MODULE-5

JOINT PROBABILITY DISTRIBUTION & SAMPLING THEORY

Joint Probability :

Let $X = \{x_1, x_2, x_3, ...x_m\}$ and $Y = \{y_1, y_2, y_3, ...y_n\}$ are two discrete random variables , then the joint probability function of X and Y is defined as

$$P(X = x_i, Y = y_j) = P(x_i, y_j) = f(x_i, y_j) = p_{ij} = f_{ij}$$

where the function $f(x, y)$ satisfy the conditions

i) $f(x, y) \geq 0$ ii) $\sum_i \sum_j f(x_i, y_j) = 1$

The joint probability table as shown below

X \ Y	y_1	y_2	y_3	.....	y_n	$f(x_i)$
x_1	p_{11}	p_{12}	p_{13}	.....	p_{1n}	$f(x_1)$
x_2	p_{21}	p_{22}	p_{23}	.....	p_{2n}	$f(x_2)$
x_3	p_{31}	p_{32}	p_{33}	.....	p_{3n}	$f(x_3)$
. . . .	. . . .	. . . .	. . . .	....	. . . .	. . . .
x_m	p_{m1}	p_{m2}	p_{m3}	.....	p_{mn}	$f(x_m)$
$g(y_i)$	$g(y_1)$	$g(y_2)$	$g(y_3)$	.....	$g(y_n)$	1

Marginal probability distributions:

In the joint probability table $f(x_1), f(x_2), f(x_3) ...... f(x_m)$ and $g(y_1)$, $g(y_2)$, $g(y_3)$ ...... $g(y_n)$ are called the marginal probability distributions respectively and represents the sum of all entries in all the rows and columns.

Independent random variables:

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

$P(X = x_i, Y = y_j) = P(X = x_i).P(Y = y_j)$ for every i, j and it is equivalent to

$$P(X = x_i, Y = y_j) = P(X = x_i).P(Y = y_j) = f(x_i).g(y_j)$$

or COV(X,Y)=0

Expectation , Variance & Covariance:

Let X be the random variable taking the random values $x_1, x_2, x_3, ...x_m$, having the probability function $f(x)$.

Then

a)The expectation of X is denoted by E(X) and is defined is $\mu_X = E(X) = \sum_i x_i f(x_i)$

b)The expectation of Y is denoted by E(Y) and is defined is $\mu_Y = E(Y) = \sum_j y_j g(y_j)$

c)The variance of X is denoted by $\sigma_X{}^2$ and is defined by $\sigma_X{}^2 = E(X^2) - [E(X)]^2$

$$\Rightarrow \sigma_X{}^2 = \sum_i x_i{}^2 f(x_i) - \mu_X{}^2$$

d)The variance of Y is denoted by $\sigma_Y{}^2$ and is defined by $\sigma_Y{}^2 = E(X^2) - [E(X)]^2$

$$\Rightarrow \sigma_Y{}^2 = \sum_i x_i{}^2 f(x_i) - \mu_X{}^2$$

e)The covariance of X and Y is defined as $COV(X,Y) = E(XY) - E(X).E(Y)$

$$COV(X,Y) = \sum_i \sum_j x_i y_j f(x_i, y_j) - \mu_X.\mu_Y$$

f) The correlation between X and Y is $\rho(X,Y) = \dfrac{COV(X,Y)}{\sigma_X \sigma_Y}$

PROBLEMS:

1) The joint distribution of two random variables X and Y is as follows. ***

X \ Y	-4	2	7
1	$1/8$	$1/4$	$1/8$
5	$1/4$	$1/8$	$1/8$

Compute the following.

i) E(X) and E(Y)

ii) E(XY)

iii) $\sigma_X \& \sigma_Y$

iv) $\rho(X,Y)$

Sol.

Given

$x_1 = 1, x_2 = 5$

$y_1 = -4, y_2 = 2, y_3 = 7$

And the probabilities are

$p_{11} = \frac{1}{8}, p_{12} = \frac{1}{4}, p_{13} = \frac{1}{8}$

$p_{21} = \frac{1}{4}, p_{22} = \frac{1}{8}, p_{23} = \frac{1}{8}$

Given the joint probability distribution is follows as

X \ Y	-4	2	7	$f(x_i)$
1	$1/8$	$1/4$	$1/8$	$1/2$
5	$1/4$	$1/8$	$1/8$	$1/2$
$g(y_i)$	$3/8$	$3/8$	$1/4$	1

The marginal distribution of X and Y are

x_i	1	5
$f(x_i)$	$1/2$	$1/2$

y_i	-4	2	7
$g(y_i)$	$3/8$	$3/8$	$1/4$

i) $\mu_X = E(X) = \sum_i x_i f(x_i) = \left(1 \times \frac{1}{2}\right) + \left(5 \times \frac{1}{2}\right) = 3$

$\mu_Y = E(Y) = \sum_j y_j g(y_j) = \left(-4 \times \frac{3}{8}\right) + \left(2 \times \frac{3}{8}\right) + \left(7 \times \frac{1}{4}\right) = 1$

$E(XY) = \sum_i \sum_j x_i y_j f(x_i, y_j)$

ii) $\Rightarrow E(XY) = \left(1 \times (-4) \times \frac{1}{8}\right) + \left(1 \times 2 \times \frac{1}{4}\right) + \left(1 \times 7 \times \frac{1}{8}\right) + \left(5 \times (-4) \times \frac{1}{4}\right) + \left(5 \times 2 \times \frac{1}{8}\right) + \left(5 \times 7 \times \frac{1}{8}\right)$

$\Rightarrow E(XY) = \frac{3}{2}$

iii) $\sigma_X{}^2 = E(X^2) - \mu_X{}^2$

$$\Rightarrow \sigma_X{}^2 = \sum_i x_i{}^2 f(x_i) - \mu_X{}^2$$

$$\Rightarrow \sigma_X{}^2 = \left(1^2 \times \frac{1}{2}\right) + \left(5^2 \times \frac{1}{2}\right) - 9 = 13 - 9 = 4 \Rightarrow \sigma_X = 2$$

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

$$\Rightarrow \sigma_Y{}^2 = \sum_j y_j{}^2 g(y_j) - \mu_Y{}^2$$

$$\Rightarrow \sigma_X{}^2 = \left((-4)^2 \times \frac{3}{8}\right) + \left(2^2 \times \frac{3}{8}\right) + \left(7^2 \times \frac{1}{4}\right) - 1^2 \Rightarrow \sigma_Y{}^2 = \frac{75}{4} \Rightarrow \sigma_Y = 4.33$$

iv) $COV(X,Y) = E(XY) - \mu_X \mu_Y$

$$\Rightarrow COV(X,Y) = \frac{3}{2} - (3)(1) = -\frac{3}{2}$$

v) $\rho(X,Y) = \dfrac{COV(X,Y)}{\sigma_X \sigma_Y}$

$$\Rightarrow \rho(X,Y) = \frac{-3/2}{2 \times 4.33} = -0.1732$$

2) Determine (i) marginal distribution (ii) covariance between the discrete random variables *X* and *Y*, using the joint probability distribution:

X \ Y	3	4	5
2	$1/6$	$1/6$	$1/6$
5	$1/12$	$1/12$	$1/12$
7	$1/12$	$1/12$	$1/12$

Sol.

Given

$$x_1 = 2, x_2 = 5, x_3 = 7$$

$$y_1 = 3, y_2 = 4, y_3 = 5$$

And the probabilities are

$$p_{11}=\frac{1}{6},\ p_{12}=\frac{1}{6},\ p_{13}=\frac{1}{6}$$
$$p_{21}=\frac{1}{12},\ p_{22}=\frac{1}{12},\ p_{23}=\frac{1}{12}$$
$$p_{31}=\frac{1}{12},\ p_{32}=\frac{1}{12},\ p_{33}=\frac{1}{12}$$

The joint distribution table is as follows

X \ Y	3	4	5	$f(x_i)$
2	$1/6$	$1/6$	$1/6$	$1/2$
5	$1/12$	$1/12$	$1/12$	$1/4$
7	$1/12$	$1/12$	$1/12$	$1/4$
$g(y_i)$	$1/3$	$1/3$	$1/3$	1

The marginal distributions of X and Y are

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

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

$$\mu_X = E(X)=\sum_i x_i f(x_i)$$
$$\Rightarrow \mu_X=\left(2\times\frac{1}{2}\right)+\left(5\times\frac{1}{4}\right)+\left(7\times\frac{1}{4}\right)$$
$$\Rightarrow \mu_X = 4$$

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

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

$$\Rightarrow \mu_Y = 4$$

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

$$\Rightarrow E(XY) = \left(2\times3\times\frac{1}{6}\right)+\left(2\times4\times\frac{1}{6}\right)+\left(2\times5\times\frac{1}{6}\right)+\left(5\times3\times\frac{1}{12}\right)+\left(5\times4\times\frac{1}{12}\right)+\left(5\times5\times\frac{1}{12}\right)$$

$$+\left(7\times3\times\frac{1}{12}\right)+\left(7\times4\times\frac{1}{12}\right)+\left(7\times5\times\frac{1}{12}\right)=16$$

$$\therefore Cov(X,Y) = E(XY) - \mu_X \mu_Y$$

$$\Rightarrow Cov(X,Y) = 16 - 4\times4$$

$$\Rightarrow Cov(X,Y) = 0$$

3) The joint probability distribution of desecrate random variables X and Y is given below:

X \ Y	1	3	6
1	$1/9$	$1/6$	$1/18$
3	$1/6$	$1/4$	$1/12$
6	$1/18$	$1/12$	$1/36$

Determine (i) marginal distribution of X and Y (ii) Are X and Y statistically independent?

