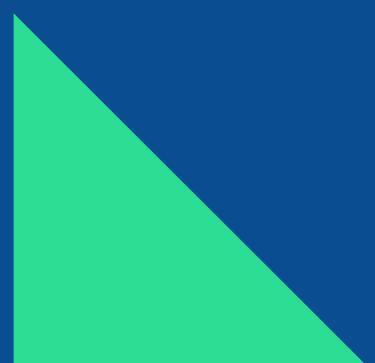




A NEW & EASY WAY TO LEARN

# AZ Documents.in

*Get All Vtu 18th Scheme  
notes for all branches here*



**THIS NOTES WAS  
DOWNLOADED FROM  
AZDOCUMENTS.IN**

[Click here to visit the website](#)

## MODULE-5

### JOINT PROBABILITY DISTRIBUTION & SAMPLING THEORY

#### **Joint Probability :**

Let  $X = \{x_1, x_2, x_3, \dots, x_m\}$  and  $Y = \{y_1, y_2, y_3, \dots, y_n\}$  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

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

The joint probability table as shown below

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

#### **Marginal probability distributions:**

In the joint probability table  $f(x_1), f(x_2), f(x_3), \dots, f(x_m)$  and  $g(y_1), g(y_2), g(y_3), \dots, g(y_n)$  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

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

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

$$\text{or } \text{COV}(X, Y) = 0$$

### Expectation , Variance & Covariance:

Let X be the random variable taking the random values  $x_1, x_2, x_3, \dots, x_m$ , 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  $\sigma_y^2 = E(Y^2) - [E(Y)]^2$   
 $\Rightarrow \sigma_y^2 = \sum_i y_i^2 f(y_i) - \mu_y^2$

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

$$COV(X, Y) = \sum_i \sum_j x_i y_j f(x_i, y_j) - \mu_x \cdot \mu_y$$

The correlation between X and Y is  $\rho(X, Y) = \frac{COV(X, Y)}{\sigma_x \sigma_y}$

### 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	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
5	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$
$g(y_i)$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{4}$	1

The marginal distribution of X and Y are

$x_i$	1	5
$f(x_i)$	$\frac{1}{2}$	$\frac{1}{2}$

$y_i$	-4	2	7
$g(y_i)$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{4}$

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

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

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

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

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

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

$$\Rightarrow \sigma_x^2 = \sum_i x_i^2 f(x_i) - \mu_x^2$$

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

$$\sigma_y^2 = E(Y^2) - \mu_y^2$$

$$\Rightarrow \sigma_y^2 = \sum_j y_j^2 g(y_j) - \mu_y^2$$

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

iv)  $COV(X, Y) = E(XY) - \mu_x \mu_y$

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

v)  $\rho(X, Y) = \frac{COV(X, Y)}{\sigma_x \sigma_y}$

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

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

$X \backslash Y$	3	4	5
2	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
5	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
7	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

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	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{2}$
5	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$
7	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$
$g(y_i)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

The marginal distributions of X and Y are

$x_i$	2	5	7
$f(x_i)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

$y_i$	3	4	5
$g(y_i)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$$\begin{aligned}\mu_x &= E(X) = \sum_i x_i f(x_i) \\ \Rightarrow \mu_x &= \left(2 \times \frac{1}{2}\right) + \left(5 \times \frac{1}{4}\right) + \left(7 \times \frac{1}{4}\right) \\ \Rightarrow \mu_x &= 4\end{aligned}$$

$$\begin{aligned}
 \mu_Y &= E(Y) = \sum_j y_j g(y_j) \\
 \Rightarrow \mu_Y &= \left(3 \times \frac{1}{3}\right) + \left(4 \times \frac{1}{3}\right) + \left(5 \times \frac{1}{3}\right) \\
 \Rightarrow \mu_Y &= 4 \\
 E(XY) &= \sum_{i,j} x_i y_j p_{ij} \\
 \Rightarrow E(XY) &= \left(2 \times 3 \times \frac{1}{6}\right) + \left(2 \times 4 \times \frac{1}{6}\right) + \left(2 \times 5 \times \frac{1}{6}\right) + \left(5 \times 3 \times \frac{1}{12}\right) + \left(5 \times 4 \times \frac{1}{12}\right) + \left(5 \times 5 \times \frac{1}{12}\right) \\
 &+ \left(7 \times 3 \times \frac{1}{12}\right) + \left(7 \times 4 \times \frac{1}{12}\right) + \left(7 \times 5 \times \frac{1}{12}\right) = 16 \\
 \therefore Cov(X, Y) &= E(XY) - \mu_X \mu_Y \\
 \Rightarrow Cov(X, Y) &= 16 - 4 \times 4 \\
 \Rightarrow Cov(X, Y) &= 0
 \end{aligned}$$

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

$X \backslash Y$	1	3	6
1	$\frac{1}{9}$	$\frac{1}{6}$	$\frac{1}{18}$
3	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{12}$
6	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{36}$

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	$\frac{1}{9}$	$\frac{1}{6}$	$\frac{1}{18}$	$\frac{1}{3}$
3	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{2}$
6	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{36}$	$\frac{3}{18}$
$g(y_i)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{3}{18}$	1

The marginal distributions of X and Y are

$x_i$	1	3	6
$f(x_i)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{3}{18}$

$y_i$	1	3	6
$g(y_i)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{3}{18}$

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

$P(X = x_1, Y = y_1) = P(X = x_1) \times P(Y = y_1)$  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	$\frac{1}{8}$	$\frac{1}{24}$	$\frac{1}{12}$
4	$\frac{1}{4}$	$\frac{1}{4}$	0
6	$\frac{1}{8}$	$\frac{1}{24}$	$\frac{1}{12}$

