Joint Probability distributions

<u>Defn</u> :- If X and Y are 2 discrete random variables, we define the Joint probability fn of X and Y by
$P(X = x, Y = y) = f(x, y)$ where $f(x, y)$ satisfies the condns

i) $f(x, y) \geq 0$

ii) $\sum_x \sum_y f(x, y) = 1$.

<u>Joint Probability distribution</u> :- The set of values of the fn $P(X = x_i, Y = y_j) = J_{ij}$ for $i = 1, 2, \ldots m$ and $j = 1, 2, \ldots n$ is called the joint Probability distribution of X and Y. These values are presented in the foll Joint probability table :

$X \backslash Y$	y_1	y_2	----	y_n	Sum
x_1	J_{11}	J_{12}	----	J_{1n}	$f(x_1)$
x_2	J_{21}	J_{22}	----	J_{2n}	$f(x_2)$
----	---	---	----	---	----
x_m	J_{m1}	J_{m2}	---	J_{mn}	$f(x_m)$
Sum	$g(y_1)$	$g(y_2)$	----	$g(y_n)$	1

$$\Rightarrow \sum_{i=1}^{m} \sum_{i=1}^{n} f(x_i, y_j) = \sum_{1}^{m} \sum_{1}^{n} J_{ij} = 1.$$

<u>Expectation</u> :- (or) Mean

$$E(X) = \mu_X = \sum_i x_i f(x_i)$$

$$E(Y) = \mu_Y = \sum_j y_j g(y_j)$$

$$E(XY) = \sum_{i,j} x_i y_j J_{ij}.$$

P.T.O

Standard Deviation :

Variance $\sigma_X^2 = E(X^2) - \mu_X^2$ and $\sigma_Y^2 = E(Y^2) - \mu_Y^2$

where $E(X^2) = \sum_i x_i^2 f(x_i)$ and $E(Y^2) = \sum_j y_j^2 g(y_j)$.

S.D of X, $\sigma_X = \sqrt{V(X)}$ $\left(\because V(X) = \sigma_X^2 \;;\; V(Y) = \sigma_Y^2\right)$

$\sigma_Y = \sqrt{V(Y)}$

Covariance :- $\text{cov}(X,Y) = \sum_i \sum_j x_i y_j J_{ij} - \mu_X \mu_Y$

$\text{cov}(X,Y) = E(XY) - \mu_X \mu_Y = E(XY) - E(X) \cdot E(Y)$

Correlation : $\rho(X,Y) = \dfrac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$

Note: If X and Y are independent random variables,

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

ii) $\text{cov}(X,Y) = 0$ and hence $\rho(X,Y) = 0$.

1) The discrete random variables X and Y are said to be independent random variables if

~~$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$ and conversely.~~

$f(x_i)\, g(y_j) = J_{ij}$ and conversely.

2) Marginal distribution of X & Y are obtained by adding all the respective row entries (for X) & all the respective column entries (for Y).

PROBLEMS ON JOIN PROBABILTY DISTRIBUTION

1. The joint Probability distribution table for two random variables X and Y is as follows:

X \ Y	-2	-1	4	5
1	0.1	0.2	0	0.3
2	0.2	0.1	0.1	0

Determine the marginial Probability distribution of X and Y also compute (i) E(x), E(Y) and E(XY) (ii) standard deviation X, Y (iii) covariance (X, Y) (iv) $\rho(X, Y)$ (v) verify X and Y are dependent random variables

Soln Mariginal distribution of X and Y are obtained by adding all the respective row and entries for X, adding all the respective column entries for Y.

x_i	1	2
$f(x_i)$	0.6	0.4

y_j	-2	-1	4	5
$g(y_j)$	0.3	0.3	0.1	0.3

(i) $E(x) = \mu_x = \sum x_i f(x_i)$

$= (1 \times 0.6) + (2 \times 0.4)$

$\mu_x = 1.4$

(ii) $E(Y) = \mu_Y = \sum y_j\, g(y_j) = (-2\times0.3)+(-1\times0.3)+(4\times0.1)+(5\times0.3)$

$$\underline{\mu_Y = 1}$$

$$E(XY) = \sum_{i,j} x_i y_j J_{ij}$$

$$= (1\times-0.2\times0.1)+(1\times-1\times0.2)+(1\times4\times0)+(1\times0.3\times5)+(2\times-2\times0.2)+(2\times-1\times0.1)+(2\times4\times0.1)+(2\times5\times0)$$

$$\underline{E(XY) = 0.9}$$

(ii) $\sigma_X^2 = E(X^2) - \mu_X^2$ where

$$E(X^2) = \sum x_i^2 f(x_i) = (1^2\times0.6)+(2^2\times0.4) = 2.2$$

$$\therefore \sigma_X^2 = 2.2-(1.4)^2 = 0.24$$

$$\sigma_X = \sqrt{0.24}$$

$$\boxed{\sigma_X = 0.4899}$$

$\sigma_Y^2 = E(Y^2) - \mu_Y^2$ where

$$E(Y^2) = \sum y_j^2 g(y_j) = (-2^2\times0.3)+(-1^2\times0.3)+(4^2\times0.1)+(5^2\times0.3) = 10.6$$

$$\sigma_Y^2 = 10.6-(1)^2 = 9.6$$

$$\sigma_Y = \sqrt{9.6}$$

$$\boxed{\sigma_Y = 3.0984}$$

(iii) $\text{Cov}(X,Y) = E(XY) - E(X)\cdot E(Y) = 0.9 - 1.4\times1$

$$\boxed{\text{Cov}(X,Y) = -0.5}$$

(iv) $\rho(X,Y) = \dfrac{cov(X,Y)}{\sigma_X \cdot \sigma_Y}$

$= \dfrac{-0.5}{0.4899 \times 3.0984}$

$\boxed{\rho(X,Y) = -0.3294}$

(v) If X and Y are independent Random variables we must have $f(x_i)\, g(y_i) = J_{ij}$

Here, we have

$f(x_1) \cdot g(y_1) = 0.6 \times 0.3 = 0.18$

& $J_{11} = 0.1$

$\therefore f(x_1) \cdot g(y_1) \neq J_{11}$

In general $f(x_i) \cdot g(y_j) \neq J_{ij}$

$\therefore$ X, Y are dependent random variable

2. Joint Probability distribution X and Y are given below.

X \ Y	1	3	9
2	1/8	1/24	1/12
4	1/4	1/4	0
6	1/8	1/24	1/12

(i) Find mariginal dist. of X and Y

(ii) Find cov (X, Y)