Sol.

Given

$$x_1 = 1, x_2 = 3, x_3 = 6$$

$$y_1 = 1, y_2 = 3, y_3 = 6$$

And the probabilities are

$$p_{11} = \frac{1}{9}, p_{12} = \frac{1}{6}, p_{13} = \frac{1}{18}$$

$$p_{21} = \frac{1}{6}, p_{22} = \frac{1}{4}, p_{23} = \frac{1}{12}$$

$$p_{31} = \frac{1}{18}, p_{32} = \frac{1}{12}, p_{33} = \frac{1}{36}$$

The joint distribution table is as follows

X \ Y	1	3	6	$f(x_i)$
1	$1/9$	$1/6$	$1/18$	$1/3$
3	$1/6$	$1/4$	$1/12$	$1/2$
6	$1/18$	$1/12$	$1/36$	$3/18$
$g(y_i)$	$1/3$	$1/2$	$3/18$	1

The marginal distributions of X and Y are

x_i	1	3	6
$f(x_i)$	$1/3$	$1/2$	$3/18$

y_i	1	3	6
$g(y_i)$	$1/3$	$1/2$	$3/18$

$$P(X=x_1,Y=y_1)=\frac{1}{9}\ ,\ P(X=x_1)\times P(Y=y_1)=\frac{1}{3}\times\frac{1}{3}=\frac{1}{9}$$

$P(X=x_1,Y=y_1)=P(X=x_1)\times P(Y=y_1)$ and similarly

$$P(X=x_i,Y=y_j)=P(X=x_i)\times P(Y=y_j)$$

$\therefore$ *X* and *Y* statistically independent

4) Determine (i) marginal distribution (ii) covariance between the discrete random variables *X* and *Y* along with correlation using the joint probability distribution:

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

Sol.

Given

$$x_1=2, x_2=4, x_3=6$$
$$y_1=1, y_2=3, y_3=9$$

And the probabilities are

$$p_{11}=\frac{1}{8}, p_{12}=\frac{1}{24}, p_{13}=\frac{1}{12}$$
$$p_{21}=\frac{1}{4}, p_{22}=\frac{1}{4}, p_{23}=0$$
$$p_{31}=\frac{1}{8}, p_{32}=\frac{1}{24}, p_{33}=\frac{1}{12}$$

The joint distribution table is as follows

X \ Y	1	3	9	$f(x_i)$
2	$1/8$	$1/24$	$1/12$	$1/4$
4	$1/4$	$1/4$	0	$1/2$
6	$1/8$	$1/24$	$1/12$	$1/4$
$g(y_i)$	$1/2$	$1/3$	$1/6$	1

The marginal distributions of X and Y are

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

y_i	3	4	5
$g(y_i)$	$1/2$	$1/3$	$1/6$

$$\mu_X = E(X) = \sum_i x_i f(x_i)$$
$$\Rightarrow \mu_X = \left(2\times\frac{1}{4}\right)+\left(4\times\frac{1}{2}\right)+\left(6\times\frac{1}{4}\right)$$
$$\Rightarrow \mu_X = 4$$
$$\mu_Y = E(Y) = \sum_j y_j g(y_j)$$
$$\Rightarrow \mu_Y = \left(1\times\frac{1}{2}\right)+\left(3\times\frac{1}{3}\right)+\left(9\times\frac{1}{6}\right)$$
$$\Rightarrow \mu_Y = 3$$

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

$$\Rightarrow E(XY) = \left(2 \times 1 \times \frac{1}{8}\right) + \left(2 \times 3 \times \frac{1}{24}\right) + \left(2 \times 9 \times \frac{1}{12}\right) + \left(4 \times 1 \times \frac{1}{4}\right) + \left(4 \times 3 \times \frac{1}{4}\right) + (4 \times 9 \times 0)$$

$$+ \left(6 \times 1 \times \frac{1}{8}\right) + \left(6 \times 3 \times \frac{1}{24}\right) + \left(6 \times 9 \times \frac{1}{12}\right) = 12$$

$$\therefore Cov(X,Y) = E(XY) - \mu_X \mu_Y$$

$$\Rightarrow Cov(X,Y) = 12 - 4 \times 3$$

$$\Rightarrow Cov(X,Y) = 0$$

The correlation $\rho(X,Y) = \dfrac{COV(X,Y)}{\sigma_X \sigma_Y} = 0$

5) Determine (i) marginal distribution (ii) covariance between the discrete random variables X and Y along with correlation using the joint probability distribution: (Home Work)

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

SAMPLING THEORY

Population:

A large collection of individuals or attributes or numerical data can be understand as population or universe.

Sample:

A finite subset of the universe is called a sample.

Sample size:

The number of individuals in a sample is called a sample size .If the sample size n is less than or equal to 30, then the sample is aid to be small , otherwise it is called a large sample.

Sampling:

The process of selecting a sample from the population is called as sampling.

Sampling with & without replacement:

Sampling where a number of the population may be selected more than once is called as sampling with replacement. If a number cannot be chosen more than once is called as sampling without replacement.

Standard Error:

The standard deviation of a sampling distribution is called the Standard Error(SE).s The reciprocal of the standard error is called Precision..

Null hypothesis :

The **null hypothesis** is a general statement or default position that there is no relationship between two measured phenomena or no association among groups.

Example : Given the test scores of two random samples, one of men and one of women, does one group differ from the other? A possible null hypothesis is that the mean male score is the same as the mean female score:

H_0: $\mu_1 = \mu_2$

where

H_0 = the null hypothesis,

μ_1 = the mean of population 1, and

μ_2 = the mean of population 2.

A stronger null hypothesis is that the two samples are drawn from the same population, such that the variances and shapes of the distributions are also equal.

Confidence intervals:

A confidence interval, in statistics, refers to the probability that a population parameter will fall between two set values for a certain proportion of times. Confidence intervals measure the degree of uncertainty or certainty in a sampling method. A confidence interval can take any number of probabilities, with the most common being a 95% or 99% confidence level.

Type I error

The first kind of error is the rejection of a true null hypothesis as the result of a test procedure. This kind of error is called a type I error and is sometimes called an error of the first kind .In terms of the courtroom example, a type I error corresponds to convicting an innocent defendant.

Type II error

The second kind of error is the failure to reject a false null hypothesis as the result of a test procedure. This sort of error is called a type II error and is also referred to as an error of the second kind.

Test of hypothesis :

Let x be the observed number of successes in a sample size of n and $\mu = np$ be the expected number of successes .Then the standard normal variate Z is defined as

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

Test of hypothesis for means:

Let μ_1 , μ_2 be the means , σ_1, σ_2 be the standard deviations of two populations and $\bar{x}_1, \bar{x}_2$ are the means of the samples , then

$Z = \dfrac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\dfrac{\sigma_1^{\ 2}}{n_1} + \dfrac{\sigma_2^{\ 2}}{n_2}}}$ If the samples are drawn from the same population ,

then $\sigma_1 = \sigma_2 = \sigma$ we have $Z = \dfrac{(\bar{x}_1 - \bar{x}_2)}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

Student's t-distribution:

Let μ be the mean of population , $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ be the mean and $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ be the standard deviation of a sample , then the Student's t-distribution is defined as

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

Another formula for t - test of two samples is

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

Where $s^2 = \dfrac{n_1\sigma_1^{\ 2} + n_2\sigma_2^{\ 2}}{n_1 + n_2 - 2}$

Chi-square distribution:

Let $O_i\,(i = 1,2,3 \ldots n)$ and $E_i\,(i = 1,2,3 \ldots n)$ be the set of observed frequencies and expected frequencies respectively , then the Chi-square distribution is defined as

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} + \ldots\ldots + \frac{(O_n - E_n)^2}{E_n}$$

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

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

Sol.

