}_1 = 980$, $\bar{x}_2 = 1012$

$\sigma_1 = 75$, $\sigma_2 = 80$

i> Null hypothesis H_0 : $\mu_1 = \mu_2$

ii> Alternate hypothesis : $\mu_1 \neq \mu_2$

iii> Calculation of test statistic :

$$S.E = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{(75)^2}{15} + \frac{(80)^2}{8}}$$

$$= \sqrt{375 + 800} = 34.28$$

Now, $$t = \frac{\bar{x}_1 - \bar{x}_2}{S.E.} = \frac{980 - 1012}{34.28} = -0.93$$

$\therefore$ $|t| = 0.93$

iv) Level of significance : $\alpha = 0.05$ (5%)

v) critical value : The table value of t at 5% level of significance is 1.96

vi) Decision : Note that the computed value $|t| = 0.93$ is less than the table value 1.96

Hence, the Null hypothesis is accepted.

$\therefore$ The population means are equal $\mu_1 = \mu_2$

Example 2. The means of two random samples of size 9 and 7 are 196.42 and 198.82 respectively. The sum of the squares of the deviations from the means are 26.94 and 18.73 respectively. Can the samples be considered to have been drawn from the same population ?

Solution: Given: $n_1 = 9$, $n_2 = 7$

$\bar{X}_1 = 196.42$, $\bar{X}_2 = 198.82$

and $\sum (X_{i1} - \bar{X}_1)^2 = 26.94$, $\sum (X_{i2} - \bar{X}_2)^2 = 18.73$

i) Null Hypothesis H_0 : $\mu_1 = \mu_2$

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

iii) Calculation of test statistic :

The samples standard deviation is

$$S_p = \sqrt{\frac{\Sigma (x_{i1}-\bar{x}_1)^2 + \Sigma (x_{i2}-\bar{x}_2)^2}{n_1+n_2-2}} = \sqrt{\frac{26.94+18.73}{9+7-2}}$$

$\therefore$ $S_p = 1.81$

Now the standard error is

$$S.E = S_p \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (1.81)\sqrt{\frac{1}{9} + \frac{1}{7}}$$

$$= 0.91$$

Now, $$t = \frac{\bar{x}_1 - \bar{x}_2}{S.E.} = \frac{196.42 - 198.82}{0.91} = -2.64$$

$\therefore$ $|t| = 2.64$

iv) Level of significance : $\alpha = 0.05$

v) Critical value : The table value of t at 5 % level of significance corresponding to the degree of freedom $9+7-2=14$ is 2.145

vi) Decision: Note that the computed value $|t|=2.64$ is greater than the table value 2.145

Hence, The Null hypothesis is rejected.

$\therefore$ The samples cannot be considered to have been drawn from the same population

Example 3 Two independent samples of sizes 8 and 7 gave the following results.

sample 1 :	19	17	15	21	16	18	16	14
sample 2 :	15	14	15	19	15	18	16	

Is the difference between sample mean significant

Hint:

* find $\bar{X}_1$ from sample 1 and $\bar{X}_2$ from sample 2 $\left(\bar{X}_1 = 17, \; \bar{X}_2 = 16 \right)$
* find s_1 using sample 1 and s_2 using sample 2 $\left(s_1 = 2.12, \; s_2 = 1.69 \right)$
* find S_p and S.E (S.E = 1.073)
* 't'

  (t = 0.93)

(Decision: Accepted)

Case 2: If samples are not independent

* formulae :
$$s^2 = \frac{\Sigma (x_i - \overline{x})^2}{n}$$

$$t = \frac{\overline{x} - \mu}{s / \sqrt{n-1}}$$

Example ① A certain injection administered to 12 patients resulted in the following changes of blood pressure : 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4

Can it concluded that the injection will be in genral accompanied by an increases in blood pressure ?

solution: Given : $n = 12$ and the values of x_i's are 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4

$$\therefore \overline{x} = \frac{\Sigma x_i}{n} = \frac{5+2+8+(-1)+3+0+6+(-2)+1+5+0+4}{12}$$

$$= 2.58$$

x	5	2	8	-1	3	0	6	-2	1	5	0	4
$x_i - \overline{x}$	2.42	0.58	5.42	-3.58	0.42	-2.58	3.42	-4.58	-1.58	2.42	-2.58	1.42
$(x_i - \overline{x})^2$	5.86	0.34	29.38	12.82	0.18	6.66	11.70	20.98	2.50	5.86	6.66	2.02