(iii) Find $\rho(X,Y)$

Soln

mariginal distribution

x_i	2	4	6
$f(x_i)$	0.25 1/4	1/2	1/4

y_i	1	3	9
$g(y_i)$	1/2	1/3	1/6

$$E(X) = \mu_X = \sum x_i f(x_i)$$

$$= (2 \times 1/4) + (4 \times 1/2) + (6 \times 1/4)$$

$$\mu_X = 4$$

$$E(Y) = \mu_Y = \sum y_j g(y_j)$$

$$= (1 \times 1/2) + (3 \times 1/3) + (9 \times 1/6)$$

$$\mu_Y = 3$$

$$E(XY) = \sum_{ij} x_i y_j J_{iy}$$

$$= (2 \times 1 \times 1/8) + (2 \times 3 \times 1/24) + (2 \times 9 \times 1/12)$$
$$+ (4 \times 1 \times 1/4) + (4 \times 3 \times 1/4) + (4 \times 9 \times 0) +$$
$$(6 \times 1 \times 1/8) + (6 \times 3 \times 1/24) + (6 \times 9 \times 1/12)$$
$$= 12$$

(ii) $\text{cov}(X, Y) = E(XY) - E(X) \cdot E(Y)$

$$= 0$$

(iii) If $\text{cov}(X, Y)$ is 0

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = 0$$

0 by anything is 0

$$= 0$$

3. The Joint Probability distribution is given below

x \ y	2	3	4
1	0.06	0.15	0.09
2	0.14	0.35	0.21

Determine mariginal distribution of X and Y verify that X and Y are <u>independent variables</u>.

Soln Marginal distribution of x and y are:

x_i	1	2
$f(x_i)$	0.3	0.7

y_j	2	3	4
$g(y_j)$	0.2	0.5	0.3

x and y are set be independent variable $f(x_i)$ and $g(y_j)$ should be equal to J_{ij}

$$f(x_i) \cdot g(y_j) = J_{ij}$$
$$f(x_1) \cdot g(y_1) = 0.3 \times 0.2 = 0.06 = J_{11}$$
$$f(x_1) \cdot g(y_2) = 0.3 \times 0.5 = 0.15 = J_{12}$$
$$f(x_1) \cdot g(y_3) = 0.3 \times 0.3 = 0.09 = J_{13}$$

$$f(x_2) \cdot g(y_1) = 0.7 \times 0.2 = 0.14 = J_{21}$$
$$f(x_2) \cdot g(y_2) = 0.7 \times 0.5 = 0.35 = J_{22}$$
$$f(x_3) \cdot g(y_3) = 0.7 \times 0.3 = 0.21 = J_{23}$$

Thus for every i, j we observe that $f(x_i) \cdot g(y_j) = J_{ij}$, therefore x and y are independent random variables.

4. The Joint Probability distribution of two discrete random variables X and Y $f(x,y) = K(2x+y)$ where x and y are integers such that $0 \le x \le 2$, $0 \le y \le 3$
(i) Find K (ii) Find mariginal distribution of X and Y (iii) Show that X and Y are dependent random variables.

Soln
$X = \{x_i\} = \{0, 1, 2\}$ from $f(x,y) = K(2x+y)$
$Y = \{y_j\} = \{0, 1, 2, 3\}$

Joint Probability distubution table is as below:

$x \backslash y$	0	1	2	3	sum
0	0	k	2k	3k	6k
1	2k	3k	4k	5k	14k
2	4k	5k	6k	7k	22k
sum	6k	9k	12k	15k	42k

$k(2\times 0+2)$
$k(2\times 1+4)$
$k(2\times 1+0)$

(i) we must have $42k = 1$

$$k = 1/42$$

(ii)

x_i	0	1	2
$f(x_i)$	$6/42$	$14/42$	$22/42$

$\frac{42}{42} = 1$

y_j	0	1	2	3
$g(y_j)$	$6/42$	$9/42$	$12/42$	$15/42$

(iii) If x and y are independent random variable we must have

$$f(x_i)\cdot g(y_j) = J_{ij}$$

$$\frac{6}{42} \times \frac{6}{42} \neq 0$$

$$\therefore f(x_i)\cdot g(y_j) \neq J_{ij}$$

Thus x and y are dependent random variable.

5. Compute (i) $P(x=1, y=2)$
(ii) $P(x \geq 1, y \leq 2)$, (iii) $P(x \leq 1, y \leq 2)$
(iv) $P(x+y \geq 2)$ using the joint Probability disturbution for X and Y.

$x \backslash y$	0	1	2	3	sum
0	0	$1/8$	$1/4$	$1/8$	$1/2$
1	$1/8$	$1/4$	$1/8$	0	$1/2$
sum	$1/8$	$3/8$	$3/8$	$1/8$	1

Soln From joint Probability distribution table :-

$J_{11} = 0$, $J_{12} = 1/8$, $J_{13} = 1/4$, $J_{14} = 1/8$

$J_{21} = 1/8$, $J_{22} = 1/4$, $J_{23} = 1/8$, $J_{24} = 0$

(i) $P(x=1, y=2) = J_{23} = 1/8$

(ii) $P(x \geq 1, y \leq 2) = J_{21} + J_{22} + J_{23}$

$= \frac{1}{8} + \frac{1}{4} + \frac{1}{8} = \frac{1}{2}$

(iii) $P(x \leq 1, y \leq 2) = J_{11} + J_{12} + J_{13} + J_{21} + J_{22} + J_{23}$

$= 0 + \frac{1}{8} + \frac{1}{4} + \frac{1}{8} + \frac{1}{4} + \frac{1}{8}$

$= \frac{7}{8}$

(iv) $P(x+y \geq 2)$

x, y takes the values $(0,2)$ $(0,3)$ $(1,1)$ $(1,2)$ $(1,3)$

$\therefore P(x+y \geq 2) = J_{13} + J_{14} + J_{22} + J_{23} + J_{24}$

$= \frac{1}{4} + \frac{1}{8} + \frac{1}{4} + \frac{1}{8} + 0$

$= \frac{3}{4}$

6. X and Y are <u>independent random variables</u>. X takes the values 2, 5, 7 with Probabilities 1/2, 1/4, 1/4 respectively, Y takes the values 3, 4, 5 with Probabilities 1/3, 1/3, 1/3 respectively. (i) Find the Probability Distribution of X and Y

(ii) Show $\text{cov}(X, Y) = 0$

(iii) Find the Probability distribution $Z = X+Y$

Soln Given mariginal distribution is shown has below

x_i	2	5	7
$f x_i$	$1/2$	$1/4$	$1/4$

y_i	3	4	5
$g(y_i)$	$1/3$	$1/3$	$1/3$

(i) Joint Probability distribution table can be obtained from the condition

$$J_{ij} = f(x_i) \cdot g(y_j) \quad \text{where } i, j = 1, 2, 3$$

($\therefore$ Because x and y are independent random variables.)