Let us suppose that the coin is unbiased .

and let p= the probability of getting a head in one toss=1/2=0.5

Since p+q=1, q=1-p=1/2=0.5

Expected number of heads in 1000 tosses=np=1000x0.5=500 , npq=250

$\therefore$ The difference is $x-\mu$=540-500=40

$\therefore$ Consider
$$Z=\frac{x-\mu}{\sigma}=\frac{x-np}{\sqrt{npq}}$$
$$\Rightarrow Z=\frac{40}{\sqrt{250}}=2.53<2.58$$

Thus we can say that the coin is unbiased

2) In324 throws of a six faced 'die' , an odd number turned up 181 times. Is it possible to think that the 'die' is an unbiased one?

Sol.

Let us suppose that the die is unbiased .

and let p= the probability of the turn up of an odd number is=3/6=1/2=0.5

Since p+q=1, q=1-p=1/2=0.5

Expected number of successes=np=324x0.5=162 , npq=81

$\therefore$ The difference is $x-\mu$=181-162=19

$\therefore$ Consider
$$Z=\frac{x-\mu}{\sigma}=\frac{x-np}{\sqrt{npq}}$$
$$\Rightarrow Z=\frac{19}{\sqrt{81}}=\frac{19}{9}2.11<2.58$$

Thus we can that the die is unbiased

3) A die is thrown 9000 times and a throw of 3 or 4 was observed 3240 times. Show that the die cannot be regarded as an unbiased one.

Sol.

The probability of getting 3 or 4 in a single through is $p={}^{2}\!/_{6}={}^{1}\!/_{3}$

And $q=1-p=1-\frac{1}{3}=\frac{2}{3}$

$\therefore$ Expected number of success $=\frac{1}{3}\times 9000=3000$

$\therefore$ The difference =3240-3000=240

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

$$\Rightarrow Z = \frac{(3240) - \left(9000 \times \frac{1}{3}\right)}{\sqrt{9000 \times \frac{1}{3} \times \frac{2}{3}}}$$

Consider $\Rightarrow Z = \frac{240}{\sqrt{2000}}$

$$\Rightarrow Z = 5.37$$

Since Z=5.37>2.58 , We conclude that the die is biased.

4) A manufacture claimed that at least 95%of the equipment which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 of them were faulty. Test his claim at a significance level of 1% and 5%.

Sol.

Let p be the probability of success which being the probability of the equipment supplied to the factory conformal to the specifications.

$\therefore$ By the data $p = 0.95$ and hence $q = 0.05$

$H_0 : p = 0.95$ and the claim is correct

$H_1 : p < 0.95$ and the claim is false.

We choose the one tailed test to determine whether the supply is conformal to the specification.

$$\therefore \mu = np = 200 \times 0.95 = 190$$

$$\sigma = \sqrt{npq} = \sqrt{200 \times 0.95 \times 0.05} = 3.082$$

Expected number of equipments according to the specification=190
Actual number=182 , since 18 out of 200 were faulty.

$\therefore$ Difference=190-182=8

Now

$$Z = \frac{x - np}{\sqrt{npq}} = \frac{8}{3.082} = 2.6$$

The value of Z is greater than the critical value 1.645 at 5% level and 2.33at 1% level of significance.

The claim of the manufacturer is rejected at 5% as well as 1% level of significance in the accordance with the one tailed test.

5) A certain stimulus administered to each of the 12 patients resulted in the following change in the blood pressure 5,3,8,-1,3,0,6,-2,1,5,0,4. Can it be concluded that the stimulus will increase the blood pressure ? (*Note* : t0.05 for 11 d.f. is 2.201).

Sol.

Given the change in blood pressure

x :5,3,8,-1,3,0,6,-2,1,5,0,4

$$\therefore \bar{x} = \frac{1}{n}\sum x = \frac{31}{12} = 2.5833$$

Variance

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

$$\Rightarrow s^2 = \frac{1}{11}\begin{Bmatrix} (5-2.58)^2 + (3-2.58)^2 + (8-2.58)^2 + (-1-2.58)^2 + (0-2.58)^2 + (6-2.58)^2 \\ + (-2-2.58)^2 + (1-2.58)^2 + (5-2.58)^2 + (0-2.58)^2 + (4-2.58)^2 \end{Bmatrix}$$

$$\Rightarrow s^2 = 9.538 \Rightarrow s = 3.088$$

Let us suppose that the stimulus administration is not accompanied with increase in blood pressure , we can take $\mu = 0$

we have

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

$$\Rightarrow t = \frac{2.5833 - 0}{\left(\frac{3.088}{\sqrt{12}}\right)}$$

$$\Rightarrow t = 2.8979 \approx 2.9 > 2.201$$

Hence the hypothesis is rejected at 5%level of significance .We conclude with 95% confidence that the stimulus in general is accompanied with increase of blood pressure.

6) A random sample of 10 boys had the following I.Q. : 70,120,110,101,88,83,95,98,107,100. Does this data support the assumption of a population mean I.Q. of 100 at 5% level of significance? (*Note*: $t_{0.05}$ =2.262 for 9 d.f.).

Sol.

Given the I.Q. of 10 boys

x :70,120,110,101,88,83,95,98,107,100

$$\therefore \bar{x} = \frac{1}{n}\sum x = \frac{972}{10} = 97.2$$

Variance

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

$$\Rightarrow s^2 = \frac{1}{9} \times 1833.6$$

$$\Rightarrow s^2 = 203.73333$$

$$\Rightarrow s = 14.2735$$

Given the mean of population $\mu = 100$

We have

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

$$\Rightarrow t = \frac{97.2 - 100}{\left(\frac{14.2735}{\sqrt{10}}\right)}$$

$$\Rightarrow t = \frac{-2.8}{4.5136} \approx -0.6203 < 2.262$$

7) Ten individuals are chosen 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 66inches ($t_{0.05}$ =2.262 for 9 d.f.)

Sol.

Given the heights of the population in inches

x:63, 63,66,67,68,69,70,70,71,71

$$\therefore \bar{x} = \frac{1}{n}\sum x = \frac{678}{10} = 67.8$$

Variance

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

$$\Rightarrow s^2 = \frac{1}{9}\begin{Bmatrix} (63-67.8)^2 + (63-67.8)^2 + (66-67.8)^2 + (67-67.8)^2 + (68-67.8)^2 + (69-67.8)^2 \\ + (70-67.8)^2 + (70-67.8)^2 + (71-67.8)^2 + (71-67.8)^2 \end{Bmatrix}$$

$$\Rightarrow s^2 = 9.067 \Rightarrow s = 3.011$$

And given the mean of population $\mu = 66$

we have

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

$$\Rightarrow t = \frac{67.8 - 66}{\left(\frac{3.011}{0}\right)}$$

$$\Rightarrow t = 2.8979 \approx 1.89 > 2.262$$

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

8) Two types of batteries are tested for their length of life and the following results are obtained:

Battery A: $n_1 = 10, \bar{x}_1 = 500hrs., \sigma_1{}^2 = 100$

Battery A: $n_2 = 10, \bar{x}_2 = 560hrs., \sigma_2{}^2 = 121$

Compute Student's t and test whether there is a significant difference in the two means.

Sol.

Given

Battery A: $n_1 = 10, \bar{x}_1 = 500hrs., \sigma_1{}^2 = 100$

Battery A: $n_2 = 10, \bar{x}_2 = 500hrs., \sigma_2{}^2 = 121$

We know that

$$s^2 = \frac{n_1\sigma_1{}^2 + n_2\sigma_2{}^2}{n_1 + n_2 - 2}$$

$$\Rightarrow s^2 = \frac{(10\times100) + (10\times121)}{10+10-2}$$

$$\Rightarrow s^2 = 122.78$$

$$\Rightarrow s = 11.0805$$

we have

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

$$\Rightarrow t = \frac{560 - 500}{11.0805\sqrt{0.1+0.1}}$$

$$\Rightarrow t = 12.1081 \approx 12.11$$

The value of t is greater than the table value of t for 18d.f.at all levels of significance.

9) Fit a binomial distribution for the data and test the goodness of fit given that $\chi 0_{.05}{}^2$= 7.815 for 3 d.f.

No. of Heads	0	1	2	3	4
Frequency	122	60	15	2	1

Sol.

Given

Observed frequencies are 122,60,15,2,1

The expected frequencies by using Binomial distribution are 121,61,15,3,0

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

$$\therefore \chi^2 = \sum_i \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

$$\Rightarrow \chi^2 = \frac{1}{121} + \frac{1}{61} + 0 + 0$$

$$\Rightarrow \chi^2 = 0.025 < \chi^2_{0.05} = 7.815$$

Hence the fitness is good.