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	$\frac{1}{8}$	$\frac{1}{24}$	$\frac{1}{12}$	$\frac{1}{4}$
4	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{2}$
6	$\frac{1}{8}$	$\frac{1}{24}$	$\frac{1}{12}$	$\frac{1}{4}$
$g(y_i)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$	1

The marginal distributions of X and Y are

$x_i$	2	4	6
$f(x_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$y_i$	3	4	5
$g(y_i)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

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

$$\begin{aligned}
 E(XY) &= \sum_{i,j} x_i y_j p_{ij} \\
 \Rightarrow E(XY) &= \left(2 \times 1 \times \frac{1}{8}\right) + \left(2 \times 3 \times \frac{1}{24}\right) + \left(2 \times 9 \times \frac{1}{12}\right) + \left(4 \times 1 \times \frac{1}{4}\right) + \left(4 \times 3 \times \frac{1}{4}\right) + \left(4 \times 9 \times 0\right) \\
 &+ \left(6 \times 1 \times \frac{1}{8}\right) + \left(6 \times 3 \times \frac{1}{24}\right) + \left(6 \times 9 \times \frac{1}{12}\right) = 12 \\
 \therefore Cov(X, Y) &= E(XY) - \mu_X \mu_Y \\
 \Rightarrow Cov(X, Y) &= 12 - 4 \times 3 \\
 \Rightarrow Cov(X, Y) &= 0 \\
 \text{The correlation } \rho(X, Y) &= \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = 0
 \end{aligned}$$

- 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)

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

$\mu_1$  = the mean of population 1, and

$\mu_2$  = 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  $\mu_1, \mu_2$  be the means,  $\sigma_1, \sigma_2$  be the standard deviations of two populations and  $\bar{x}_1, \bar{x}_2$  are the means of the samples, then

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

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

### Student's $t$ -distribution:

Let  $\mu$  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

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

Another formula for  $t$ -test of two samples is

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

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

### Chi-square distribution:

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

$$\begin{aligned} \chi^2 &= \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} + \dots + \frac{(O_n - E_n)^2}{E_n} \\ \Rightarrow \chi^2 &= \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \end{aligned}$$

- 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= $np=1000 \times 0.5=500$  ,  $npq=250$

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

$$\begin{aligned} Z &= \frac{x - \mu}{\sigma} = \frac{x - np}{\sqrt{npq}} \\ \therefore \text{ Consider} \quad & \\ \Rightarrow Z &= \frac{40}{\sqrt{250}} = 2.53 < 2.58 \end{aligned}$$

Thus we can say that the coin is unbiased

- 2) In 324 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= $np=324 \times 0.5=162$  ,  $npq=81$

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

$$\begin{aligned} Z &= \frac{x - \mu}{\sigma} = \frac{x - np}{\sqrt{npq}} \\ \therefore \text{ Consider} \quad & \\ \Rightarrow Z &= \frac{19}{\sqrt{81}} = \frac{19}{9} 2.11 < 2.58 \end{aligned}$$

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

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

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

$$\therefore \text{ The difference } = 3240 - 3000 = 240$$

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

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

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.

$$\therefore \text{Difference} = 190 - 182 = 8$$

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 2.33 at 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 :  $t_{0.05}$  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

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

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

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

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

$$\begin{aligned} t &= \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \\ \Rightarrow t &= \frac{97.2 - 100}{\left( \frac{14.2735}{\sqrt{10}} \right)} \\ \Rightarrow t &= \frac{-2.8}{4.5136} \approx -0.6203 < 2.262 \end{aligned}$$

- 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 66 inches ( $t_{0.05} = 2.262$  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

$$\begin{aligned} s^2 &= \frac{1}{n-1} \sum (x - \bar{x})^2 \\ \Rightarrow s^2 &= \frac{1}{9} \left\{ (63 - 67.8)^2 + (63 - 67.8)^2 + (66 - 67.8)^2 + (67 - 67.8)^2 + (68 - 67.8)^2 + (69 - 67.8)^2 \right. \\ &\quad \left. + (70 - 67.8)^2 + (70 - 67.8)^2 + (71 - 67.8)^2 + (71 - 67.8)^2 \right\} \\ \Rightarrow s^2 &= 9.067 \Rightarrow s = 3.011 \end{aligned}$$

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

we have

$$\begin{aligned} t &= \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \\ \Rightarrow t &= \frac{67.8 - 66}{\left( \frac{3.011}{\sqrt{10}} \right)} \\ \Rightarrow t &= 2.8979 \approx 1.89 > 2.262 \end{aligned}$$

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:

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

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

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

Sol.