Now, $$s^2 = \frac{\Sigma (x_i - \overline{x})^2}{n} = \frac{104.92}{12} = 8.74$$

i) Null hypothesis : H_0 : $\mu = 0$

ii) Alternate hypothesis H_a : $\mu \neq 0$

iii) Calculation of statistic :

The t - distribution is

$$t = \frac{\bar{x} - \bar{u}}{s/\sqrt{n-1}} = \frac{2.58 - 0}{\sqrt{8.74}/\sqrt{12-1}} = 2.89$$

$\therefore \quad |t| = 2.89$

iv) Level of significance : $\alpha = 0.05$

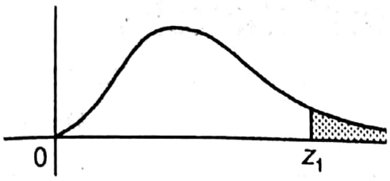
v) critical value : The value of t_α at 5 % level of significance corresponds to the degree of freedom $12-1 = 11$ is 2.201

vi) Decision: Note that the computed value $|t| = 2.83$ is greater that critical value 2.201

Hence, Null hypothesis is rejected.

$\therefore$ There is rise in blood pressure.

Percentage Points of χ^2 - Distribution



Example

For $\Phi = 10$ d. o. f.

$P(\chi^2 > 15.99) = 0.10$

Φ \ P	0 = .99	0.95	0.50	0.10	0.05	0.02	0.01
1	.000157	.00393	.455	2.706	3.841	5.214	6.635
2	.0201	.103	1.386	4.605	5.991	7.824	9.210
3	.115	.352	2.366	6.251	7.815	9.837	11.341
4	.297	.711	3.357	7.779	9.488	11.668	13.277
5	.554	1.145	4.351	9.236	11.070	13.388	15.086
6	.872	1.635	5.348	10.645	12.592	15.033	16.812
7	1.339	2.167	6.346	12.017	14.067	16.622	18.475
8	1.646	2.733	7.344	13.362	15.507	18.168	20.090
9	2.088	3.325	8.343	14.684	16.919	19.679	21.666
10	2.558	3.940	9.340	15.987	18.307	21.161	23.209
11	3.053	4.575	10.341	17.275	19.675	22.618	24.725
12	3 571	5.226	11.340	18.549	21.026	24.054	26.217
13	4.107	5.892	12.340	19.812	22.362	25.472	27.688
14	4.660	6.571	13.339	21.064	23.685	26.873	29.141
15	4.229	7.261	14.339	22.307	24.996	28.259	30.578
16	5.812	7.962	15.338	23.542	26.296	29.633	32.000
17	6.408	8.672	16.338	24.769	27.587	30.995	33.409
18	7.015	9.390	17.338	25.989	28.869	32.346	34.805
19	7.633	10.117	18.338	27.204	30.144	33.687	36.191
20	8.260	10.851	19.337	28.412	31.410	35.020	37.566
21	8.897	11.591	20.337	29.615	32.671	36.349	38.932
22	9.542	12.338	21.337	30.813	33.924	37.659	40.289
23	10.196	13.091	22.337	32.007	35.172	38.968	41.638
24	10.856	13.848	23.337	32.196	36.415	40.270	42 980
25	11.524	14.611	24.337	34.382	37.652	41.566	44.314
26	12.198	15.379	25.336	35.363	38.885	41.856	45.642
27	12.879	16.151	26.336	36.741	40.113	44.140	46.963
28	13.565	16.928	27.336	37.916	41.337	45.419	48.278
29	14.256	17.708	28.336	39.087	42.557	46.693	49.588
30	14.953	18 493	29.336	40.256	43.773	47.962	50.892

* Chi-square Test : (χ^2 - Distribution)

1). To Test independent of attributes :

— χ^2 - test is widely used to test whether there is association between two or more attributes.

In this case the Null hypothesis is like " there is no assoeiation between the attributes".

for ex. It can be used to determine whether there is association between the colour of mother's eye and daughter's eye between Inoculation and prevention of a diseases.