x \ y	3	4	5	$f(x_i)$
2	$\frac{1}{2}\times\frac{1}{3}$ $1/6$	$\frac{1}{2}\times\frac{1}{3}$ $1/6$	$1/6$	$1/2$
5	$\frac{1}{4}\times\frac{1}{3}$ $1/12$	$1/12$	$1/12$	$1/4$
7	$\frac{1}{4}\times\frac{1}{3}$ $1/12$	$1/12$	$1/12$	$1/4$
$g(y_j)$	$1/3$	$1/3$	$1/3$	1

(ii) $\text{Cov}(X, Y) = E(XY) - E(X) \cdot E(Y)$

$$E(X) = \mu_X = \sum x_i f(x_i)$$

$$= (2 \times 1/2) + (5 \times 1/4) + (7 \times 1/4)$$

$$E(X) = 4$$

$$E(Y) = \sum \cancel{x} y_j \, g(y_j)$$

$$= (3 \times 1/3) + (4 \times 1/3) + (5 \times 1/3)$$

$$E(Y) = 4$$

$$E(XY) = \sum_{ij} x_i y_j J_{ij}$$

$$= \left(2 \times 3 \times \frac{1}{6}\right) + \left(2 \times 4 \times \frac{1}{6}\right) + \left(2 \times 5 \times \frac{1}{6}\right) +$$

$$\left(5 \times 3 \times \frac{1}{12}\right) + \left(5 \times 4 \times \frac{1}{12}\right) + \left(5 \times 5 \times \frac{1}{12}\right) +$$

$$\left(7 \times 3 \times \frac{1}{12}\right) + \left(7 \times 4 \times \frac{1}{12}\right) + \left(7 \times 5 \times \frac{1}{12}\right)$$

$= 4 + 5 + 7$

$\boxed{E(XY) = 16}$

$cov(X, Y) = 16 - 4 \times 4$

$= 16 - 16 = 0$

(iii) $Z = X + Y$

$Z_i = \{5, 6, 7, 8, 9, 10, 11, 12\}$

5+5 & 7+3=10 → $\frac{1}{12}+\frac{1}{12} = \frac{1}{6}$

Z	5	6	7	8	9	10	11	12
P(Z)	1/6	1/6	1/6	1/12	1/12	1/12 + 1/12 = 1/6	1/12	1/12

Observe that $\Sigma P(Z) = 1$

7. A fair coin is tossed thrice. The random variables X and Y are defined as follows:
X = 0 or 1 according as head or tail occurs on the first toss, Y = no of heads.
(i) Determine the distributions of X and Y.
(ii) Determine the joint dist. of X and Y
(iii) Obtain the expectations of X, Y and XY. Also, find S.D's of X and Y.
(iv) Compute Cov (X, Y) and $\rho(X, Y)$

soln
We have the sample space S and the association of random variables X and Y is given below:

S	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
X	0	0	0	0	1	1	1	1
Y	3	2	2	1	2	1	1	0

(i) To find distribution X and Y

$X = \{x_i\} = \{0, 1\}$ and $Y = \{y_j\} = \{0, 1, 2, 3\}$

$P(X=0)$ is $4/8 = 1/2$; $P(X=1)$ is $4/8 = 1/2$

$P(Y=0)$ is $1/8$; $P(Y=1)$ is $3/8$

$P(Y=2)$ is $3/8$; $P(Y=3)$ is $1/8$

x_i	0	1
$f(x_i)$	$1/2$	$1/2$

y_j	0	1	2	3
$g(y_j)$	$1/8$	$3/8$	$3/8$	$1/8$

(ii) To find joint distribution of X and Y:

$J_{ij} = P(X = x_i, Y = y_j)$

$J_{11} = P(X=0, Y=0) = 0$ from Table

$J_{12} = P(X=0, Y=1) = 1/8$

$J_{13} = P(X=0, Y=2) = 2/8 = 1/4$ etc

$\therefore$ Joint Probability distribution of X and Y as below:

X \ Y	0	1	2	3	Sum
0	0	$1/8$	$1/4$	$1/8$	$1/2$
1	$1/8$	$1/4$	$1/8$	0	$1/2$
Sum	$1/8$	$3/8$	$3/8$	$1/8$	1

(iii) $\mu_X = E(X) = \sum x_i f(x_i) = 1/2$

$\mu_Y = \sum y_j g(y_j) = 3/2$

$E(XY) = \sum x_i y_j J_{ij}$

$= 1/2$

$\sigma_X^2 = E(X^2) - \mu_X^2$ and $\sigma_Y^2 = E(Y^2) - \mu_Y^2$

where $E(X^2) = \sum x_i^2 f(x_i)$

$= 0 + 1/2 = \frac{1}{2}$

$E(Y^2) = \sum y_j^2 g(y_j)$

$= 0 + \frac{3}{8} + \frac{3}{2} + \frac{9}{8} = 3$

$\sigma_X^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$ and $\sigma_Y^2 = 3 - \frac{9}{4} = \frac{3}{4}$

$$\sigma_X = \frac{1}{2} \quad \text{and} \quad \sigma_Y = \frac{\sqrt{3}}{2}$$

(iv) $$\text{Cov}(X,Y) = E(XY) - \mu_X \mu_Y$$

$$= \frac{-1}{4}$$

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{-1}{\sqrt{3}}$$

9. The Joint Probability distribution of X and y is below: Find (i) $E(x)$ and $E(y)$ (ii) $E(xy)$ (iii) $\mathrm{Cov}(X, Y)$

X \ Y	-3	2	4
1	0.1	0.2	0.2
2	0.3	0.1	0.1

soln

$$E(x) = \mu_x = \sum x_i f(x_i)$$

x_i	1	2
$f(x_i)$	0.5	0.5

y_i	-3	2	4
$g(y_i)$	0.4	0.3	0.3

$$E(x) = (1 \times 0.5) + (2 \times 0.5)$$
$$= 1.5$$
$$E(y) = (-3 \times 0.4) + (2 \times 0.3) + (4 \times 0.3)$$
$$= 0.6$$

Date ______
Page ______

(ii) $E(XY) = \sum_{ij} x_i y_j J_{ij}$

$= 0.3$

(iii) $Cov(X, Y) = E(XY) - E(X) \cdot E(Y)$

$= 0.3 - 1.5 \times 0.6$

$Cov(X, Y) = -0.6$

Sampling theory

* The Process of selecting a sample from the population is called a sampling.

* The selection of an individual or an item from the population in such a way that each has the same chance of being selected is called as random sampling.

Testing of hypothesis :

imp ** defn

* To arrive at a decision regarding the population thru' a sample of the population, we have to make certain assumption referred to as hypothesis which may or may not be true.

** * The hypothesis formulated for the purpose of its rejection under the assumption that it is true is called the Null hypothesis, denoted by H_0.

