

Statistical Decision Theory

- ▷ **Statistical Decision** :- Very often in practice we are called upon to make decision about population on the basis of sample information, such decision called statistical decision.
- ▷ **Statistical Hypothesis** :- In attempting to reach decisions, it is useful to make assumption (guess) about the population, such assumption, which may or may not be true are called statistical Hypothesis.

Type of Hypothesis:

i) **Null Hypothesis** :- If we want to decide whether one procedure is better than others, so we formulate the hypothesis that there is no difference between the procedure, such hypothesis is called as null hypothesis and it is denoted by (H_0).

ii) **Alternative Hypothesis** :- Any hypothesis that differs from the given hypothesis is called an Alternative Hypothesis.

For Example:- If one hypothesis is $P=0.5$ (H_1).

In H_1 , $P=0.7$, $P \neq 0.5$ or $P > 0.5$.

A hypothesis alternative to null hypothesis is denoted by H_1 .

* **Test of Hypothesis and Significance** :- If we suppose that a particular hypothesis but find that results of the observed in a random sample differ from the result expected under the hypothesis, then we would say that the observe difference are significant and reject the hypothesis.

* Simple and Composite hypothesis:

A statistical hypothesis that completely specifies the population parameter is called a simple hypothesis, and the hypothesis that does not specify the population parameter is called a composite hypothesis.

Example: If x_1, x_2, \dots, x_n is a random sample from normal with mean μ and variance σ^2 then $H_0: \mu = \mu_0$ and $\sigma^2 = \sigma_0^2$ is a simple hypothesis. The full fully not specified hypothesis is called a composite hypothesis.

$H_1: \mu \neq \mu_0$ 2) $H_0: \sigma^2 \neq \sigma_0^2$ 3) $H_0: \mu = \mu_0$ and $\sigma^2 > \sigma_0^2$ etc.

* Errors in the test of significance: The main objective in the sampling theory is to draw a valid inference about the population parameters based on sample results.

i) Type-I error: Rejecting H_0 when H_0 is true.

ii) Type-II error: Accepting H_0 when it is false (Accepting H_0 when H_1 is true).

iii) Size of Type-I and Type-II errors:

$$P[\text{Reject } H_0 \text{ when it is true}] = \alpha$$

$$P[\text{Accept } H_0 \text{ when it is wrong}] = \beta$$

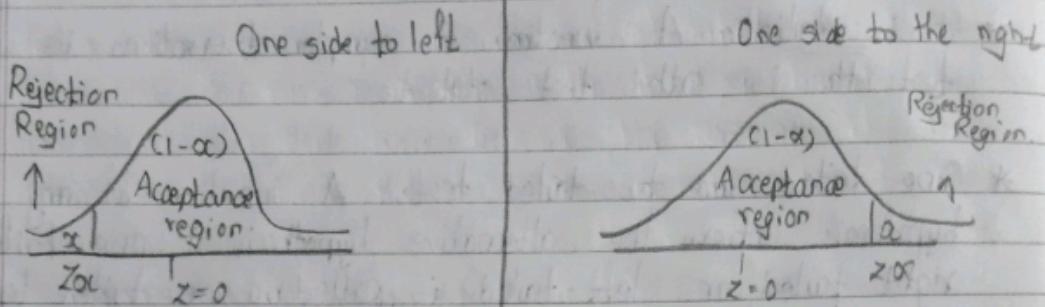
* Critical region: A region in sample space S which amounts to a rejection of H_0 is called a critical region or rejection region of H_0 .

* Level of significance: The probability ' α ' is that the value of the test statistic belongs to the critical region, known as 'level of significance'. That is the probability of occurrence of the type I error is the level of significance. Usually, we use the level of significance of 5% or 1%.

- * Test statistic: A function of sample observations is used to test H_0 is called test statistic.
- * One-tailed and two-tailed tests: A function of any statistical hypothesis where the alternative hypothesis is one-tailed (right-tailed or left-tailed) is called a one-tailed test.
 Eg: A test for testing the mean of a population $H_0: \mu = \mu_0$ versus $H_1: \mu > \mu_0$ (Right Tailed) or $H_1: \mu < \mu_0$ (Left Tailed), is a one-tailed test.
- * Central limit theorem: In many cases, the exact probability distribution of the test statistics T cannot be obtained. The difficulty is overcome using the normal approximation. The probability distribution of standardized T is assumed to be $N(0,1)$ as the sample size $n \rightarrow \infty$ (i.e., n is sufficiently large). The corresponding theorem in support of the normal approximation is known as the central limit theorem.

$$\therefore \left[Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1) \right]$$

- * Critical value: The value of the test statistic, which separates the critical (rejection) region and acceptance region, is called the 'critical value'. It depends upon (1) The level of significance α used and (2) The alternative hypothesis, whether it is two-tailed or one-tailed.