10) Four coins are tossed 100 times and the following results were obtained:

No. of Heads	0	1	2	3	4
Frequency	5	29	36	25	5

Fit a binomial distribution for the data and test the goodness of fit $\chi 0_{.05}{}^2$ = 9.49 for 4 d.f.

Sol.

Given the 4 coins are tossed 100 times

The probability of getting head is p=0.5, q=0.5

The probability mass function of a binomial distribution is

$$P(X = x) = 4_{Cx}(0.5)^x (0.5)^{4-x}$$

$$P(1) = 4_{C1}(0.5)^1 (0.5)^{4-1} = 0.25$$

$$P(2) = 4_{C2}(0.5)^2 (0.5)^{4-2} = 0.375$$

$$P(3) = 4_{C3}(0.5)^3 (0.5)^{4-3} = 0.25$$

$$P(4) = 4_{C4}(0.5)^4 (0.5)^{4-4} = 0.0625$$

$$\therefore E_0 = 100 \times 0.0625 = 6.25$$

$$E_1 = 100 \times 0.25 = 25$$

$$E_2 = 100 \times 0.375 = 37.5$$ Where 100 is the sum of frequency

$$E_3 = 100 \times 0.25 = 25$$

$$E_4 = 100 \times 0.0625 = 6.25$$

O_i	5	29	36	25	5
E_i	6.25	25	37.5	25	6.25

$$\therefore \chi^2 = \sum_i \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

$$\Rightarrow \chi^2 = \frac{1.5625}{6.25} + \frac{16}{25} + \frac{2.25}{37.5} + 0 + \frac{1.5625}{6.25}$$

$$\Rightarrow \chi^2 = 0.25 + 0.64 + 0.06 + 0.25$$

$$\Rightarrow \chi^2 = 1.2 < \chi^2_{0.05} = 9.49$$

Hence the fitness is good.

***

Thank you

By:

Purushotham.P

Sampling Theory & Stochastic process

Stochastic process

Probability vector

A vector whose component are non-negative and their sums are equal to 1 $\quad v_i \geq 0 \;\&\; \sum_{i=1}^{n} v_i = 1$

Stochastic Matrix

A square matrix P is called stochastic matrix if all the entries of the P are non-negative & sum of the entries of any row is 1

Ex :- $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1/3 & 2/3 \\ 0 & 1 \end{bmatrix}$

Ex: $[1/4, 1/4, 1/2]$.

Fixed vector or fixed Point :-

A vector 'V' is said to be fixed vector or a fixed point of a matrix 'A' if

$$\boxed{VA = V}$$

Regular Stochastic matrix

A stochastic matrix P said to be regular if all the entries of sum of P^m are +ve [positive]

$P \Rightarrow$ sum of P^m are +ve.

$A = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix} \quad A^2 = A \cdot A = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$

Note 1. Let 'P' be a regular stocastic matrix

i) P has a unique fixed ~~vector~~ probability vector.

ii) P is a stocastic matrix

2. A stochastic matrix is not regular if 1 occurs in the ~~principle~~ main diagonal.

Ex:- $\begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}$ not regular.

Examples

1. Find unique Fixed probability vector of Regular stochastic matrix $P = \begin{bmatrix} 1/4 & 3/4 \\ 1/4 & 1/2 \end{bmatrix}$

Sol

Let $V = (x, y)$ be a unique fixed probability vector associated with P

To Prove $\boxed{VA = V}$ ie $x + y = 1$

W.K.T by the property of regular Stochastic matrix
ie $VA = V$

$\Rightarrow (x, y) \begin{bmatrix} 1/4 & 3/4 \\ 1/2 & 1/2 \end{bmatrix} = (x, y)$.

$= \left[\frac{x}{4} + y/2, \; 3x/4 + y/2\right] = (x, y)$.

$\cancel{\frac{x+y}{4}}$

$\Rightarrow \frac{x}{4} + y/2 = x$ & $\frac{3x}{4} + y/2 = y$.

$\Rightarrow y/2 = x - x/4$ & $\Rightarrow y = 3x/2$.

$y/2 = \frac{3x}{4}$.

∴ substitute $y = 3x/2$ to get x

$\cancel{\frac{3x}{4}} + \cancel{\frac{3x}{4}}$

$\cancel{\frac{3x}{4} + \frac{3x}{4} = 3x/2}$.

$\cancel{\frac{3x}{2} = 3x/2}$.

$x+y=1 \Rightarrow x+3x/2=1$;

$\Rightarrow \frac{5x}{2} = 1 \Rightarrow x = 2/5$.

$y = 3x/2 \Rightarrow y = 3/2 \cdot 2/5 = 3/5$.

$V(x,y) = V(2/5, 3/5)$.

Hence unique probability vector is $2/5$ & $3/5$

2. Find the unique fixed probability ~~wh~~ vector of Regular Stochastic matrix

$$A = \begin{bmatrix} 0 & 3/4 & 1/4 \\ 1/2 & 1/2 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Let $v = (x, y, z)$ be unique fixed prob. vector with A

A ie $x+y+z=1$.

W.K.T by the prob of regular stochastic matrix

ie $VA = V$.

$$(x,y,z)\begin{bmatrix} 0 & 3/4 & 1/4 \\ 1/2 & 1/2 & 0 \\ 0 & 1 & 0 \end{bmatrix} = (x,y,z).$$

$$\left[\tfrac{1}{2}y,\ \tfrac{3}{4}x + \tfrac{1}{6}y + z,\ \tfrac{1}{4}x\right] = (x, y, z)$$

$\Rightarrow x = y/2$, $y = 3/4 x + 1/6 y + z$, $z = 1/4 x$

$x+y+z=1$

$x + 2x + x/4 = 1$

$\frac{13}{4} x = 1$

$\therefore x = 4/13$.

$\therefore y = 2x = 8/13$.

$lly \quad z = 1/4 \times 4/13 = 1/13$

$\Rightarrow V = [4/13, 8/13, 1/13]$

H.W

3. $A = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$

4. $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.3 & 0.5 & 0 \end{bmatrix}$

$V = (x, y, z)$

$x + y + z = 1$

To prove.

$VA = V$

$(x, y, z)\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.5 & 0.5 & 0 \end{bmatrix} = (x, y, z)$

$[z(0.5), \; x + z(0.5), \; y] = (x, y, z)$

$z(0.5) = x, \quad x + 0.5z = y, \quad y = z.$

$x + y + z = 1$

$\frac{1}{2}z + z + z = 1 \qquad = \frac{5}{2}z = 1$

$z = 2/5, \quad y = 2/5, \quad x = \frac{2}{5} \times \frac{1}{2} = 1/5.$

$V = [1/5, 2/5, 2/5].$

$$A = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$$

$V = (x, y, z)$

To prove $VA = V$.

$$(x, y, z)\begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix} = (x, y, z).$$

$$\left[x/2 + y/2,\ x/4 + z,\ x/4 + y/2\right] = (x, y, z).$$

$x/2 + y/2 = x$, $x/4 + z = y$, $x/4 + y/2 = z$.

$x + y + z = 1$.

$1/2\, y = 1/2\, x$ $\qquad$ $1/4\, x + 1/2\, y = z$.

$3/4\, x = z$.

$x + y + z = 1$

$x + x + 3/4\, x = 1$

$2x + 3/4\, x = 1$

$11/4\, x = 1$

$x = 4/11$

Transition probability matrix or Stochastic matrix

The probabilities of moving from one state to another state or ~~tm~~ remaining in the same state are called transition probability ~~matrix~~.

Transition probability from a square matrix is called ∴ Transition probability matrix or Stochastic matrix.

~~Markov chain~~

$$P = [P_{ij}] = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix}$$

Markov chain

The stochastic process which is such that the generation of p the probability distribution depends only on the present state is called Markov chain.

Then entry P_{ij} in the transition probability matrix P of the markov chain is the probability that the system changes from the state a_i to a_j in a single step.

$$p^{(1)} = p^{(0)}, p^{(1)}$$
$$p^{(2)} = p^{(0)}\ p^{(2)}$$

Note

Markov chain is irrecible if the associated transition probability matrix is regular

explain the following

1. Absorbing state :- In a Markov Chain, the process Reaches to a certain state after which it continous to remain in the same State such a State is called as Absorbing state.

Ex P.T.O

Ex:- $\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & \phi & 0 \\ 1/2 & 1/2 & 1/2 \end{bmatrix}$

2. Transient State :-

A state i said to be transiend iff if there is a positive prob that the process will not return to this state

Ex If we model a program as a Markov chain, then all except final step of the program are transient state.