2). To test the Goodness to fit :

χ^2 - test is very commonly known as χ^2 - test of goodness to fit because it enables us to ascertain how well the theoretical distribution such as poeston or Normal fit the observed frequencies. In this case the Null hypothesis is like " The theory supports the observations.

* Examples on Independence of attributes

Ex 1. A sample of 400 students of under-graduate and 400 students of post graduate classes was taken to know their opinion about atonomous colleges. 290 of the undergraduate and 310 of the post-graduate students favoured the autonomous status. present these facts in the form of a table and test at 5 % level, that the opinion regarding autonomous status of colleges in independent of the level of classes of students.

solution:

opinion about autonomous colleges

	Favoured	Not-Favoured	Total
Under-graduate	290	110	400
post-graduate	310	90	400
Total	600	200	800

i) Null hypothesis Ho : There is no association between the classes and the opinion.

Alternative hypothesis Ha : There is association

ii) Calculation of test statistic :

on the basis of hypothesis,

The Number in first cell $= \frac{A \times B}{N}$

where, A = Number of under-graduate students.

B = Number who ferovured

N = Total Number of students.

$\therefore$ The Number in first cell $= \dfrac{400 \times 600}{800} = 300$

opinion about the autonomous colleges.

	Favoured	Not-favoured	Total
Under-graduate	300	100	400
Post-graduate	300	100	400
	600	200	800

Calculation of $(O-E)^2/E$

O	E	$(O-E)^2$	$(O-E)^2/E$
290	300	100	0.33
310	300	100	0.33
110	100	100	1.00
90	100	100	1.00
		Total. (χ^2)	2.66

iii) level of significance : $\alpha = 0.05$ (5%)

Degree of freedom $= (r-1)(c-1) = (2-1)(2-1) = 1$

iv) Critical value : The Table value at 5% level of significance corresponds to 1 DOF is

$\chi^2 = 3.84$

v) Decision: clearly, the calculated value of $\chi^2 = 2.66$ is less than the table value of $\chi^2 = 3.84$

Hence, the Null hypothesis is accepted.

$\therefore$ There is no association between the opinion and the level of classes.

<u>Ex 2.</u> To test the effect of a new drug, a controlled experiment was conducted. 300 patients were given the new drug while 200 patients were given no drug. ~~on~~ the basis of examination of these person, the following result were obtained.

	Cured	condition worsoned	No effect	Total
Given the new drug	200	40	60	300
Not given drug	120	30	50	200
Total	320	70	110	500

Use χ^2 test to find effect of new drug.

<u>Solution:</u> i) Null Hypothesis H_0: The drug is not effective.

Alternative Hypothesis H_a: The drug is effective

ii) Calculation of test statistic:

The Number in the first cell $= \frac{A \times B}{N}$

where, A = total in the first column

B = total in first Row

N = Total number of observations

$\therefore$ Number in first cell $= \frac{320 \times 300}{500} = 192$

IIIy, The Number in second cell $= \frac{70 \times 300}{500} = 42$

$\therefore$ Table of calculated frequencies.

	Cured	Condition worsoned	No effect	Total
Given the new drug	192	42	66	300
Not given the drug	128	28	44	200
Total	320	70	110	500

Calculation of $(O-E)^2/E$

O	E	O - E	$(O-E)^2$	$(O-E)^2/E$
200	192	8	64	0.333
40	42	−2	4	0.095
60	66	−6	36	0.545
120	128	−8	64	0.500
30	28	2	4	0.143
50	44	6	36	0.818
		Total	(χ^2)	2.434

iii) Level of significance; $\alpha = 0.05$

Degree of freedom $= (r-1)(c-1) = (2-1)(3-1) = 2$

iv) Critical value: The Table value at 5% level of significance corresponds to 2 DOF is $\chi^2 = 5.991$

v) Decision: Note that the calculated value of $\chi^2 = 2.435$ is less than the table value $\chi^2 = 5.991$

Hence, the Null hypothesis is Accepted.

$\therefore$ The New drug is not effective.