Given

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

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

We know that

$$\begin{aligned} s^2 &= \frac{n_1 \sigma_1^2 + n_2 \sigma_2^2}{n_1 + n_2 - 2} \\ \Rightarrow s^2 &= \frac{(10 \times 100) + (10 \times 121)}{10 + 10 - 2} \\ \Rightarrow s^2 &= 122.78 \\ \Rightarrow s &= 11.0805 \end{aligned}$$

we have

$$\begin{aligned} t &= \frac{(\bar{x}_2 - \bar{x}_1)}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\ \Rightarrow t &= \frac{560 - 500}{11.0805 \sqrt{0.1 + 0.1}} \\ \Rightarrow t &= 12.1081 \approx 12.11 \end{aligned}$$

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  $\chi^2_{0.05} = 7.815$  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

$$\begin{aligned}\therefore \chi^2 &= \sum_i \left[ \frac{(O_i - E_i)^2}{E_i} \right] \\ \Rightarrow \chi^2 &= \frac{1}{121} + \frac{1}{61} + 0 + 0 \\ \Rightarrow \chi^2 &= 0.025 < \chi^2_{0.05} = 7.815\end{aligned}$$

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

$$\begin{aligned}P(X = x) &= 4_{Cx} (0.5)^x (0.5)^{4-x} \\ P(1) &= 4_{C1} (0.5)^1 (0.5)^{4-1} = 0.25 \\ P(2) &= 4_{C2} (0.5)^2 (0.5)^{4-2} = 0.375 \\ P(3) &= 4_{C3} (0.5)^3 (0.5)^{4-3} = 0.25 \\ P(4) &= 4_{C4} (0.5)^4 (0.5)^{4-4} = 0.0625\end{aligned}$$

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

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

$$E_2 = 100 \times 0.375 = 37.5 \quad \text{Where } 100 \text{ is the sum of frequency}$$

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

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

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

$$\begin{aligned}\therefore \chi^2 &= \sum_i \left[ \frac{(O_i - E_i)^2}{E_i} \right] \\ \Rightarrow \chi^2 &= \frac{1.5625}{6.25} + \frac{16}{25} + \frac{2.25}{37.5} + 0 + \frac{1.5625}{6.25} \\ \Rightarrow \chi^2 &= 0.25 + 0.64 + 0.06 + 0.25 \\ \Rightarrow \chi^2 &= 1.2 < \chi^2_{0.05} = 9.49\end{aligned}$$

Hence the fitness is good.

\*\*\*

*Thank you*

*By:*

*Purushotham.P*

PRP@SUCIT

Sampling Theory | Stochastic Process

Stochastic process

Probability vector

A vector whose components are non-negative and their sum is equal to 1  
 $v_i \geq 0 \quad \sum_{i=1}^n v_i = 1$

Stochastic Matrix

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

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

Ex:  $\begin{bmatrix} 1/4 & 1/4 & 1/2 \end{bmatrix}$

Fixed vector or fixed point :-  
A vector ' $V$ ' is said to be fixed vector or a fixed point of a matrix ' $A$ ' if

$VA = V$

Regular Stochastic matrix

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

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

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

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

Note :- i) Let ' $P$ ' be a regular stochastic matrix  
ii)  $P$  has a unique fixed probability vector.

iii)  $P$  is a stochastic matrix

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

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

### Examples

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

Sol

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

To Prove

$$VA = V \quad \text{i.e. } x + y = 1$$

W.R.T by the property of regular stochastic matrix

i.e.

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

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

$$\Rightarrow \frac{x}{4} + \frac{y}{2} = x \quad \text{&} \quad \frac{3x}{4} + \frac{y}{2} = y$$

$$\Rightarrow \frac{y}{2} = x - \frac{x}{4} \quad \text{&} \quad y = \frac{3x}{2}$$

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

$\therefore$  substitute  $y = \frac{3x}{2}$  to get  $x$

$$\cancel{\frac{3x}{4}} + \cancel{\frac{y}{2}}$$

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

$$\cancel{\frac{3x}{4}} = \cancel{\frac{3x}{2}}$$

$$x+y=1 \Rightarrow x+3x/2 = 1 \text{ (dividing by } 2\text{)}.$$

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

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

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

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

2. Find the unique fixed probability 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 w.r.t  $A$

i.e.  $x+y+z=1$  (sum of probabilities is 1)

w.r.t by the prob of regular stochastic matrix

$$\text{i.e. } VA = V.$$

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

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

$$\Rightarrow x = \frac{y}{2}, y = \frac{3}{4}x + \frac{1}{6}y + z, z = \frac{1}{4}x$$

$$x+y+z=1$$

$$\Rightarrow x + \frac{y}{2} + \frac{x}{4} = 1$$

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

$$\therefore x = \frac{4}{13}.$$

$$\therefore y = 2x = \frac{8}{13}.$$

$$\therefore z = \frac{1}{4} \times \frac{4}{13} = \frac{1}{13}$$

$$\Rightarrow V = \left[ \frac{4}{13}, \frac{8}{13}, \frac{1}{13} \right]$$

H.W

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

$$(s, v, r) = V$$

$V = Av$  where  $\vec{v}$

$$4: A = \begin{bmatrix} (s, v, r) &= & \begin{bmatrix} sV & vV & rV \\ sV & 0 & sV \\ 0 & 1 & 0 \end{bmatrix} (s, v, r) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.3 & 0.5 & 0 \end{bmatrix}$$

$$(s, v, r) = \begin{bmatrix} sV + vV + rV \\ sV + vV \\ sV + vV \end{bmatrix} V$$

$$x + y + z = 1 \quad v = s + vV \quad x = sV + vV$$

To prove:

$$VA = V$$

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

Explain

$$\begin{bmatrix} z(0.5), x+z(0.5), y \end{bmatrix} = (x, y, z)$$

$$z(0.5) = x, x + 0.5z = y, y = z$$

$$x + y + z = 1$$

$$\frac{1}{2}z + z + z = 1 \quad = \frac{3}{2}z = 1$$

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

$$V = \left[ \frac{1}{5}, \frac{2}{5}, \frac{2}{5} \right]$$

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

$$V = (x, y, z)$$

To prove  $VA = V$ .