(1.) Describe the hypothesis H_0 & Alternate hypothesis H_1

- Given a die with six number 1, 2, 3, 4, 5, 6 on its faces. Test whether the die is unbiased or biased.

$$H_0: P(x) = \frac{1}{6}, \text{ if } x = 1, 2, 3, 4, \dots, 6.$$

$$H_1: P(x) \neq \frac{1}{6}, \text{ if } x = 1, 2, \dots, 6.$$

- The new process technology manufacturer guarantees increase in mean production atleast by 5% and the benefits of doubts is to be given to him.

$$H_0: \mu \geq 0.05$$

$$H_1: \mu < 0.05$$

Ex:- For a certain test of coin, $H_0: p = 1/2$ against $H_1: p = 1/3$ can be retained where P represent the probability of getting a tail. To decide this coin is tossed 4 times & H_0 reject only if the no. of head observe 0 or 1. Find the probability of both type of Error.

Let X : Number of tails observed in 4 tosses

$$H_0: p = \frac{1}{2}$$

$$H_1: p \neq \frac{1}{2} \quad (\text{i.e. } p = \frac{1}{3})$$

$$n=4, p=\frac{1}{2}, q=1-p=\frac{1}{2}$$

$$P(X) = \frac{n!}{x!(n-x)!} p^x q^{n-x} \dots \text{Binomial Distribution Formula}$$

$$= \frac{4!}{x!(4-x)!} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{4-x}, \quad x=0,1,2,3,4$$

$$P(\text{Type I-error}) = P(H_0 \text{ is rejected when } H_0 \text{ is true})$$

$$= P(X=0 \text{ or } 1 \text{ when } P)$$

$$= P(0) + P(1)$$

$$= {}^4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{4-0} + {}^4C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^3$$

$$= 1 \times 1 \times \frac{1}{16} + 4 \left(\frac{1}{16}\right) = \frac{1+4}{16} = \boxed{\frac{5}{16}}$$

$$P(\text{Type II-error}) = P(H_0 \text{ is accepted when } H_0 \text{ is false})$$

$$= P(X=2 \text{ or } 3 \text{ or } 4 \mid H_1 \text{ is True})$$

$$= P(2) + P(3) + P(4)$$

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

$$= {}^4C_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^2 + {}^4C_3 \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^1 + {}^4C_4 \left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)^0$$

$$= 6 \left(\frac{1}{9}\right)\left(\frac{4}{9}\right) + 4 \left(\frac{1}{27}\right)\left(\frac{8}{27}\right) + 1 \left(\frac{1}{81}\right)\left(\frac{16}{81}\right)$$

$$= 6 \times \frac{4}{81} + 4 \times \frac{8}{729} + 1 \times \frac{16}{6561}$$

$$= \frac{24}{81} + \frac{32}{729} + \frac{16}{6561}$$

$$= \frac{38}{81} = \boxed{\frac{11}{27}}$$

- Q.) In a box there are 3 bangles of which ($m > 0$) are red. Two test $H_0: m \leq 2$ against $H_1: m > 2$, 1 bangle is drawn at random from box & H_0 rejected if the drawn is red. Find the level of significance for the test.

$$H_0: m \leq 2 \quad (m=1 \text{ or } 2)$$

$$H_1: m > 2$$

This is a composite hypothesis.

Level of significance = max $m.p(H_0)$ (when it is true).

$P(H_0 \text{ is reject when } m=1)$

$$= \frac{{}^1C_1 \times {}^2C_0}{{}^3C_1} = \frac{1 \times 1}{3} = \boxed{\frac{1}{3}}$$

Also, $P(H_0 \text{ is reject when } m=2)$

$$= \frac{{}^2C_1 \times {}^1C_0}{{}^3C_1} = \frac{2 \times 1}{3} = \boxed{\frac{2}{3}}$$

Level of significance = $\max\left(\frac{1}{3}, \frac{2}{3}\right)$

$$= \boxed{\frac{2}{3}}$$

- Q.) In box there are B balls in which n balls are red, it is known that no. of red balls n is either three or four but the exact no. is not known. However it is $n=3$ is much more likely than $n=4$. To take a decision on the value of n , two balls are drawn & $n=3$ is rejected only if both the balls are drawn red. Find H_0, H_1 ,

③ State whether the hypothesis is simple or composite

④ Define critical region.