* Any hypothesis which is complimentary to the null hypothesis is called alternative hypothesis, denoted by H_1.

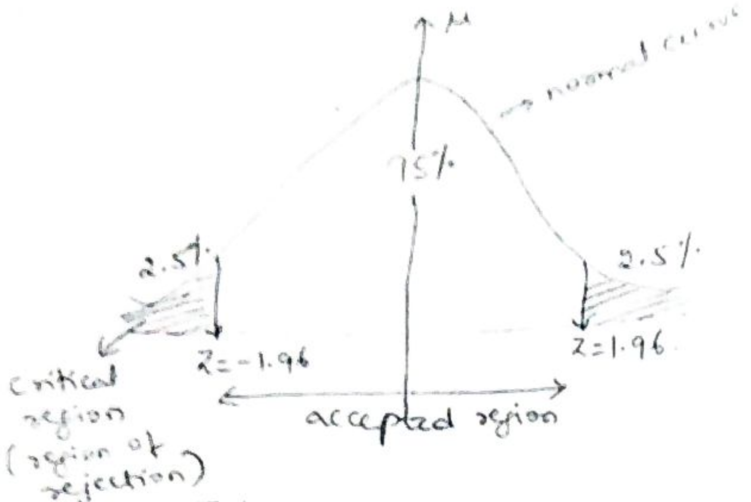
** * Significance level : The probability level, below which leads to the rejection of the hypothesis is known as the significance level. This probability is conventionally fixed at 0.05 or 0.01 ie 5% or 1%. These are called significance levels.

** * The process which helps us to decide about the acceptance or rejection of the hypothesis is called the test of significance.

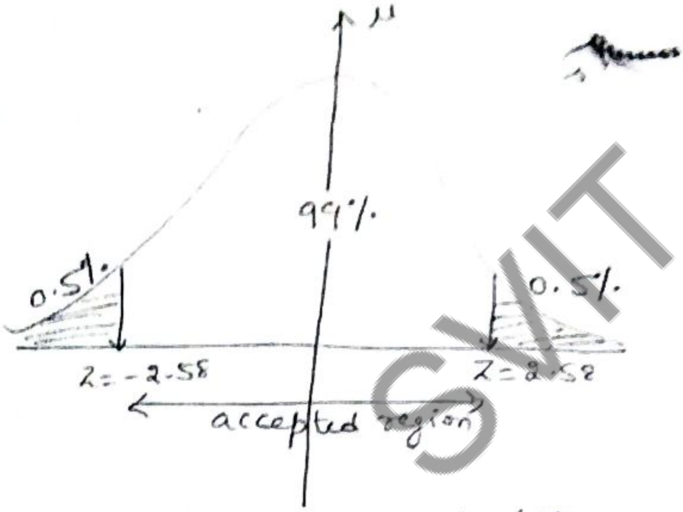
Standard table values of z

$z_{0.05} = 1.96$

2.58



5% level of significance



1% level of significance

Note :-

1) If calculated value of a test is less than table value, then H_0 is accepted.
2) If Calculated value of a test is more than table value, then H_0 is rejected.
3) for a small sample, ie $n \leq 30$, apply t-test.
4) for a large sample, ie $n > 30$, apply z-test.

imp
** Confidence intervals :-

From fig 1 we see that 95% of the area lies between $Z = -1.96$ & $Z = +1.96$ that is with 95% confidence we say that z lies between ie -1.96 and $+1.96$. similarly from fig 2 99% of the area lies between $Z = -2.58$ & $Z = +2.58$ that is $-2.58 \leq Z \leq 2.58$ this are confidence intervals.

** Type 1 and Type 2 errors :-

In a test process, there can be four possible situations of which 2 of the situations leads to two types of errors as given below.

	Accepting the hypothesis	Rejecting the hypothesis
Hypothesis true	Correct decision	wrong decision (Type - I error)
Hypothesis False	Wrong decision (Type - II error)	Current decision

TEST OF SIGNIFICANCE OF PROPORTIONS:

Standard normal variant, $z = \frac{x - np}{\sqrt{npq}}$ where

n is a sample size, x is observed number of success, p is probability of success, q is probability of failure.

NOTE :-

1. $p \pm 2.58 \sqrt{\frac{pq}{n}}$ are the Probable limits at 1% level of significance.
2. $p \pm 1.96 \sqrt{\frac{pq}{n}}$ are the Probable limits at 5% level of significance. where $\sqrt{\frac{pq}{n}}$ is the standard error proportion of success.

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

Soln unbiased coin is fair coin whose Probabilites are ½.

Let us, suppose that coin is unbiased

∴ Probability of getting a head in one toss is equal to $p = 1/2$

$$q = 1 - p$$
$$= 1 - 1/2$$
$$q = 1/2$$

By data, sample size $n = 1000$

observed number of success $x = 540$

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

$$= \frac{540 - 1000 \times 1/2}{\sqrt{1000 \times 1/2 \times 1/2}}$$

$$z = 2.53 \begin{cases} > z_{0.05} = 1.96 \\ < z_{0.01} = 2.58 \end{cases}$$

Thus the hypothesis is accepted that 1% and rejected that 5% level of significance therefore coin is unbiased 1% level of significance.

2. In 324 rows of 6 face die, odd number turned up 181 times. Is it reasonable to think that die is unbiased one.

Soln
Let us assume that die is unbiased

therefore Probability of odd number turning up in a single throw is $p = 3/6 = 1/2$

$$q = 1 - p$$
$$= 1 - 1/2 = 1/2$$

$n = 324$

observed number of success $x = 181$

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

$$= \frac{181 - 324 \times 1/2}{\sqrt{324 \times 1/2 \times 1/2}}$$

$Z = 2.11$ $\begin{cases} > Z_{0.05} = 1.96 \\ < Z_{0.01} = 2.58 \end{cases}$

Thus the die unbiased at 1% level of significance

3. A die thrown 9000 times and a throw off 3 or 4 was observed 3240 times. shown that the die cannot be regarded has a unbiased one.

3 or 4 two choices out of 6

Soln

$n = 9000$ times

$x = 3240$ times

$p = 2/6 = 1/3$

$q = 1 - p$

$= 1 - 1/3 = 2/3$

$$Z = \frac{x - np}{\sqrt{npq}}$$

$$= \frac{3240 - 9000 \times 1/3}{\sqrt{9000 \times 1/3 \times 2/3}}$$

$Z = 5.366$ $\begin{cases} > Z_{0.05} = 1.96 \\ > Z_{0.01} = 2.58 \end{cases}$

Thus die cannot be regarded has a unbiased one.