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

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

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

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

$$\text{test } 1/2y = 1/2z \Rightarrow [1/4x + 1/2y + 1/2z] = [x, y, z]$$

$$(x + y + z) + 2 = 1$$

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

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

$$1/4x = 1$$

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

$$x = 1 \times \frac{1}{2} = x, e^x = e^x, e^{1/2} = e^{1/2} = 1$$

$$[e^x, e^{1/2}, e^{1/2}] = V$$

## Transition probability matrix or Stochastic matrix

The probabilities of moving from one state to another state or 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

$$P = \{P_{ij}\} = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{bmatrix}$$

Markov chain is a kind of simple stochastic process.

## Markov chain

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

The entry  $P_{ij}$  in the transition probability matrix  $P$  of the markov chain is the probability that the system changes from the state  $\alpha_i$  to  $\alpha_j$  in a single step.

$$P^{(1)} = P^{(0)} P^{(1)}$$

$$P^{(2)} = P^{(0)} P^{(2)}$$

INSTITUTE OF TECHNOLOGY

## (SOURCE DIGINOTES)

### Note

Markov chain is irreducible 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 continues to remain in the same state such a state is called as absorbing state.

Ex

P.T.O

Ex:-

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

Answer to this question is  
Transient state since it is not possible to return to state 3 from state 1.

2. Transient State :- A state  $i$  is said to be transient 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.

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \rightarrow P^2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

The system remains in the same state after two steps.

4. Periodic State :- The recurrent state  $i$  is said to be periodic state if  $\text{GCD}(d_i) > 1$

Ex Identity matrix

### Problems

### (SOURCE DIGINOTES)

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

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

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

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

$\therefore$  It is regular.

2. Define a stochastic matrix and show that

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

(is a regular stochastic matrix)

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

Definition is written backside.

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

$$P^3 = \begin{bmatrix} 1/4 & 5/16 & 1/8 \\ 3/8 & 9/32 & 11/32 \\ 3/8 & 19/64 & 21/64 \end{bmatrix}$$

3. A student's study habits are as follows

If he studies one night, he is 70% sure not to study the next night. On 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  $S = \{\text{studying}(A), \text{not studying}(B)\}$

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

We have to find unique fixed probability vector.

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

$$\Rightarrow x + y = 1 \rightarrow \textcircled{1}$$

$\therefore$  from stochastic matrix we have

$$NP = v$$

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

$$\left[ 0.3x + 0.4y, 0.7x + 0.6y \right] = [x, y]$$

$$0.3x + 0.4y = x, 0.7x + 0.6y = y$$

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

$$4y = 7x$$

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

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

$$x + y = 1$$

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

$$11x = 4$$

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

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

Hence required fixed probability is  $(4/11, 7/11)$ .

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

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

Q. A man's smoking habits are as follows.

If he smokes filter cigarettes one week

he switches to non-filter cigarettes 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 cigarettes?

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

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

$$x + y = 1 \rightarrow ①$$

$$VP = V$$

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

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

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

$$0.8x + 0.3y = x$$

$$0.3y = 0.2x$$

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

$$x + y = 1$$

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

$$3x + 2x = 3 \Rightarrow 5x = 3 \Rightarrow x = \frac{3}{5}$$

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

$$y = \frac{2}{5}$$

$$x = \frac{3}{5}, y = \frac{2}{5}$$

$$V = \left[ \frac{3}{5}, \frac{2}{5} \right]$$

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 B
- ii) If the person has visited club B on Monday, find the probability that he visits club A on Tuesday.

SOL<sup>n</sup>

$$P = \begin{bmatrix} A & B \\ \text{Mon} & \text{Tue} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$\begin{aligned} i) \quad & x+y=1 \\ & VP = V \\ \Rightarrow & [x,y] \cdot \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = [x,y] \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = V \end{aligned}$$

- ii) If the person visited club B on Monday, i.e. Day 1 then he visits club A on Thursday day 3.

$$P^3 = \begin{bmatrix} 0.25 & 0.75 \\ 0.375 & 0.625 \end{bmatrix}$$

- 6) Every year a man trades his car for a new car. If he has Maruti he trades it for an Ambassador, 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 Maruti or an Ambassador. In the year 2000, he bought his first car which was a Santro. Find the probability that he has

i) 2002 Maruti

ii) 2003 Santro

$$\begin{bmatrix} 0.5 & 0.5 \end{bmatrix} = V$$

SOL  $P = m \begin{bmatrix} M & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

After 2 years.

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

$$A_{3,1} = 0.11$$

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

$$A_{3,3} = 0.592$$

Q. 3. bags 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 has the 1st person to throw the ball, find the probability that:

i) A has a ball

ii) b has a ball

iii) C has a ball for the  $n^{th}$  throw

SOL  $P = A \begin{bmatrix} A & B & C \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$

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

i)  $0.25 = \frac{1}{4}$

ii)  $0.25 = \frac{1}{4}$

iii)  $0.5 = \frac{1}{2}$

8. & 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?