3. Recurrent State :-

A state i is said to be Recurrent state iff starting from the state i, the process eventually return to state i with the probability 1

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \longrightarrow P^3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

The system remains in the same state after two steps.

4. Periodic state :- The recurrente state i is said to be periodic state if $GCD(d_i > 1)$

ex Identity matrix

Problems

1. Define regular Stochastic matrix & show that

$$A = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$ is irreducible [Regular] matrix

Ans A stochastic matrix P said to be regular if all the entries of sum of P^m are +ve

$P \Rightarrow$ sum of P^m are +ve.

$$A = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

$$A^2 = \begin{bmatrix} 0.5 & 1/6 & 1/3 \\ 1/4 & 7/12 & 1/6 \\ 1/4 & 1/3 & 5/12 \end{bmatrix}$$

$\therefore$ It is Regular.

2. Define a stochastic matrix and show that

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/4 & 1/4 \end{bmatrix}$$ is a regular stochastic matrix

Ans. ~~st~~ ~~st~~ Defination is written Backside.

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/4 & 1/4 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 1/4 & 5/8 & 1/8 \\ 1/4 & 9/16 & 3/16 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 1/4 & 5/8 & 1/8 \\ 3/8 & 9/32 & 11/32 \\ 3/8 & 19/64 & 21/64 \end{bmatrix}$$

3. A student's study habits are as follows
If he studies One night, he is 70% sure Not ~~to~~ study the next night, One the other hand if he does not study One night, he is 60% sure not to study the next night.
In the Long Run how often does he study?

sol

Let $S = \{ \text{Studying}(A) , \text{ not studying } (B) \}$

$$\Rightarrow \quad P = \begin{bmatrix} 0.3 & 0.7 \\ 0.4 & 0.6 \end{bmatrix}$$

we have to find unique fixed probability vector

Let 'v' be a unique fixed prob vector $v = (x, y)$

$\Rightarrow \quad x + y = 1 \longrightarrow (1)$

$\because$ from stochastic matrix we've

$$vP = v$$

$$(x, y) \begin{bmatrix} 0.3 & 0.7 \\ 0.4 & 0.6 \end{bmatrix} = (x, y)$$

$$\begin{bmatrix} 0.3x + 0.4y , & 0.7x + 0.6y \end{bmatrix} = \begin{bmatrix} x , & y \end{bmatrix}$$

$$0.3x + 0.4y = x , \quad 0.7x + 0.6y = y .$$

$$3x + 4y = 10x .$$

$$4y = 7x .$$

$$\boxed{y = \frac{7}{4} x}$$

sub $y = \frac{7x}{4}$

$$x + y = 1$$

$$x + \frac{7x}{4} = 1$$

$$11x = 4$$

$$x = 4/11 \Rightarrow y = \frac{7x}{4} \Rightarrow 7/11$$

$$x = 4/11 , \quad y = 7/11$$

Hence require fixed probability is

$$v = \begin{bmatrix} 4/11 , & 7/11 \end{bmatrix}$$

Oftenly he does study $4/11$ or $36.36\ \%$.

4. A man's smoking habits are as follows.
If he smokes filter cigarettes one week he switches to non-filter cigerates the next week with probability 0.2. If he smokes ~~non~~ non-filter cigarettes one week, there is a probability of 0.7 that he will smoke non-filter cigarette the next week as well. In the long run how often does he smoke filter cigarette?

Let $S = \{\text{filter cig } (A), \text{ Non-filter cig } (B)\}$

$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$$

$$x + y = 1 \longrightarrow (1)$$

$$VP = V$$

$$(x, y)\begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix} = (x, y)$$

$$[0.8x + 0.3y, \quad 0.2x + 0.7y] = (x, y)$$

$$0.8x + 0.3y = x, \quad 0.2x + 0.7y = y$$

$$8x + 3y = 10x$$

$$3y = 2x$$

$$y = \frac{2x}{3}$$

$$x + y = 1$$

$$x + \frac{2x}{3} = 1$$

$$\frac{3x + 2x}{3} = 1$$

$$3x + 2x = 3 \implies y = \frac{2 \times 3}{3 \times 5} = 2/5$$

$$5x = 3$$

$$x = 3/5$$

$$V = [3/5, \ 2/5]$$

15/6/17

5. A habitant gambler is a member of two clubs A & B. He visits either of the clubs everyday for playing cards. He never visits club A on two consecutive days. But if he visits club B on a particular day, then the next day, he is as likely to visit club B or club A

i) After long run how often does he visit club A

ii) If the person has visited club B on monday, find the probability that he visit club A on tuesday.

Sol^n

$$P = \begin{array}{c|cc} & A & B \\ \hline A & 0 & 1 \\ B & 1/2 & 1/2 \end{array}$$

i) $x + y = 1$

$VP = V$

$\Rightarrow [x, y] \cdot [\quad] = [x, y]$

ii) If the person visited club B on Monday, i.e Day 1 the he visit club A on thursday day 3

$$P^3 = \begin{bmatrix} 0.25 & 0.75 \\ 0.375 & 0.675 \end{bmatrix}$$

$= 0.375$

6) Every year a man trades his car for a new car. If he has Maruthi he trades it for an Ambassdor. If he has an Ambassador he trades it for Santro. However if he has a Santro he is just as likely to trade it for a new santro as to trade it for maruthi or an Ambassador. In the year 2000, he bought his first car which was a santro. find the probability that he has

i) 2002 maruthi

ii) 2003 Santro

$\underline{\text{Sol}^n}$

$$P = \begin{array}{c} \\ m \\ n \\ s \end{array}\begin{array}{c} \begin{array}{ccc} M & A & J \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \end{array}$$

After 2 years.

i) $P^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0.33 & 0.33 & 0.33 \\ 0.11 & 0.44 & 0.44 \end{bmatrix}$

$A_{31} = 0.11$

ii) $P^3 = \begin{bmatrix} 0.33 & 0.33 & 0.33 \\ 0.11 & 0.44 & 0.44 \\ 0.148 & 0.259 & 0.592 \end{bmatrix}$

$A_{33} = 0.592$.

7. 3 boys A, B, C are throwing the ball to each other. A always throws a ball to B & B always throws ball to C. But C is just as likely to throw a ball to B as to A. If C was the 1st person to throw the ball, find the ~~to~~ probability that:

i) A has a ball

ii) B has a ball

iii) C has a ball for the 4th throw

$\underline{\text{Sol}^n}$

$$P = \begin{array}{c} \\ A \\ B \\ C \end{array}\begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{bmatrix} \end{array}$$

$P^3 = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$

i) $0.25 = 1/4$

ii) $0.25 = 1/4$

iii) $0.5 = 1/2$

8. 2 Boys B_1, B_2 & 2 girls G_1, G_2 are throwing the ball from one to another. Each boy throws the ball to other boy with the probability 1/2 & to each girl with the probability 1/4 on the other hand throws a ball to each boy with probability 1/2 & never to the other girl. In the long run how often does each receive the ball?

$$P = \begin{array}{c} \\ B_1 \\ B_2 \\ G_1 \\ G_2 \end{array}\begin{array}{c} \begin{array}{cccc} B_1 & B_2 & G_1 & G_2 \end{array} \\ \begin{bmatrix} 0 & 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/4 & 1/4 \\ 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix} \end{array}$$

To find the long run we have to use UFP vector

Let $V = (a, b, c, d)$ - are U.F.P

$\Rightarrow a+b+c+d = 1 \longrightarrow (1)$

From stochastic matrix we have

$VP = V$

$$[a \; b \; c \; d] \begin{bmatrix} 0 & 1/2 & 1/4 & 1/2 \\ 1/2 & 0 & 1/4 & 1/4 \\ 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix} = [a \; b \; c \; d]$$

$a+b+c+d = 1-a$

$\Rightarrow \frac{b}{2} + c/2 + d/2 = a$

$a/2 + c/2 + d/2 = b$

$a/4 + b/4 = c$

$a/4 + b/4 = d$

$\Rightarrow c = d.$

$\Rightarrow b/2 + c/2 + c/2 = a$

$\Rightarrow a/2 + c/2 + c/2 = b$

$\frac{1}{2}[b+c+d] = a$

$\frac{1}{2}[1-a] = a$

$1/2 - a/2 = a$

$1/2 = a + a/2$

$1/2 = 3a/2$

$1/3 = a$

$$\Rightarrow \quad b/2 + c = a$$

$$\begin{array}{l} a/2 + c = b \\ (-) \quad (-) \quad (-) \\ \hline \end{array}$$

$$b/2 - a/2 = a - b$$

$$\frac{b}{2} - a/2 = a - b$$

$$b/2 + b = a + a/2$$

$$\frac{\cancel{3}b}{\cancel{2}} = \cancel{3}/\cancel{2}\, a$$

$$a = b$$

$$a = 1/3, \quad b = 1/3$$

$$c = 1/6, \quad d = 1/6$$

SAMPLING THEORY

A large collection of data of individuals or attributes is known as population or universe.