⑤ Compute the size of error of both.

⑥ Level of Significance

$$\Rightarrow ① H_0 : n=3$$

$$② H_1 : n=4$$

③ One value is specified

\therefore It is a simple hypothesis.

④ Decision Criteria

Reject H_0 if $x=2$, otherwise do not reject.

⑤ Type I error: $P(\text{Type I error}) = P(\text{Reject } H_0 \text{ when } H_0 \text{ is true})$

$= P(x=2 \text{ when } n=3)$

$$= \frac{3C_2}{6C_2} = \frac{6}{35} = \boxed{\frac{1}{5}}$$

$P(\text{Type II error}) = P(\text{Accept } H_0 \text{ when } H_1 \text{ is true})$

$= P(x \neq 2 \text{ when } n=4)$

$= P(x=0 \text{ or } x=1 \text{ when } n=4)$

$$= \frac{4C_0 \cdot {}^2C_2}{6C_2} + \frac{4C_1 \cdot {}^2C_1}{6C_2}$$

$$= \frac{1 \times 1}{15} + \frac{4 \times 2}{15}$$

$$= \frac{1}{15} + \frac{8}{15}$$

$$= \boxed{\frac{3}{5}}$$

H_1	Decision criteria repeat H_0 iff	
	$\alpha = 0.05$	$\alpha = 0.01$
>	$t > 1.64$	$t > 2.33$
<	$t < -1.64$	$t < -2.33$
\neq	$t > 1.96 \text{ or}$ $t < -1.96$	$t > 2.38 \text{ or}$ $t < -2.58$
		$t > t_{\alpha/2} \text{ or}$ $t < -t_{\alpha/2}$

Sr. no.	Test	Notation (Sample)	H_1	Test of statics
1)	One population mean	n : size \bar{x} : mean μ : population mean s : s_d , σ : p.s.d	H_1 H_1 H_1 H_1	$t = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$
2)	One population proportion	n : sample size p : sample population proportion P : population proportion	$p > p_0$ $p < p_0$ $p \neq p_0$	$t = \frac{p - p_0}{\sqrt{p_0(1-p_0)}} / n$
3)	Two population mean	n_1, n_2 : Sample size \bar{x}_1, \bar{x}_2 : Sample mean μ_1, μ_2 : population mean σ_1, σ_2 : population S.D.	$H_1 - H_2 > 0$ $H_1 - H_2 < 0$ $H_1, H_2 \neq 0$	$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

- Steps :-
- 1) State H_0 & H_1 ,
 - 2) level of significance (α)
 - 3) Mention decision criteria
 - 4) Find test statistic
 - 5) Conclusion.

Q1) The mean lifetime of samples of 100 light bulbs produced by company is found to be 1570 hrs with S.D 120 hrs. Test the hypothesis that the mean lifetime bulbs produced by the company is 1600 hrs. The alternative hypothesis is less than at 5% level of significance.

→ Given:- $n = 100$, $\bar{x} = 1570$, $\sigma = 120$, $H_0 = 1600$

$$\text{Step 1: } H_0 : \mu \geq 1600$$

$$H_1 : \mu < 1600$$

$$\text{Step 2: LOS, } \alpha = 5\% = 0.05$$

Step 3:- Decision criteria :-

$$\text{Reject } H_0 \text{ iff } t < -1.64$$

$$\begin{aligned} \text{Step 4:- Test statistic: } t &= \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{1570 - 1600}{120 / \sqrt{100}} \\ &= \frac{-30}{120 / 10} = \frac{-5}{2} = \boxed{-2.5} \\ \therefore t &= \boxed{-2.5} \end{aligned}$$

Step 5:- Conclusion:-

$$\text{Since: } -2.5 < -1.64$$

Reject H_0 at 5% LOS.

Q.2) A certain coin is showed up head 270 occasions in 500 tosses test the claim that coin is unbiased.

→ Step 1:- $H_0 : p = 0.5$
 $H_1 : p \neq 0.5$

Step 2:- Level of significance is not given:
∴ we assume 5% LOS / $\alpha = 0.05$

Step 3:- Decision criteria:-

Reject H_0 iff $t > 1.96$ or $t < -1.96$

Step 4:- Test statistics:-

$$P = \frac{270}{500} = 0.54, P_0 = 0.5, Q_0 = 0.5, n = 500$$

$$\therefore t = \frac{P - P_0}{\sqrt{\frac{P_0 Q_0}{n}}} = \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{500}}} \approx 1.80$$

$$\therefore \boxed{t = 1.80}$$

Step 5:- Conclusion:-

since, $1.80 < 1.96$

$$t < t\alpha$$

∴ Accept H_0 at 5% LOS.