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

To find the long run we have to use UFP vector

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

$$\Leftrightarrow a+b+c+d = 1 \rightarrow ①$$

From stochastic matrix we have  
 $V P = V$

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

$$\Rightarrow \frac{b}{2} + \frac{c}{4} + \frac{d}{2} = a$$

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

$$\frac{a}{4} + \frac{b}{4} = c$$

$$\frac{a}{4} + \frac{b}{4} = d$$

$$\Rightarrow c = d$$

$$\Rightarrow b_1 + c_1 + d_1 = a$$

$$\Rightarrow a_1 + c_1 + d_1 = b$$

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

$$\underline{a/2 + c = b}$$

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

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

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

$$\frac{3b}{2} = \frac{3}{2}a$$

$$3b = 3a \quad \therefore b = a$$

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

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

**CAMBRIDGE**  
INSTITUTE OF TECHNOLOGY  
**(SOURCE DIGINOTES)**

## 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 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 before the next draw. In this method

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

CAMBRIDGE  
INSTITUTE OF TECHNOLOGY  
(SOURCE DIGINOTES)

16/5/17

## Testing of Hypothesis

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 acceptance is called alternate hypothesis denoted by  $H_1$ .

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

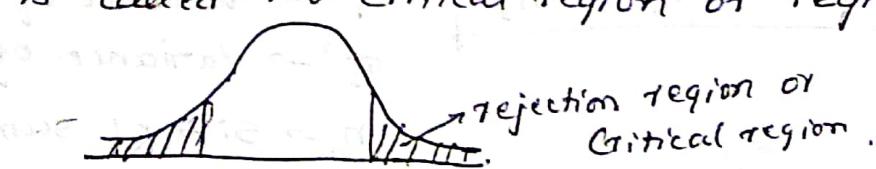
## ERRORS

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

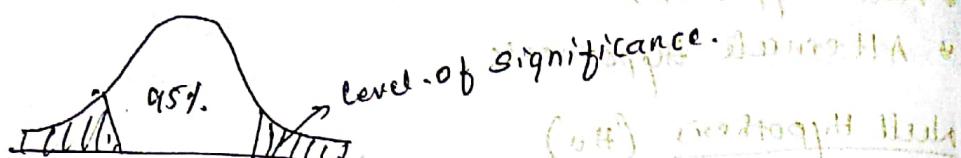
## Critical Region

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



## Significance level or Level of Significance [LOS]

The probability level below which leads to rejection is called Significance Level.  
 → Generally 1% or 5% is the significance level.



if we have (5% or 1%). Then (5% or 0.5%)  
 $(-1.96 \text{ or } +1.96)$  and  $(-2.58 \text{ or } +2.58)$

Note :- critical value of  $Z$  (two-tailed test)

at 5% LOS	at 1% LOS
+1.96 & -1.96	+2.58 & -2.58

Testing of Population mean ( $\mu$ ) is given by

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad \text{where } \bar{x} \rightarrow \text{The sample mean}$$

$\mu \rightarrow \text{population mean}$

$\sigma \rightarrow \text{S.D.}$

Testing of proportion ( $p$ )

$$Z = \frac{\bar{x} - np}{\sqrt{npq}} \quad \text{where } \bar{x} \rightarrow \text{Observed no. of success}$$

$np \rightarrow \text{expected no. of success}$

$q = 1-p \rightarrow \text{probability of success}$

Testing of mean

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \begin{aligned} \bar{x}_1 &\rightarrow \text{mean of sample 1} \\ \bar{x}_2 &\rightarrow \text{mean of sample 2} \\ \sigma_1^2 &\rightarrow \text{variance of sample 1} \\ \sigma_2^2 &\rightarrow \text{variance of sample 2} \\ n_1 &\rightarrow \text{size of sample 1} \\ n_2 &\rightarrow \text{size of sample 2} \end{aligned}$$

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  $M$ .

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

99% C.I.  $\bar{x} \pm 2.58 \left( \frac{\sigma}{\sqrt{n}} \right)$  with a loss of 99% of confidence level.

6. confidence intervals for  $(P)$ .

95% C.I.  $P \pm 1.96 \left( \sqrt{\frac{PQ}{n}} \right)$  with a loss of 95% of confidence level.

99% C.I.  $P \pm 2.58 \sqrt{\frac{PQ}{n}}$

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

### Example

1. The mean life of a sample of 100 fluorescent light bulbs produced by a company is computed to be 1570 hours with  $SD = 120$  hours. If  $\mu$  is the life time of all bulb's produced by the company, test the hypothesis  $H_0 = 1600$  hours against the alternate hypothesis  $H_a \neq 1600$  hours.

using LOS as 0.01

Ans  $LOS = 0.01$   
 $= 1.7 \cdot = 2.58$

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

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

$$|Z|_{\text{cal}} = 2.5 \cdot \left( \frac{1}{\sqrt{n}} + \frac{1}{\sigma} \right) \text{pp}$$

level of significance

$$Z_{0.01} = 2.58$$

$$|Z|_{\text{cal}} < Z_{0.01}$$

$H_0$  is accepted

2. 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  $n = 9000$ ,  $x = 3240$

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

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

$$\therefore Z = \frac{x - np}{\sqrt{npq}} = \frac{3240 - 9000 \times \frac{1}{3}}{\sqrt{9000 \times \frac{1}{3} \times \frac{2}{3}}} = 5.37$$

INSTITUTE OF TECHNOLOGY

(SOURCE DIGINOTES)

$$|Z|_{\text{cal}} = 5.37 > Z_{\text{tab}} = 2.58$$

$\therefore H_0$  is rejected

$|Z|_{\text{cal}} > Z_{\text{tab}}$   $\Rightarrow H_0$  is rejected

3. A dice was thrown 1200 times & the no. 6 is obtained 236 times, can the dice be fair at 0.01 level?

$$\text{Ans} \quad n = 1200, \quad x = 236.$$

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

$$q = 1 - \frac{1}{6} = \frac{5}{6}$$

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

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

$$|Z|_{\text{cal}} = 2.8$$

$$|Z|_{\text{tab}} = 2.58$$

$|Z|_{\text{cal}} > |Z|_{\text{tab}} \Rightarrow H_0$  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.