4. A manufacture claimed that atleast 95% of equipment which he supplied to a factory conformed two specification. An Examination of a sample of 200 Pieces of equipment revealed that 18 of them were faulty. Test is claim ~~at the~~ a significance level of 1% and 5%.

Soln

95% of equipment conformed to specification implies 5% of them may be faulty.

Let x denote observed number of success that is number of faulty item

$\therefore \quad x = 18 \quad , \quad n = 200$

$p = 5\% = 5/100 = 0.05$

$q = 1 - p$

$= 1 - 0.05 = 0.95$

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

$$= \frac{18 - 200 \times 0.05}{\sqrt{200 \times 0.05 \times 0.95}}$$

$$z = 2.59 \begin{cases} > z_{0.05} = 1.96 \\ > z_{0.01} = 2.58 \end{cases}$$

Thus, the manufacture claim is not supported

5. A biased (unfair) coin is tossed 500 times and head turns up 120 times. Find the 95% confident limits for the Proportion of head turning up in infintely many tosses. Given that $z_c = 1.96$

Soln

Since the coin is biased, the Probability of head turning up $p = \frac{120}{500} = 0.24$

$q = 1 - p$

$= 1 - 0.24 = 0.76$

By data, sample $n = 500$

$\therefore$ 95% confidence limits are

$$p \pm 1.96 \sqrt{\frac{pq}{n}}$$

$$= 0.24 \pm 1.96 \sqrt{\frac{0.24 \times 0.76}{500}}$$

$$= 0.24 \pm 0.0374$$

$$= 0.24 - 0.0374 \quad \& \quad 0.24 + 0.0374$$

$= 0.2026$ and 0.2774

Therefore, 20 % to 28 % Proportion of heads turns up in infinetly many tosses under 5% level of significance.

6. A sample of 900 days was taken in a coastal town and it was found that on 100 days the weather was very hot. Obtain the Probable limits of the Percentage of very hot weather given $Z_C = 2.58$

Soln

By data, sample size $n = 900$

Probability of very hot weather $= p = \frac{100}{900}$

$p = 0.1111 \qquad p = \frac{1}{9}$

$q = 1 - p$

$= 1 - 1/9$

$q = 8/9$

Probable limits $= p \pm 2.58 \sqrt{\frac{pq}{n}}$

$= \frac{1}{9} \pm 2.58 \sqrt{\frac{1/8 \times 8/9}{900}}$

$= \frac{1}{9} \pm 0.027$

$= \frac{1}{9} - 0.027$ and $\frac{1}{9} + 0.027$

$= \begin{matrix} 0.0841 \\ 8.4\% \end{matrix}$ to 0.1381

Probable limits of very hot weather is 8.4% to 13.8% or 8% to 14%

7. A survey conducted in slam localty of 2000 families by selecting a sample of size 800

it has revealed that 180 families were illetrate find the Probable limits of the illetrate families in the Population of 2000.

Sol) By data, sample size $n = 800$

Probability illetrates $p = \frac{180}{800}$

$p = 0.225$

$q = 1 - p$

$= 1 - 0.225$

$q = 0.7750$

Probable limits $= 1.96 \pm \sqrt{\frac{pq}{n}}$

$= p \pm 1.96\sqrt{\frac{pq}{n}}$

$= 0.225 \pm 1.96\sqrt{\frac{0.225 \times 0.775}{800}}$

$= 0.225 \pm 0.0289$

$= 0.225 - 0.0289$ and $0.225 + 0.0289$

$= 0.1961$ and 0.2539

$= 19.61$ and 25.39

For a Population of 2000 Probable limits illertate families are $0.1961 \times 2000 = 392$

$0.2539 \times 2000 = 508$

(or)

Probable limits $= p \pm 2.58\sqrt{\frac{pq}{n}}$

$= 0.225 \pm 2.58\sqrt{\frac{0.225 \times 0.775}{800}}$

$= 0.225 \pm 0.0381$

$= 0.225 - 0.0381$ and $0.225 + 0.0381$

$= 0.1869$ and 0.2631

$0.1869 \times 2000 = 374$

$0.2631 \times 2000 = 526$ are the Probability illetrate families out of 2000 families.

8. A die is tossed 960 time and 5 appears 184 times is the die biased at 5% of significance.

Soln By data, $n = 960$

$x = 184$ times

$p = \frac{1}{6}$

$q = 1 - p$

$= 1 - 1/6 = 5/6$

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

$$= \frac{184 - 960 \times 1/6}{\sqrt{960 \times 1/6 \times 5/6}}$$

$$= 2.07 \begin{cases} > z_{0.05} = 1.96 \\ < z_{0.01} = 2.58 \end{cases}$$

TEST OF SIGNIFICANCE OF SINGLE MEAN

Standard normal variant $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ where

σ is the standard deviation of Population

μ mean of the Population, $\bar{x}$ is the mean of the sample.

NOTE :-

1. If σ is not known, then we use alternate formula for z given by $z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ where s is the standard deviation of sample.
2. This z is to test whether difference between the sample mean $\bar{x}$ and population mean μ is significant or not. So we assume the null hypothesis H_0 as there is no significant

difference between the two means

3. 95% confidence limits are given by

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

4. 99% confidence limits are given by

$$\bar{x} - 2.58\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.58\frac{\sigma}{\sqrt{n}}$$

1. A random sample of 400 items choosen from an infinite population is found to have a mean of 82 and standard deviation of 18. Find the 95% confidence limits for the mean of the population from which the sample is drawn.

Soln

By data, $n = 400$ (sample size)

mean of sample $\bar{x} = 82$

standard deviation of the sample $s = 18$

95% confidence limits are given by

$$\bar{x} - 1.96\frac{s}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{s}{\sqrt{n}}$$

$$82 - 1.96\frac{18}{\sqrt{400}} < \mu < 82 + 1.96\frac{18}{\sqrt{400}}$$

$$80.2360 < \mu < 83.7640$$

Thus 95% confidence limits are 80.236 to 83.764.

2. A sample of 100 tyres is taken from lot. The mean life of tyres is found to be 39350 kms with standard deviation 3260 can it be consider as a true random sample from a Population with mean life of 40000 kms?

Date
Page

use 0.05 log. Establish 99% confidence limits within which mean life of tyres expected to lie.

Soln Let H_0 : there is no significant difference between the two means. By data, $n = 100$.