A small part of the population is called a ~~Simp~~ sample. The process of selecting a sample from the population is called sampling.

There are 2-types of sampling.

1) Random sampling with replacement

Here the items are drawn 1 by 1 and are put back to the population before the next draw. In this method mean $= \mu$ & variance $= \frac{\sigma^2}{n}$.

2) Random sampling without Replacement

Here the items are drawn 1 by 1 but are not put back to the population befor the next draw. In this method

$$\text{Variance} = \left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n}.$$

16/5/17

Testing of Hyphothesis

Hypothesis is a decision making statement which is true or false. There are 2 types of hypothesis

- Null Hypothesis
- Alternate Hypothesis

Null Hypothesis (H_0)

The hypothesis formulated for the purpose of rejection is called Null Hypothesis

denoted by H_0

Alternate Hypothesis (H_1)

Any hypothesis which is not Null or exceptance is called alternate hypothesis denoted by H_1.

Confidence interval :- that acts as good estimates of the unknown population parameter

ERRORS

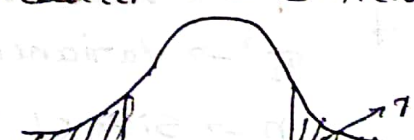
There are 2 types of Errors of testing hypothesis.

1. **Type-I error** :- rejecting the null hypothesis H_0 and accepting the alternate hypothesis H_1 when actually H_0 is true is called type-II error.

2. **Type-I error** :- accepting H_0 and rejecting H_1, actually H_1 is true.

Critical Region.

A region which amounts to the rejection of Null Hypothesis is called the critical region or region of rejection

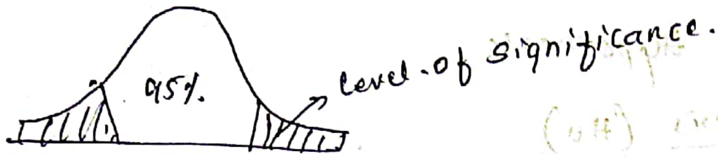


rejection region or Critical region.

Signifance level . or Level of significance [LOS]

The probability level below which leads to regection is called significance level.

→ generally 1% or 5% is the significance level



if we're (5% or 1%). (5% or 1%.)

(−1.96 or −2.58) (1.96 or 2.58)

Note :- critical value of z (two-tailed test)

5% LOS	1% LOS
−1.96 & 1.96	−2.58 & 2.58.

Testing of population mean (μ) is given by

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

where $\bar{x}$ → The sample mean
μ → population mean
σ → S.D.

Testing of propotion (p)

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

np → Expected no of success
x → observed no of success
p → probability of success

$q = 1 - p$

Testing of mean :

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

$\bar{x}_1$ → mean of sample 1
$\bar{x}_2$ → —— " —— 2
σ_1^2 → variance of sample 1
σ_2^2 → variance of sample 2.
n_1 → size of sample 1
n_2 —— " —— 2

Testing of significance for the difference of properties of two samples is given by

$$z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}.$$

5. confidence intervals for μ.

95% C.I $\bar{x} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$. confidence interval.

99% C.I $\bar{x} \pm 2.58\left(\frac{\sigma}{\sqrt{n}}\right)$.

6. confidence intervals for (p)

95% CI $p \pm 1.96\left(\sqrt{\frac{pq}{n}}\right)$.

99% CI $p \pm 2.58\sqrt{\frac{pq}{n}}$.

7. C.I for difference of two means.

95% C.I $(\bar{x}_1 - \bar{x}_2) \pm 1.96\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

99% C.I $(\bar{x}_1 - \bar{x}_2) \pm 2.58\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

C

Example

1. The mean life time of the sample ~~100~~ of 100 fluroscent light bulbs produced by a company is computed to be 1570 hours with SD = 120 hours. If μ is the life time of all bulb's produced by the company, thest the hypothesis $\mu = 1600$ hours against the alternate hypothesis $\mu \neq 1600$ hours. using LOS as 0.01

Ans LOS = 0.01

= 1% = 2.58.

given $n = 100$, $\bar{x} = 1570$, $\mu = 1600$.

$\sigma = 120$

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{1570 - 1600}{\frac{120}{\sqrt{100}}} = -2.5$$

$$|z|_{cal} = 2.5$$

level of significance

$$Z_{0.01} = 2.58$$

$$|z|_{cal} < Z_{0.01}$$

$\therefore H_0$ is accepted

2. ~~Ad~~ A dice was thrown 9000 times and a throw of 5 or 6 was obtained 3240 times. On the assumption of random throwing, Do the data indicate that the data is unbiased?

Ans $n = 9000$, $x = 3240$

$$p = \frac{2}{6} = \frac{1}{3}$$

$$q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$$

$$\therefore Z = \frac{x - np}{\sqrt{npq}} = \frac{3240 - 9000 \times 1/3}{\sqrt{9000 \times 1/3 \times 2/3}}$$

$$= 5.37$$

$$|Z|_{cal} = 5.37$$

1% LOS $Z_{tab} = 2.58$

$Z_{cal} > Z_{tab} \Rightarrow H_0$ is rejected

$\therefore$ the data is unbiased.

3. A dice was thrown ~~4000~~ 1200 time & the no 6 is obtained 236 times, can the dice be fair at 0.01 LOS

Ans $n = 1200$, $x = 236$.

$p = 1/6$

$q = 1 - 1/6 = 5/6$.

$$Z = \frac{x - np}{\sqrt{npq}} = 2$$

$$\frac{236 - 1200 \times 1/6}{\sqrt{1200 \times 1/6 \times 5/6}} = 2.8$$

$|Z|_{cal} = 2.8$

$|Z|_{tab} = 2.58$

$|Z|_{cal} > |Z|_{tab} \Rightarrow H_0$ is unfair [rejected]

4. A coin is tossed 400 times and the head turn up 216 times, test the hypothesis at 5% LOS that the coin is unbiased.

As $n = 400$

$p = 1/2$.

$q = 1 - 1/2 = 1/2$.

5. A Dice is tossed 960 times, it falls with 5 upwards 184 times, it the die biased at 5% of LOS ?

6. T

6. To know the mean weights of all 10 year boys in delhi, a sample of 225 was taken, the mean weight of the sample was ~~po~~ found to be 67 pounds with the S.D 12 pounds. What can we infer about the mean weight of population

A: $n = 225$, $\bar{x} = 67$. $\sigma = 12$.

We've solve this by confidence interval for μ.

→ 99% of CI $\bar{x} \pm 2.58\left(\frac{\sigma}{\sqrt{n}}\right)$.

$= 67 \pm 2.58\left(\frac{12}{\sqrt{225}}\right)$.

$\mu = 69.064, 64.936$.

∴ mean weight lies b/w $64.936 \le \mu \le 69.064$.

→ 95% of CI $\bar{x} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$

$= 67 \pm 1.96\left(\frac{12}{\sqrt{225}}\right)$.

$= 68.56 \quad , \quad 65.432$.

~~$68.56 < \mu < 65.432$~~ .

$\therefore$ The mean weigh lies b/w .

$65.432 < \mu < 68.56$.

~~The mean weight lies b/w~~ .

7. The mean & S.D of maximum load supported by 60 cables are 01.09 tonnes & 0.37 tonnes respectively . Find 99%. confidence limit for the mean of the maximum loads of all the cables produced by a company.

A: $n = 60, \quad \bar{x} = 01.09$.

$\sigma = 0.73$.

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

~~$= \bar{x} - 1.09$~~

95% of CI $= \bar{x} \pm 1.96 \left(\frac{\sigma}{\sqrt{n}}\right)$

$= 1.09 \pm 1.96 \cdot \left(\frac{0.73}{\sqrt{60}}\right)$.

$= 1.2747 \; , \; 0.9052$.

99% of CI $= \bar{x} \pm 2.58 \left(\frac{\sigma}{\sqrt{n}}\right)$

$= 1.333 , \; 0.8468$.

18/5/17

Note suppose in the question, if infer confidence limit or probable limit is mentioned, then we've to solve by confidence method.

8. A sample of 900 days was taken in a costal town & it was found that the 100 days the weather was very ~~high~~ hot, ~~optain~~ obtain the probable limit of 1%. of very hot weather .