$$\text{Ans} \quad n = 400$$

$$P = \frac{1}{2}$$

$$\text{and } q = 1 - \frac{1}{2} = \frac{1}{2}$$

INSTITUTE OF TECHNOLOGY

(SOURCE DIGINOTES)

$$D = 5 \left( \frac{216}{400} - \frac{1}{2} \right) = 5 \times 0.05 = 0.25$$

$$\left( \frac{0.25}{0.05} \right)^{0.05} = 1.05$$

Approximate,

1.05 and 1.06

is the approx. diff. & hence P-value is approx.

$$\left( \frac{0.25}{0.05} \right)^{0.05} = 1.05 \approx 0.025$$

$$\left( \frac{0.25}{0.05} \right)^{0.05} = 1.05 \approx 0.025$$

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

$\Rightarrow T$

$$\frac{184 - 5}{\sqrt{960 \cdot 5}} = 5$$

$$5 < 3.86$$

$$5 < 3.86$$

$$5 < 3.86$$

$$5 < 3.86$$

Decision Region:  $\text{A.O.F.} \leftarrow \text{Ans}(\text{S}) < \text{Ans}(\text{L})$

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 found to be 67 pounds with the S.D. 12 pounds. What can we infer about the mean weight of population?

$$\underline{\text{A.O.F.}} \quad n = 225, \bar{x} = 67, \sigma = 12.$$

We've solve this by confidence interval for  $\mu$ .

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

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

$\therefore$  mean weight lies b/w

$$\mu = 69.064, 64.936 \quad 64.936 \leq \mu \leq 69.064$$

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

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

$$= 68.56, 65.432$$

$$\underline{68.56} < \bar{x} < \underline{65.432}$$

$\therefore$  The mean weight lies b/w  $\underline{68.56}$  &  $\underline{65.432}$ .

$$65.432 < \bar{x} < 68.56$$

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.

$$A: n = 60, \bar{x} = 01.09 \\ \sigma = 0.73$$

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

$$= \frac{0.09 - 0.37}{0.73/\sqrt{60}} = 1.96 \left( \frac{\sigma}{\sqrt{n}} \right)$$

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

$$= 1.09 \pm 2.58 \left( \frac{0.73}{\sqrt{60}} \right) = 1.09 \pm 0.96 \cdot \left( \frac{0.73}{\sqrt{60}} \right)$$

$$= 1.09 \pm 0.96 \cdot \left( \frac{0.73}{\sqrt{60}} \right) = 1.09 \pm 0.96 \cdot 0.073 = 1.09 \pm 0.070 = 1.02 \text{ to } 1.16$$

18/5/17  
Note

Suppose in the question, if ~~99%~~ confidence limit or probable limit is mentioned, then we have to solve by confidence method and go to APP.

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

sol

$$n = 900$$

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

$$q = 1 - P$$

$$q = \frac{8}{9}$$

∴ 99% CI for proportion

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

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

$$= 0.084$$

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

$$= 0.13$$

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

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?

$$n = 500$$

$$P = 0.6 \Rightarrow q = 0.4$$

We 99% of level of significance, thus we

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

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

$$= 0.6565 \quad 0.5434$$

$$95\% \text{ of } CI(P) = P \pm$$

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. Test if there is a difference b/w the mean life of the bulb produced by 2 companies at 5% LOS. LOS: 1%, LOS

A/q

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.

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

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

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

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

$\Rightarrow H_0$  is rejected.

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

A/q

$n_1 = 32$  and  $n_2 = 36$  and  $\bar{x}_1 = 72$  and  $\bar{x}_2 = 75$

And  $\sigma_1 = 8$  and  $\sigma_2 = 6$

$H_0$ : there is no significant difference b/w the performance of girls and boys

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

$$Z = \frac{72 - 75}{\sqrt{\frac{(8)^2}{32} + \frac{(6)^2}{36}}}$$

$$= \frac{-3}{\sqrt{\frac{64}{32} + \frac{36}{36}}} = \frac{-3}{\sqrt{5}}$$

$$Z_{cal} = -1.73$$

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

$$Z_{cal} < Z_{tab}$$

$\therefore H_0$  is accepted.

12. In a city A, 20% of random 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 significant? [use 1% level of significance]

Ans : 20%      900  
18.5%      1600

$$n_1 = 900$$

$$n_2 = 1600$$

$$p_1 = 0.2$$

$$p_2 = 0.185$$

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

$H_0$ : There is no significant difference b/w the proportion.

$$Z_{\text{cal}} = \frac{P_1 - P_2}{\sqrt{PQ} \left[ \frac{1}{n_1} + \frac{1}{n_2} \right]} . \quad P = \frac{900(0.2) + 1600(0.185)}{900 + 1600}$$

$$P = 0.1904.$$

$$Q = 1 - P = 1 - 0.1904 = 0.8096.$$

$$\therefore Z = \frac{P_1 - P_2}{\sqrt{PQ} \left[ \frac{1}{n_1} + \frac{1}{n_2} \right]} = 0.916.$$

$$Z_{\text{cal}} = 0.916, \quad Z_{\text{tab}} \text{ at } 1\% \alpha = 2.58.$$

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

$\therefore H_0$  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 samples were from payment quota. Does the data reveal a significant difference b/w two cities w.r.t payment quota students?