$\bar{x} = 39350$, $s = 3260$, $\mu = 40000$

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{39350 - 40000}{3260/\sqrt{100}}$$

$$Z = -1.9939$$

$$\boxed{|Z| = +1.9939} > Z_{0.05} = 1.96$$

That 5% H_0 is rejected at 5% level of significance, ie that cannot be consider true random sample from a Population of $\mu = 40000$ kms

99% confidence limits is given by

$$\bar{x} - 2.58 \frac{s}{\sqrt{n}} < \mu < \bar{x} + 2.58 \frac{s}{\sqrt{n}}$$

$$39350 - 2.58 \frac{3260}{\sqrt{100}} < \mu < 39350 + 2.58 \frac{3260}{\sqrt{100}}$$

$$38508.92 < \mu < 40191.08$$

Hence μ excepted to lie between 38509 to 40191

3. It has been found that mean breaking strength of a particular brand of thread is 275.6 gms with $\sigma = 39.7$ gms. A sample of 36 pieces of thread showed a mean breaking strength of 253.2 gms. Test claim that 1% and 5% levels of significance.

Let Ho : There is no significant diff b/w the two me...

Soln By data, $\mu = 275.6$

$\sigma = 39.7$

$\bar{x} = 253.2$, $n = 36$

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

$$= \frac{253.2 - 275.6}{39.7/\sqrt{36}}$$

$$= -3.3854$$

$$|z| = +3.3854 \begin{cases} > Z_{0.05} = 1.96 \\ > Z_{0.01} = 2.58 \end{cases}$$

Ho is rejected at both the level of significance that is there is a significant difference between the 2 means.

4. Mean life of 100 fluroscent tube lights manufactured by a company 1570 hours with a standard deviation of 120 hours. Test the hypothesis that mean life time of the lights produced by the company is 1600 hours at 0.01 level of significance

Soln Ho : The mean life time of the lights produced by the company is 1600 hours

By data, $n = 100$

$\bar{x} = 1570$, $s = 120$, $\mu = 1600$

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{1570 - 1600}{120/\sqrt{100}}$$

$$z = -2.5$$

$$|z| = 2.5 < Z_{0.01} = 2.58$$

Ho is accepted that 1% level of significance.

Student's 't' distribution

I Student's 't' test for a sample mean :-

$$\boxed{t = \frac{(\bar{x} - \mu)\sqrt{n}}{s}}$$ where

$\bar{x} = \frac{\Sigma x}{n}$ is the mean of the sample.

μ is the mean of the population.

n is the sample size

s is the S.D of the sample given by

$$s^2 = \frac{1}{(n-1)}\left[\Sigma x^2 - \frac{1}{n}(\Sigma x)^2\right]$$

II Test of significance of difference b/w sample means:-

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ where $\bar{x}, \bar{y}$ are the mean of the 2 samples.

n_1, n_2 are the sample sizes.

s is the S.D of the sample

given by $$s^2 = \frac{1}{n_1+n_2-2}\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n_1} + \Sigma y^2 - \frac{(\Sigma y)^2}{n_2}\right]$$

(or) $$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}$$ where s_1, s_2 are the S.D of the 2 samples.

Note:- 1) for a single sample of size n, d.f $(\nu) = n-1$. degree of freedom $\nu = n-1$.

2) for 2 samples of sizes n_1 & n_2, d.f, $(\nu) = n_1+n_2-2$

3) If $|t| <$ table value, we accept the hypothesis.

If $|t| >$ table value, we reject " ——— "

PROBLEMS ON STUDENT'S 'T' DISTRIBUTION

1. 10 individuals are choosen at random from a population there heights in inches are found to be. Test the hypothesis that 63, 63, 66, 67, 68, 69, 70, 70, 71, 71. Test the hypothesis that mean height of the universe is 66 inches ($t_{0.05} = 2.262$ for 9 d.f)

Soln

$$t = \frac{(\bar{x}-\mu)\sqrt{n}}{s} \quad \text{--- (1)}$$

by data, $n = 10$

Let Ho, There is no significant difference the 2 means.

Assume $\mu = 66$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{678}{10}$$

$$\bar{x} = 67.8$$

$$s^2 = \frac{1}{n-1}\left[\Sigma x^2 - \frac{1}{n}(\Sigma x)^2\right]$$

$$= \frac{1}{n-1}\left[46050 - \frac{1}{10}(678)^2\right]$$

$$= \frac{1}{9}\left[46050 - \frac{1}{10}(678)^2\right]$$

$$s^2 = 9.0667$$

$$\therefore \boxed{s = 3.0111}$$

Sub in 1

$$t = \frac{(67.8 - 66)\sqrt{10}}{3.0111}$$

$$t = 1.8904 < 2.262 = t_{0.05}$$

Thus, Ho is accepted.

2. 9 items of a sample have the following values 45, 47, 50, 52, 48, 47, 49, 53, 51. Does the mean of this differ significantly from the assumed mean of 47.5 (given $t_{0.05} = 2.306$ for 8 d.f.)

Soln Let H_0 : There is no significant difference between the mean of sample and mean of population i.e. assume $\mu = 47.5$

By data, $n = 9$

$$t = \frac{(\bar{x} - \mu)\sqrt{n}}{s} \quad --- (1)$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{442}{9}$$

$$\boxed{\bar{x} = 49.1111}$$

$$s^2 = \frac{1}{n-1}\left[\Sigma x^2 - \frac{1}{n}(\Sigma x)^2\right]$$

$$= \frac{1}{8}\left[21762 - \frac{1}{9}(442)^2\right]$$

$$s^2 = 6.8611$$

$$\boxed{s = 2.6194}$$

sub in eqn 1

$$t = \frac{(49.1111 - 47.5)\sqrt{9}}{2.6194}$$

$\cancel{t = 2.2279}$ $\quad t = 1.8452 < 2.306$

Thus, H_0 is accepted

**3. A certain stimulus adminstered to each of the 12 Patients resulted in the following change in blood Pressure : 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4 can it be concluded that the stimulus will increase the blood Pressure? ($t_{0.05}$ for 11 d.f = 2.201)

Date
Page

Soln Ho : stimulus administration is not accomplied with increase in blood Pressure then. we can take $\mu = 0$

$$t = \frac{(\bar{x} - \mu)\sqrt{n}}{s} \quad \cdots (1)$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{31}{12}$$

$$\bar{x} = 2.5833$$

$$s^2 = \frac{1}{n-1}\left[\Sigma x^2 - \frac{1}{n}(\Sigma x)^2\right]$$

$$= \frac{1}{11}\left[185 - \frac{1}{12}(31)^2\right]$$

$$s^2 = 9.5379$$

$$s = 3.0883$$

sub in (1)

$$t = \frac{(2.5833 - 0)\sqrt{12}}{3.0883}$$

$$t = 2.8977 > 2.201 = t_{0.05}$$

Thus, Ho is rejected. That is Yes stimulus will increase the blood Pressure

4. A mechanist is making engine Parts with axle diameter of ~~0.07~~ 0.7 inch a random sample of 10 Parts shows mean diameter 0.742 inch with a standard deviation of 0.04 inch. On on the bais of this sample would you say that work is inferior? (for 9 degree of freedom $t_{0.05} = 2.262$)

Soln Ho : The Product is not inferior. ie there is no significant difference between $\bar{x}$ and μ.

$$t = \frac{(\bar{x} - \mu)\sqrt{n}}{s}$$

$$= \frac{(0.742 - 0.7)\sqrt{10}}{0.04}$$

$$= 3.3204 > 2.262 \quad t_{0.05}$$

Thus Ho is rejected.