Sol

$n = 900$

$P = \frac{100}{900} = 1/9$

$q = 1 - P$

$q = 8/9$

$\therefore$ 99% CI for proportion

$$99\% \text{ of } CI(P) : P \pm 2.58\sqrt{\frac{PQ}{n}}$$

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

$$= 0.084$$

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

$$= 0.13$$

$$\therefore 0.084 \leq P \leq 0.13$$

9. In a sample of 500 men, It was found that 60% of them had overweight. what can we infer about the proportion of people having overweight in the population.

$n = 500$

$P = 0.6 \Rightarrow q = 0.4$.

use 99% ~~of~~ of Level of significance.

$$99\% \text{ of } CI(P) = P \pm 2.58\sqrt{\frac{PQ}{n}}$$

$$= 0.6 \pm 2.58\sqrt{\frac{(0.6)(0.4)}{500}}$$

$$= 0.6565 \text{ or } 0.5434$$

~~95% of CI(P) = P ±~~

10. A sample of 100 bulbs produced by a company. Showed a mean life of 1190 hours. & S.D of 19 hrs. and also a sample of 75 bulbs produced by company B showed a mean life of 1230 hrs. & S.D is 120 hrs. Is there a difference b/w the mean life of the bulb produced by 2 companies at 5% LOS. 1% LOS

Sol:

Company - A.

$n_1 = 100$

$\bar{x}_1 = 1190$

$\sigma_1 = 19$.

Company - B

$n_2 = 75$

$\bar{x}_2 = 1230$

$\sigma_2 = 120$.

H_0 = there is no significant difference b/w mean life of A & B companies.

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$Z = \frac{1190 - 1230}{\sqrt{\frac{(19)^2}{100} + \frac{(120)^2}{75}}} \Rightarrow Z = -2.85$$

$\therefore |Z| = 2.85$.

$$Z_{tab} = \begin{cases} 2.58 & \text{at } 1\% \text{ LOS} \\ 1.96 & \text{at } 5\% \text{ LOS} \end{cases}$$

$\therefore |Z|_{cal} > Z_{tab}$

$\Rightarrow$ Ho is rejected.

11. In an elementary school examination, the mean rate of 32 boys was 72. with $\sigma_1 = 8$ and while the mean rate of 36 girls was 75 with $\sigma_2 = 6$. Test the hypothesis that the performance of the girls is better than the boys.

A:-

$n_1 = 32$ $\qquad$ $n_2 = 36$

$\bar{x}_1 = 72$ $\qquad$ $\bar{x}_2 = 75$

$\sigma_1 = 8$ $\qquad$ $\sigma_2 = 6$.

Ho : there is no significant differences b/w the performance of girls and boys

$$\therefore Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{\bar{x}_1} + \frac{\sigma_2^2}{\bar{x}_2}}}$$

$$Z = \frac{72 - 75}{\sqrt{\frac{(8)^2}{72} - \frac{(6)^2}{75}}}$$

$$= \frac{-3}{\sqrt{\frac{64.32}{72} - \frac{36}{75}}} = 1.73$$

2. 1.96 5%

2.58 1%

$Z_{cal} = 1.73$.

$Z_{tab} = \begin{cases} 2.58 & \text{at } 1\% \text{ LOS} \\ 1.96 & \text{at } 5\% \text{ LOS} \end{cases}$

$Z_{cal} < Z_{tab}$

$\therefore$ Ho is accepted.

12. In a city A, 20% of rand sample of 900 School boys has physical defects. In another city B 18.5% of random sample of 1600 boys has the same defect.. Is the difference b/w the proportion is Significant. [use 1% of level of Significance].

Ans. 20% $\qquad$ 900

18.5% $\qquad$ 1600

$n_1 = 900$ $\qquad$ $n_2 = 1600$

$P_1 = 0.2$ $\qquad$ $P_2 = 0.185$

$$Z = \frac{P_1 - P_2}{\sqrt{PQ\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} \quad . \quad P_0 = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}.$$

Ho : There is no significant difference b/w the proportion.

$Z_{cal} =$

$$P = \frac{900(0.2) + 1600(0.185)}{900 + 1600}$$

$P = 0.1904.$

$q = 1 - p = 1 - 0.1904 = 0.8096.$

$\therefore Z = 0.916.$

$Z_{cal} = 0.996. \quad Z_{tab}$ at 1% $= 2.58.$

$\Rightarrow Z_{cal} < Z_{tab}$

$\Rightarrow$ Ho is accepted.

13. A Random sample of 1000 Engg students from a city A and 800 from city B were taken. It was found that 400 students in each of the sample were from payment quota. ~~Then~~ Does the data reveal a significant diffence b/w two cities w.r.t payment ~~sto~~ quota students.

Ans

A: $n_1 = 1000$ B: $n_1 = 800$

400 - payment seats

$P_1 = \frac{400}{1000} = 2/5 = 0.4$, $P_2 = \frac{400}{800} = 1/2 = 0.5$.

Ho: there is no significant diff b/w two cities

P.T.O.

$$Z_1 = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{1000(0.4) + 800(0.5)}{1000 + 800}$$

$P = 4/9$

$\Rightarrow q = 5/9$

$$\therefore Z = \frac{1000 - 800}{\sqrt{4/9 \times 5/9 \left(\frac{1}{1000} + \frac{1}{800}\right)}}$$

$Z_{cal} = 4.243$

$$\therefore Z_{tab} = \begin{cases} 2.58 \text{ at } 1\% \text{ LOS} \\ 1.96 \text{ at } 5\% \text{ LOS} \end{cases}$$

$\sum (x - \bar{x})^2$

$Z_{cal} > Z_{tab}$

$\Rightarrow$ Ho is rejected.

14. A Random Sample of 1000 workers in a company has mean wage of 50 Rs/day and S.D. is 15. Another sample of 1500 workers from the another company has mean wage of 45 Rs/day & SD is 20 Rs. Find the 95% of confidence limit for the difference of the mean wages of population of 2-companies.

Sol

A

$n_1 = 1000$

$\bar{x}_1 = 50$

$\sigma_1 = 15$

B

$n_2 = 1500$

$\bar{x}_2 = 45$

$\sigma_2 = 20$

$$(\bar{x}_1 - \bar{x}_2) \pm 1.96\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(50 - 45) \pm 1.96\sqrt{\frac{(15)^2}{1000} + \frac{(20)^2}{1500}}$$

$$5 \pm 1.96\sqrt{0.225 + 0.2666}$$

$= 3.62$ & 6.3

19/5/17

t – distribution & chi – Squar test

student · t – distribution [t – test]

① $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ s – S.D

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

② $t = \dfrac{\bar{x_1} - \bar{x_2}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $s^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$.

s_1 is the S.D of 1^{st} sample

s_2 ——— " ——— 2^{nd} sample.

$\bar{x}_1$ is mean of 1^{st} sample

$\bar{x}_2$ ——— " 2^{nd} samle

n_1 - size of 1^{st} sample.

n_2 size of 2^{nd} sample.

③ 95% Confidence limits for (μ):

$$\bar{x} \pm t_{0.05} \left[\frac{s}{\sqrt{n}}\right]$$

Note :- Degree of freedom. [df]

Degree of freedom is nothing but

no of values — 1

$[df] = n - 1$

Problems

1. 10 individuals are choosen at random from the population & their height 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. at $[t_{0.05} = 2.262$ for 9df$]$

Ans

Given $n = 10$

Soln :- $\mu = 66$ inches.

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

$$= 67.8$$

x	$(x-\bar{x})^2$
63	23.04
63	23.04
66	3.24
67	0.64
68	0.04
69	1.44
70	4.84
70	4.84
71	10.24
71	10.24

$$\sum(x-\bar{x})^2 = 81.6$$

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

$$= \frac{1}{9}(81.5) = 9.067$$

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

$$S^2 = \frac{1}{n-1}\sum(x-\bar{x}) \qquad \bar{x} = \frac{\sum x}{n}$$

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{67.8-66}{3.011/\sqrt{10}} = 1.89$$

$t_{cal} < t_{tab} \Rightarrow H_0$ is accepted.

2. A certain stimulus administered to each of the 12 pateints resulted in the following change in BP
5 2 8 -1 3 0 6 -2 1 5 0 4
can it be concluded that this stimulus will increase the B.P? $[t_{0.05} = 2.201$ for 11 df$]$

Given $n = 12$, $\mu = 0$.

$$\bar{x} = \frac{\sum x}{n} = \frac{31}{12} = \cancel{2.58}\ 2.58$$

x	$(x-\bar{x})^2$
5	5.8564
2	0.3364
8	29.3764
-1	12.816
3	0.1764
0	6.6564
6	11.6964
-2	20.9764
1	2.4964
5	5.5864
0	6.6564
4	2.0164
	104.908

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

$$= \frac{1}{11} 104.908$$

$$\boxed{s^2 = 9.5370}$$

$\Rightarrow s = 3.088$

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{2.58-0}{\frac{3.088}{\sqrt{12}}}$$

$= 2.894$

$t_{cal} > t_{tab} \Rightarrow H_0$ is not accepted.