* Examples on Goodness of fit :-

Ex ① The following table gives the number of accidents in a city during a week. find whether the accidents are uniformly distributed over a week

Day :	Sun	Mon	Tue	Wed	Thu	fri	Sat.	Total
No. of accidents :	13	15	9	11	12	10	14	84

Solution: i) Null Hypothesis H_0 : Accidents are equally distributed over all the days of a week

ii) Alternative hypothesis H_a: Accidents do not occur equally.

iii) Calculation of test statistic :

$$E = \frac{\text{Total Number of Accidents}}{\text{Total days}} = \frac{84}{7} = 12$$

days.	O	E	$(O-E)^2$	$(O-E)^2/E$
Sun	13	12	1	$\frac{1}{12}$
Mon	15	12	9	$\frac{3}{4}$
Tue	9	12	9	$\frac{3}{4}$
Wed	11	12	1	$\frac{1}{12}$
Thu	12	12	0	0
fri	10	12	4	$\frac{1}{3}$
Sat.	14	12	4	$\frac{1}{3}$
			Total (χ^2)	2.33

$\therefore \ \chi^2 = 2.33$

iv) level of significance : $\alpha = 0.05$

Degree of freedom : $n-1 = 7-1 = 6$

v) Critical Value : The table value of χ^2 at 5 % level of significance corresponds to 6 DOF is 12.59

vi) Decision: Note that the calculated value of $\chi^2 = 2.33$ is less than the table value $\chi^2 = 12.59$

Hence, the Null hypothesis is Accepted.

$\therefore$ The accidents occur equally on all working days.

Ex. 2. A die was thrown 132 times and the following frequencies were observed.

No. obtained :	1	2	3	4	5	6	Total
frequency :	15	20	25	15	29	28	132

Test the hypothesis that the die is unbiased.

Solution: i) Null Hypothesis H_0 : The die is unbiased.

ii) Alternative Hypothesis H_a : The die is not unbiased.

iii) Calculation of test statistic :

$$E = \frac{\text{Total frequency}}{\text{Total Number obtained}} = \frac{132}{6} = 22$$

$$\therefore \quad E = 22$$

No.	O	E	$(O-E)^2$	$(O-E)^2/E$
1	15	22	49	2.227
2	20	22	4	0.182
3	25	22	9	0.409
4	15	22	49	2.227
5	29	22	49	2.227
6	28	22	36	1.636
			Total (χ^2)	8.91

$\therefore \chi^2 = 8.91$

iv) level of significance : $\alpha = 0.05$

Degree of freedom : $n-1 = 6-1 = 5$

v) critical Value : The table value of χ^2 at 5% level of significance corresponds to 5 DOF is 11.07

vi) Decision; Note that the calculated value $\chi^2 = 8.91$ is less than the critical value $\chi^2 = 11.07$

$\therefore$ The Null Hypothesis is accepted.

$\therefore$ The die is unbiased.

Ex.3. Theory predicts that the proportion of beans in the four groups A, B, C, D should be 9:3:3:1. In an experiment among 1600 beans the numbers in the four groups were 882, 313, 287 and 118. Does the experimental results support the theory?

solution: i) Null Hypothesis H_0: The proportion of the beans in the four groups A, B, C, D is given proportion 9 : 3 : 3 : 1

ii) Alternative Hypothesis H_a: The proportion is not as given above

iii) Calculation of test statistic:

Note that the sum $9+3+3+1=16$

$\therefore$ The number of beans in the four groups will be

$A = \frac{9}{16} \times 1600 = 900$, $B = \frac{3}{16} \times 1600 = 300$

$C = \frac{3}{16} \times 1600 = 300$, $D = \frac{1}{16} \times 1600 = 100$

O	E	$(O-E)^2$	$(O-E)^2/E$
882	900	324	0.0378
313	300	169	0.5633
287	300	169	0.5633
118	100	324	3.24
		Total (χ^2)	4.72

$\therefore \chi^2 = 4.72$

iv) level of significance : $\alpha = 0.05$

degree of freedom $= n-1 = 4-1 = 3$

v) critical value: The table value of χ^2 at 5 % level of significance corresponds to 3 DoF is 7.81

vi) Decision: Note that the calculated value $\chi^2 = 4.72$ is less than the table value $\chi^2 = 7.81$

Hence, The Null Hypothesis is accepted.

∴ The proportion 9:3:3:1 is correct.

Ex 4. In an experiment on pea breeding the following frequencies were obtained.

Round and yellow	wrinkled and yellow	Round and green	wrinkled and green	Total
315	101	108	32	556

Theory predicts that the frequencies should be in proportion of 9:3:3:1

Examine the correspondence between theory and experiment using chi-square test.