5. Two types of batteries are tested for there length of life and following results were obtained

Battery A : $n_1 = 10$, $\bar{x}_1 = 500$ hrs, $s_1^2 = 100$

Battery B : $n_2 = 10$, $\bar{x}_2 = 560$ hrs, $s_2^2 = 121$

Compute students t and test whether there is a significant difference in the two means. ($t_{0.05} = 2.101$ for 18 d.f)

soln

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{--- } (1)$$

how to calculate $df = n_1 + n_2 - 2$
$= 10 + 10 - 2$
$= 18$

where $s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$

$$= \frac{10 \times 100 + 10 \times 121}{10 + 10 - 2}$$

$s^2 = 122.7778$

$s = 11.0805$

sub in eqn 1

$$t = \frac{500 - 560}{11.0805\sqrt{\frac{1}{10} + \frac{1}{10}}}$$

$|t| = |-12.1081|$

$|t| = 12.1081 > t_{0.05} = 2.101$

6. 11 Students were given a test in statistics. They were given a month's further tution and second test of equal difficulty was held at the end

of it. Do the marks give evidence that the students have benefited by extra coaching (Give $t_{0.05} = 2.086$ for 20 df)

$n_1 = 11$, $n_2 = 11$, $11+11-2 = 20$

Boys	1	2	3	4	5	6	7	8	9	10	11
Marks I test (x)	23	20	19	21	18	20	18	17	23	16	19
Marks II test (y)	24	19	22	18	20	22	20	20	23	20	17

soln Let Ho be assumed there is no significant difference between means of two test marks ie students have not benefited by extra coaching.

$\cancel{t = \bar{x}_1 - x_2}$

By data, $n_1 = 11$, $n_2 = 11$

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{--- (1)}$$

$$\bar{x} = \frac{\Sigma x}{n_1} = \frac{214}{11} \qquad \bar{y} = \frac{\Sigma y}{n_2} = \frac{225}{11}$$

$$\bar{x} = 19.4545 \qquad \bar{y} = 20.4545$$

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma x^2 - \frac{1}{n_1}(\Sigma x)^2 + \Sigma y^2 - \frac{1}{n_2}(\Sigma y)^2\right]$$

$$= \frac{1}{11+11-2}\left[4214 - \frac{1}{11}(214)^2 + 4647 - \frac{1}{11}(225)^2\right]$$

$$= \frac{1}{20}\left[50.7273 + 44.7273\right]$$

$$s^2 = 4.7727$$

$$s = 2.1847$$

sub in (1)

$$t = \frac{19.4545 - 20.4545}{2.1847\sqrt{\frac{1}{11} + \frac{1}{11}}}$$

$|t| = |-1.0735|$

$|t| = 1.0735 < 2.086 = t_{0.05}$

H_0 is accepted at 5% significance, i.e. students are not benefited by extra coach

7. Two horses A and B were tested according to time in sec to run a particular race with the following results

Horse A (x) : 28 30 32 33 33 29 34

Horse B (y) : 29 30 30 24 27 29

Test whether it discriminate b/w two horses (Given $t_{0.05} = 2.2$ for 11 d.f)

Soln

Let H_0 be assumed as there is no significant difference between the two horses that is there is no discrimination between the two horses.

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{--- } (1)$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{219}{7} \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{169}{6}$$

$$\bar{x} = 31.2857 \qquad \bar{y} = 28.1667$$

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma x^2 - \frac{1}{n_1}(\Sigma x)^2 + \Sigma y^2 - \frac{1}{n_2}(\Sigma y)^2\right]$$

$$= \frac{1}{7+6-2}\left[6883 - \frac{1}{7}(219)^2 + 4787 - \frac{1}{6}(169)^2\right]$$

$$= \frac{1}{11}\left[31.4286 + 26.8333\right]$$

$$s^2 = 5.2965$$

$$s = 2.3014$$

sub in (1)

$t = 31.1286 - 2.6.$

$$t = \frac{31.2857 - 28.1667}{2.3019\sqrt{\frac{1}{7} + \frac{1}{6}}}$$

$t = 2.7360 > 2.2 = t_{0.05}$

Ho is rejected => there is a significant difference b/w two mean.

CHI - SQUARE DISTRIBUTION

$$\chi^2 = \frac{\Sigma (O_i - E_i)^2}{E_i} \qquad i = 1, 2, 3 \ldots n$$

where O_i and E_i are respectively the observed and estimated frequencies. (estimated / theortical)

NOTE :-

1. As a test of goodness of fit, value of χ^2 is used to study corresponds between observe and theortical frequency.
2. Degree of freedom $\nu = n-1$
3. If the calculated value of χ^2 is less than calcut corresponding tabulated value then we expect the null hypothesis conclude that there is good corresponding between theory and experiment, otherwise we reject null hypothesis conclude that experiment does not support theory
4. If the excepted frequencies are less than 5, we group them suitably for computing the value is χ^2.

1. A die is thrown 264 times and the number appearing on the face (x) follows the following frequency distribution. Calculate the value of χ^2 and test the hypothesis that die is unbiased given that $\chi^2_{0.05}(5) = 11.07$ and $\chi^2_{0.01}(5) = 15.09$

x	1	2	3	4	5	6
f	40	30	26	52	52	60

Soln: Let Ho be assumed that die is unbiased so that expected number of frequencies for the numbers 1, 2, 3, 4, 5, 6 to appear on the face is $\frac{264}{6} = 44$ each.

Thus we have

No. on the dice	1	2	3	4	5	6
observed frequency (O_i)	40	30	26	56	52	60
Estimated frequency (E_i)	44	44	44	44	44	44

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

$$= \frac{(40-44)^2}{44} + \frac{(30-44)^2}{44} + \frac{(26-44)^2}{44} + \frac{(56-44)^2}{44} + \frac{(52-44)^2}{44} + \frac{(60-44)^2}{44}$$

$$\chi^2 = 22.73 > \begin{cases} \chi^2_{0.05} = 11.07 \\ \chi^2_{0.01} = 15.09 \end{cases}$$

Therefore, Ho is rejected at both level of significance ie the die is biased.