$\bar{x} = \frac{\sum x}{n}$

3. A sample of 10 measurements of the diameter of a sphere gave a mean 12 cm & S.D is 0.15 cm find 95% of confidence limit for the actual diameter $[t_{0.05} = 2.201$ for 9 df$]$

$n = 10$ $\mu = 12$
σ 0.15

Ans
$\bar{x} = 12$ cm & $\sigma = 0.15$ cm, $n = 10$

95% of confidence limits

$$\bar{x} \pm t_{0.05} \left[s/\sqrt{n} \right]$$

$$12 \pm 2.20 \times \frac{0.15}{\sqrt{10}}$$

$\therefore$ μ lies b/w 12.104 or 11.8951

$\therefore$ $\boxed{11.895 < \mu < 12.104}$.

4. 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 S.D 0.3. Can it be said that the machine is producing as per specification? [$t_{0.05} = 2.201$ for 24df].

Ans Sol^n

$n = 25$

$\mu = 3$ $\quad t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$. $\quad S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + }$

$\bar{x} = 3.1$

$S = 0.3$. $\quad S^2 = \frac{\sum (x - \bar{x})^2}{(n-1)}$

Q5. A sample of 11 rates from a central population had an average blood viscosity of 3.92 with a S.D of 0.61. On the basis of this sample, establish 95% confidence limits for μ, the mean viscosity of the central population [~~to~~ $t_{0.05} = 2.228$ for 10df].

Sol^n. $n = 11$

$\bar{x} = 3.92$.

$S = 0.61$

$\bar{x} \pm t_{0.05}\left[S/\sqrt{n}\right] =$ $\quad \le \mu \le$

6. A group of boys and girls where given an intellige-nce test. The mean score, SD score and number in each group are as follows

	Boys	Girls		
mean	74	70		
S.D	8	10		
n.	12	10		

Is the difference b/w the means of the two groups significant at 5% LOS ? [$t_{0.05} = 2.086$ for 20df].

Sol

~~$t = \bar{x}_1 - \bar{x}_2$~~

Ho: there is no significant difference b/w girls and boys.

$\bar{x}_1 = 74 \quad \bar{x}_2 = 70$

$s_1 = 8 \quad s_2 = 10$

$n_1 = 12 \quad n_2 = 10.$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{1/n_1 + 1/n_2}}$$

$$s^2 = \frac{12 \times (8)^2 + 10(10)^2}{12 + 10.}$$

$= 80.36.$

$s = 8.964$

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

$t_{cal} =$ $\quad t_{tab} = 2.086$

$$t = \frac{74 - 70}{8.964\sqrt{1/12 + 1/10}} = 1.044$$

$t_{cal} = 1.044 \quad t_{tab} = 2.086.$

20/5/17

Chi - square test

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

where 'O' is observed frequency

E is expected frequency χ - (chi)

$\sum O = N$ $\quad F(x) = N.P(x)$

$n = df$

1. Fit a poisson distribution for the following data and test the goodness of fit given that

$\chi^2_{0.05} = 7.815$ for 3 df

x	0	1	2	3	4
$O = f$	122	60	15	2	1
E	121.306	60.65	15.16	2.52	0.31

Sol

$E = F$?

$N = \sum f = 200$

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

$$m = \frac{\sum f \times x}{\sum f} = \frac{60+30+6+4}{200} = \frac{100}{200} = 1/2$$

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

Expected frequency $F(x) = N\,P(x) = \frac{200}{x!}(0.5)^x e^{-0.5}$

~~at x=0~~ $E \rightarrow F = \frac{200}{x!}(0.5)^x e^{-0.5}$

at $x = 0$

$$= \frac{200}{1!}(0.5)^0 e^{-0.5}$$

$= 121.306$

at $x = 1$

$$\frac{200}{1}(0.5) \cdot e^{-0.5} = 60.65$$

$$\frac{(O-E)^2}{E} = \frac{(122 - 121.13)^2}{121.3} = 6.2399 \times 10^{-3}$$

$$\frac{(O-E)^2}{E} = \frac{(60 - 60.65)^2}{60.65} = 6.966 \times 10^{-3}$$

$$\frac{(O-E)^2}{E} = \frac{(15 - 15.16)^2}{15.16} = 1.688 \times 10^{-3}$$

$$\frac{(2 - 2.25)^2}{2.25} = 0.0277$$

$$\frac{(1 - 0.31)^2}{0.31} = 1.5358$$

$$\sum \frac{(O-E)^2}{E} = 6.2399 \times 10^{-3} + 6 - - -$$

$$= 1.578$$

$$\psi^2_{cal} < \psi^2_{tab}$$

$\therefore$ Hypothesis is accepted.

2. Fit the poisson distribution for the following data and test for the goodness of fit given that $\psi^2_{0.05} = 9.49$ for 4 df.

x	0	1	2	3	4
$O \to f$	419	352	154	56	19
E	404.94	366.071	165.46	49.85	11.268
$\frac{(O-E)^2}{E}$	0.4881	0.540	0.7937	0.758	5.3058

$$\Sigma \frac{(O-E)^2}{E} = 7.889$$

$N = \sum f = 1000$

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

$$m = \frac{\sum fx}{\sum f} = \frac{904}{1000} = 0.904$$

$$P(x) = \frac{(0.904)^x e^{-0.904}}{x!}$$

$$F(x) = N\,P(x) = \frac{1000}{x!}(0.904)^x e^{-0.904}$$

at $x=0$. $1000\,(1)\times e^{-0.904} = 404.94$

$x=1$ $1000\,(0.904)\,e^{-0.904} = 366.071$

$x=2$ $\frac{1000}{2!}(0.904)^2 e^{-0.904} = 165.46$

$x=3$ $\frac{1000}{3!}(0.904)^3 e^{-0.904} = 49.8599$

$x=4$ 11.2683

$$\sum \frac{(O-E)^2}{E} = 7.885$$

$\psi_{cal} < \psi_{tab}$

$\therefore$ Hypothesis is accepted.

3. Four coins were tossed 160 times and the follo result were obtained. Test the goodness of fit of the binomial distribution given that. $[\psi^2_{0.05} = 9.49 \text{ for } 4\,df]$

No of heads	0	1	2	3	4
frequency	17	52	54	31	6

$n = 4$.

$N = 160$.

$p = 1/2,\ q = 1/2$ {Probability of tossing a coin.

Binomial Distribution.

$P(x) = {}^nC_x \cdot p^x q^{n-x}$.

$P(x) = 4C_x (1/2)^x (1/2)^{4-x} = \frac{1}{2^4} \, 4C_x$.

$F = N P(x) = \frac{160}{16} \cdot 4C_x = 10\,4C_x$.

x =	0	1	2	3	4
O = ~~fre~~	17	52	54	31	6
$E \to F$ =	10	40	60	40	10
$\frac{(O-E)^2}{E}$ =	$\cancel{2/5}$ 4.9	18/5 3.6	3/5	81/40	8/5

$\sum \frac{(O-E)^2}{E} = \frac{\cancel{329}}{\cancel{20}} = \cancel{8.225}$ 12.725

$\psi_{cal} > \psi_{tab}$

$\Rightarrow$ Hypothesis is ~~accepted~~ rejected.

• 5 dice were thrown 96 times and the numbers ~~not~~ 1 or 2 or 3 appearing on the face of the dice follows the frequency distribution as follows

$[\cancel{\psi^2_{0.05} = 9.49 \text{ for } 4 \, df}]$.

No of the dice showing 1 or 2 or 3	5	4	3	2	1	0
~~f~~ frequency	7	19	35	24	8	3

S. Test the hypothesis that the data follows a binomial distribution $[\psi^2_{0.05} = 9.49 \text{ for } 5 \, df]$.

$n = 5 \qquad N = 96$

$P = 3/6 = 1/2$ {out of 6 we need only 3 sides

$\Rightarrow q = 3/6 = 1/2$.

Binomial Distribution

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

$$= {}^5C_x (1/2)^x (1/2)^{5-x}$$

$$= \frac{1}{32} \, {}^5C_x$$

$$F = N\,P(x) = \frac{96}{32} \, {}^5C_x \ldots = 3({}^5C_x)$$

x	5	4	3	2	1	0
O	7	19	35	24	8	3
F = E	3	15	30	30	15	3
$\frac{(O-E)^2}{E}$	5.333	1.066	0.833	1.2	3.26	0

$$\sum \frac{(O-E)^2}{E} = 11.692$$

$\psi_{cal} > \psi_{tab} \Rightarrow$ hypothesis is rejected.