$$\begin{array}{ll} A & B \\ \underline{n_1 = 1000} & \underline{n_2 = 800} \end{array}$$

400 - payment seats

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

$H_0$ : There is no significant diff b/w two cities

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

$$\frac{q_1 - q_2}{\sqrt{\frac{q_1(1-q_1)}{n_1} + \frac{q_2(1-q_2)}{n_2}}} \sim Z$$

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

$$P = 4/9.$$

$$\Rightarrow q = S/P = 5/9$$

$$\therefore Z = \frac{1000 - 800}{\sqrt{4/9 \times 5/9 \left( \frac{1}{1000} + \frac{1}{800} \right)}} = 2.0$$

$$Z_{cal} = 2.43$$

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

$$Z_{cal} > Z_{tab}$$

$\Rightarrow H_0$  is rejected.

14. A Random Sample of 1000 workers in a company has mean wage of 50 Rs/day and SD 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 two companies.

Sol

$$n_1 = 1000$$

$$\bar{x}_1 = 50$$

$$\sigma_1 = 15$$

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

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

$$5 \pm 1.96 \sqrt{0.225 + 0.2666} = 3.62 \text{ E. 6.3}$$

# t-distribution & Chi-Square test

student  $\cdot t \sim$  distribution [t-test]

$$\textcircled{1} \quad t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad s = S.D.$$

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

$$\textcircled{2} \quad t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$

$s_1$  is the S.D of 1<sup>st</sup> sample

$s_2$  — u — 2<sup>nd</sup> sample.

$\bar{x}_1$  is mean of 1<sup>st</sup> sample

$\bar{x}_2$  — II — 2<sup>nd</sup> sample

$n_1$  — size of 1<sup>st</sup> sample.

$n_2$  — size of 2<sup>nd</sup> sample.

(3) 95% Confidence limits for ( $\mu$ ):

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

Note :-

Degree of freedom. [df]

Degree of freedom is nothing but

no of values — 1

$$[df] = n - 1$$

## Problems

- 10 individuals are chosen 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  $[t_{0.05} = 2.262 \text{ for } 9 \text{ df}]$

Ans

Given  $n = 10$

~~so~~  $\bar{x} = 66$  inches.

$$\bar{x} = \frac{\sum x}{n} = \frac{63 + 63 + 66 + 67 + 68 + 69 + 70 + 70 + 71 + 71}{10} \\ = 67.8$$

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

$$63 \cdot 23.04$$

$$63 \cdot 23.04$$

$$66 \cdot 3.24$$

$$67 \cdot 0.64$$

$$68 \cdot 0.04$$

$$69 \cdot 1.04$$

$$70 \cdot 4.84$$

$$70 \cdot 4.84$$

$$71 \cdot 10.24$$

$$71 \cdot 10.24$$

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

$$= 81.6$$

$$s^2 = \frac{1}{n-1} \sum (x - \bar{x})^2 \\ = \frac{1}{9} (81.6) = 9.067$$

$$\therefore s = 3.011$$

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

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

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

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

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

Given  $n = 12$ ,  $\bar{x} = 0.58$

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

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

$$s^2 = \frac{1}{11} \sum (x - \bar{x})^2$$

$$s^2 = \frac{1}{11} (104.908)$$

$$s^2 = 9.5370$$

$$s = \sqrt{9.5370} = 3.088$$

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

$$t = \frac{2.58}{0.894}$$

$$t = 2.894$$

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

$$t_{\text{cal}} = 2.894$$

$$t_{\text{tab}} = 2.201$$

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

$$t_{\text{cal}} = 2.894$$

$$t_{\text{tab}} = 2.201$$

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

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

$$[t_{0.05} = 2.201 \text{ for } 9 \text{ df}]$$

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

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

95% of confidence limits

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

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

Estimate of  $\mu$  is  $11.895$

$11.895 < \mu < 12.104$

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 24 d.f.]

Sol  $n = 25$

$$\bar{x} = 3.1 \quad t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{3.1 - 3}{0.3/\sqrt{25}} = \frac{0.1}{0.06} = 1.67$$

$$\sum (x_i - \bar{x})^2 = 0.025$$

5. A sample of 11 rats 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  $\mu$ , the mean viscosity of the central population [ $t_{0.05} = 2.228$  for 10 d.f.]

Sol  $n = 11$

$$\bar{x} = 3.92 \quad S = 0.61$$

$$\bar{x} \pm t_{0.05} [S/\sqrt{n}] = 3.92 \pm 2.228 \times \frac{0.61}{\sqrt{11}} = 3.92 \pm 0.42$$

6. A group of boys and girls where girls are intelligent 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

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

$H_0$ : there is no significant difference b/w girls and boys.