2. The following table gives number of aircraft accident that occur during various days of week. find whether the accidents are uniformly distributed over the week. Given $\chi^2_{0.05} = 12.59$ for 6 degree of freedom

Days	Sun	Mon	Tue	Wed	Thur	Fri	Sat	Total
No of accidents	14	16	8	12	11	9	14	84

Soln Let Ho assume that there is no significant difference between O_i and E_i. That is accident are uniformly distributed over the week

The number of accidents on different days of the week are observed frequencies. excepted frequency assuming the accidents are uniformly distributed $\frac{84}{7} = 12$ each

Days	Sun	Mon	Tue	Wed	Thur	Fri	Sat
observed frequency(O_i)	14	16	8	12	11	9	14
estimated frequency(E_i)	12	12	12	12	12	12	12

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

$$= \frac{(14-12)^2}{12} + \frac{(16-12)^2}{12} + \frac{(8-12)^2}{12} + \frac{(12-12)^2}{12}$$

$$+ \frac{(11-12)^2}{12} + \frac{(9-12)^2}{12} + \frac{(14-12)^2}{12}$$

$$\chi^2 = 4.17 < \chi^2_{0.05} = 12.59$$

$\therefore$ Ho is accepted at 5% level of significance

Yes uniformly distributed week.

3. A sample analysis of examination results 500 students was make. It was found that 220 students had failed, 170 has secured 3rd class, 90 secured 2nd class and 20 had secured 1st class. Do this figures support the general examination result in the ratio 4:3:2:1 for the respective categories ($\chi^2_{0.05} = 7.81$ for 3 d.f)

Soln observed frequencies 220, 170, 90, 20 respectively

Let us take Ho be this figures support to the general result in the ratio 4:3:2:1

Therefore expected frequencies in the respective category are $\frac{4}{10} \times 500$, $\frac{3}{10} \times 500$, $\frac{2}{10} \times 500$, $\frac{1}{10} \times 500$

ie 200, 150, 100, 50

Thus we have:

O_i	220	170	90	20
E_i	200	150	100	50

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

$$= \frac{(220-200)^2}{200} + \frac{(170-150)^2}{150} + \frac{(90-100)^2}{100} + \frac{(20-50)^2}{50}$$

$$\chi^2 = 2 + 2.6667 + 1 + 18$$

$$\chi^2 = 23.67 > \chi^2_{0.05} = 7.81$$

Ho is rejected that 5% level of significance.

*4. Genetic theory state that children having one Parent of blood type M and other of blood type N will always be r of 1/3 times M, MN, N that the propotions of this types will be on an

Average 1 : 2 : 1. Report says that out of 300 children having one M Parent and one N Parent, 30% were found to be of type M 45% of type MN and remaining of type N. Test the theory of χ^2 test given that $\chi^2_{0.05} = 5.99$ for 2 d.f

Soln Let Ho assumed as there is a good corresponde between observed and theortical frequencies.

By data, observed frequencies are:-

	M	MN	N
Oi	$\frac{30}{100} \times 30$	$\frac{45}{100} \times 300$	$\frac{25}{100} \times 300$
	$= 90$	$= 135$	$= 75$

Estimated frequencies are: 1 : 2 : 1, given by

$\frac{1}{4} \times 300$, $\frac{2}{4} \times 300$, $\frac{1}{4} \times 300$

$= 75 \qquad = 150 \qquad = 75$ respectively.

Thus we have:

Oi	90	135	75
Ei	75	150	75

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

$$= \frac{(90-75)^2}{75} + \frac{(135-150)^2}{150} + \frac{(75-75)^2}{75}$$

$$\chi^2 = 4.5 < 5.99 = \chi^2_{0.05}$$

Thus Ho hypothesis is accpeted for 5% level of significance.

5. Fit a Poisson distribution for the following data test the goodness of fit given that $\psi^2_{0.05} = 7.815$ for 3 degree of freedom.

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

Soln

Let Ho be the fitness is good

To find theortical frequencies Ei

Mean $\mu = \frac{\Sigma f.x}{\Sigma f}$

$= \frac{0+60+30+6+4}{200}$

$\boxed{\mu = 0.5 = m}$ for a Poisson distribution

Poisson distribution is given by

$P(x) = \frac{m^x e^{-m}}{x!}$

$P(x) = \frac{(0.5)^x e^{-0.5}}{x!}$

Let $f(x) = 200 \times P(x)$

$= \frac{200 \times (0.5)^x e^{-0.5}}{x!}$

$= 200 \times e^{-0.5} \times \frac{(0.5)^x}{x!}$

$= 121.3061 \times \frac{(0.5)^x}{x!}$ Put $x = 0,1,2,3,4$

We have

x	0	1	2	3	4
O_i	60	60	15	2	1
E_i	121	61	15	3	0

Since the last two expected frequency are less than 5 we club them with Previous one & rewrite has

O_i	122	60	$15+2+1=18$
E_i	121	61	$15+3+0=18$

$$\psi^2 = \frac{\Sigma(O_i - E_i)^2}{E_i}$$

$$= \frac{(122-121)^2}{121} + \frac{(60-61)^2}{61} + \frac{(18-18)^2}{18}$$

$$\psi^2 = 0.0247 < 7.815 = \psi^2_{0.05}$$

$\therefore$ fitness is considered to be good, Ho Hypthesis accepted

6. Four coins are tossed 100 times and following results were obtained fit a binomial distribution for the data, test the goodness of fit given $\psi^2_{0.05} = 9.49$ for 4 d.f

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

Soln Let x denote number of heads and f corresponding frequency. Let Ho be assumed as fitness is good.

$$\text{Mean } \mu = \frac{\Sigma f x}{\Sigma f}$$

$$= \frac{0+29+72+75+20}{100}$$

$\boxed{\mu = 1.96 = np}$ for binomial distribution

$\mu = np$

$1.96 = 4 \times p$

$p = \frac{1.96}{4}$ $\quad$ $p = 0.49$

$q = p - 1$

$q = 0.51$

Binomial distribution : $^nC_x \, p^x q^{n-x}$

$$P(x) = {}^4C_x (0.49)^x (0.51)^{4-x}$$

Let $f(x) = 100 \times P(x)$

$$= 100 \times {}^4C_x (0.49)^x (0.51)^{4-x}$$

Put $x = 0, 1, 2, 3, 4 \ldots$ To get E_i

we have

x	0	1	2	3	4
O_i	5	29	36	25	5
E_i	7	26	37	24	6

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

$$= \frac{(5-7)^2}{7} + \frac{(29-26)^2}{26} + \frac{(36-37)^2}{37} + \frac{(25-24)^2}{24} + \frac{(5-6)^2}{6}$$

$$\chi^2 = 1.1529 < 9.49 = \chi^2_{0.05}$$

H_0 Hypothesis is accepted.

NOTE:-

For a Poisson distribution degree of freedom d.f, $\nu = n-2$

For a binomial distribution $\nu = n-1$

For a normal distribution d.f $\nu = n-3$