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

$$S_1 = 8 \quad S_2 = 10$$

$$n_1 = 12 \quad n_2 = 10$$

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

$$S^2 = \frac{12 \times (8)^2 + 10 \times (10)^2}{12 + 10} \\ = 80.36$$

$$S = \sqrt{80.36} \\ = 8.964$$

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

$$t_{tab} = 2.086$$

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

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

2015/17

### Chi-square test

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

Where  $O_i$  is observed frequency

$E_i$  is expected frequency

$\chi^2$  - (chi)

$$\sum O = N$$

$$F(x) = N \cdot P(x)$$

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

$$\chi^2_{0.05} = 7.815 \text{ for } 3 \text{ d.f}$$

$x$	0	1	2	3	4
$O = f$	122	60	15	2	1
$E$	121.301	60.65	15.16	2.52	0.31
$O - E$	-1.301	-0.65	-0.16	-0.52	-0.31
$(O - E)^2 / E$	0.08	0.01	0.01	0.02	0.01

$$E = f ?$$

$$N = \sum f = 200$$

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

$$m = \frac{\sum f \cdot x}{\sum f} = \frac{60 + 30 + 6 + 4}{200} = \frac{100}{200} = 0.5$$

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

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

$$\text{at } x=0 \quad E \rightarrow F = \frac{200 \cdot (0.5)^0 \cdot e^{-0.5}}{0!} = 200 \cdot 1 \cdot e^{-0.5}$$

$$\cancel{\frac{200}{1!}} = \text{at } x=0 \\ = \frac{200}{1!} \cdot (0.5)^0 \cdot e^{-0.5}$$

$$\text{At } x = 1$$

$$\frac{d^2O}{dx^2} (0.5) \cdot e^{-0.5} = 60.65$$

$$\sum \frac{(O-E)^2}{E} = \frac{(122 - 120.65)^2}{120.65} = 6.2399 \times 10^{-3}$$

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

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

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

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

$$\sum \frac{(0 - 0.05)^2}{0.05} = 1.578$$

$$\sum \frac{(O-E)^2}{E} = 6.2399 \times 10^{-3} + 6.966 \times 10^{-3} + 1.688 \times 10^{-3} + 0.0277 + 1.5358 + 1.578 = 1.578$$

$$Q^2_{\text{cal}} < Q^2_{\text{tab}}$$

Hypothesis is accepted.

INSTITUTE OF TECHNOLOGY

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

$$\chi^2_{0.05} = 9.49 \text{ for } 4 \text{ d.f.}$$

$O \rightarrow f$	1	2	3	4
0 → f	419	352	154	56

$E$	404.94	366.071	165.46	49.85	11.268
$\frac{(O-E)^2}{E}$	0.4881	0.540	0.7934	0.758	5.3058

$$N = \sum f = 1000$$

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

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

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

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

$$\text{At } x=0, 1000 (1) e^{-0.904} = 404.94$$

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

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

$$x=3, \frac{1000}{3!} (0.904)^3 e^{-0.904} = 49.8591$$

$$x=4, 11.268 \dots$$

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

$$Y_{\text{cal}} < Y_{\text{tab}}$$

∴ Null hypothesis is accepted.

### INSTITUTE OF TECHNOLOGY

3. Four coins were tossed 160 times and the following result were obtained. Test the goodness of fit of the binomial distribution.

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

$$\left[ \chi^2_{0.05} = 9.49 \text{ for } 4 \text{ df} \right]$$

$$n=4, P = \frac{1}{2}, Q = \frac{1}{2}$$

$$N=160, P = \frac{1}{2}, Q = \frac{1}{2} \quad \left\{ \begin{array}{l} \text{Probability of tossing a coin.} \\ \text{Exptd. freq.} \end{array} \right.$$

Binomial Distribution

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

$$P(x) = 4 C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{4-x} = \frac{1}{2^4} \cdot 4 C_x$$

$$E = NP(x) = \frac{160}{16} \cdot 4 C_x = 104 C_x$$

$$\textcircled{1} x = 0, 1, 2, 3, 4$$

$$O = 17, 52, 54, 31, 6$$

$$B \rightarrow E = 10, 40, 60, 40, 10$$

$$\frac{(O - E)^2}{E} = \frac{21}{40}, \frac{18}{5}, \frac{3}{5}, \frac{81}{40}, \frac{6}{5}$$

$$\sum \frac{(O - E)^2}{E} = \frac{324}{40} = 8.1 \approx 12.725$$

$\chi^2_{\text{cal}} > \chi^2_{\text{tab}}$   
 $\Rightarrow$  Hypothesis is rejected.

5 dice were thrown 96 times and the numbers showing 1 or 2 or 3 appearing on the face of the dice follows the frequency distribution as follows

$$\left[ \chi^2_{0.05} = 9.49 \text{ for 4 df} \right]$$

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

8. Test the hypothesis that the data follows a binomial distribution  $\left[ \chi^2_{0.05} = 9.49 \text{ for 5 df} \right]$

$$\eta = .5 \quad N = 96$$

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

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

Binomial Distribution

$$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

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

$$= \frac{1}{32} {}^5 C_3 x^3 (1-x)^2$$

$$E = Np(x) = \frac{96}{32} {}^5 C_3 = 3 ({}^5 C_3)$$

$x$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32								
$E$	0	1.5	3	4.5	6	7.5	9	10.5	12	13.5	15	16.5	18	19.5	21	22.5	24	25.5	27	28.5	30	31.5	33	34.5	36	37.5	39	40.5	42	43.5	45	46.5	48	49.5	51	52.5	54	55.5	57	58.5	60

$$\frac{(O-E)^2}{E} = 5.333 \quad 1.066 \quad 0.833 \quad 1.2 \quad 3.26 \quad 0$$

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

STATISTICAL INSTITUTE OF TECHNOLOGY

$\chi_{\text{cal}}^2 > \chi_{\text{tab}}^2 \Rightarrow \text{hypothesis is rejected.}$

Source diginotes.in

binomial distribution  
discrete probability distribution  
discrete random variable

Source diginotes.in