Q.3) A random sample of 100 taken from the items produced by a certain manufacturing process have mean length of 2.5 cm with S.P of 0.05 cm. It was necessary to decrease the length of the items and a group of engineers applied some modification claiming to achieve success. A Random sample then drawn gave a mean of 2.48 cm with $\sigma = 0.06$ cms. We are ready to accept the success of the engineers only if the above observation indicates so. what is our conclusion at 5% LOS.

$$\Rightarrow n = 100, \bar{x}_1 = 2.5, \sigma_1 = 0.05, \alpha = 0.05, \bar{x}_2 = 2.48, \sigma_2 = 0.06$$

$$\begin{aligned} \text{Step 1: } H_0: \mu_1 &< \mu_2 \\ \mu_1 - \mu_2 &< 0 \\ H_1: \mu_1 - \mu_2 &\geq 0 \end{aligned}$$

$$\begin{aligned} \text{Step 2: } L.O.S, \alpha &= 5\% \\ \alpha &= 0.05 \end{aligned}$$

$$\begin{aligned} \text{Step 3: Decision criteria} \\ \text{Reject } H_0 \text{ iff } |t| > 1.64 \end{aligned}$$

Step 4:- Test statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{2.5 - 2.48}{\sqrt{\frac{(0.05)^2}{100} + \frac{(0.06)^2}{100}}} = 2.56$$

$$\begin{aligned} \text{Step 5:- Conclusion: } t &> 1.64 \\ 2.56 &> 1.64 \end{aligned}$$

H_0 is reject at 0.05% level of significance.

The Chi-Square Test

- * Degrees of freedom: The number of degrees of freedom of a statistic, generally denoted by v , is defined as the number N of independent observations in the sample (i.e. the sample size) minus the number k of population parameters, which must be estimated from sample observations.
In symbols, $v = N - k$.

Eg:- The number of independent observations in the sample is N , from which we compute \bar{x} & s . However, since we must estimate $M, k=1$ and $v=N-1$.

- * Introduction:- If x is a standard normal variate (a continuous random variable) so that its probability density function is given by:-

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, -\infty < x < \infty = 0, \text{ otherwise.}$$

then $y = x^2$ is also a continuous random variable whose probability density function is given by,

$$g(y) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-\frac{y}{2}}, y \geq 0$$

$$= 0, \text{ otherwise.}$$

Here, the distribution of y is known as Chi-Square distribution with one degree of freedom.

- * Properties of Chi-Square variate with n degrees of freedom:

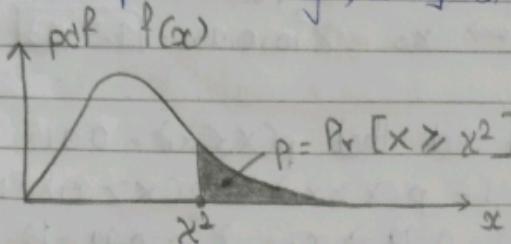
Let χ^2 denote a Chi-square variate with n degrees of freedom.
Then we have the following properties:-

- The mean of χ^2 is $E(\chi^2) = n$

- The variance of χ^2 is $V(\chi^2) = 2n$

- The mode of $\chi^2 = n - 2$

- The frequency curve is given by $y = f(\chi^2)$ lies in the first quadrant and it is positively skewed curve, its tail on the right extends upto infinity, as given in the below diagram:



$p = \Pr [X \geq x^2]$ is denoted by α and $x^2 = \chi^2_{\alpha}$

- The total area under the Chi-square curve is 1.

- $p(\chi^2 > c) = \text{area under the curve } y = f(\chi^2) \text{ to the right of } \chi^2 = c$

- For a chi square variate with n degrees of freedom, if $p(\chi^2 > c) = \alpha$, then c is denoted by $\chi^2_{\alpha, n}$, i.e. $p(\chi^2 > \chi^2_{\alpha, n}) = \alpha$

- $\chi^2_{\alpha, n}$ is called a probability point of chi square distribution with n degree of freedom.

Q1) If a random variable X follows chi square distribution with 10 degrees of freedom such that $p(X > x_0) = 0.95$,
 $p(X \leq x_1) = 0.01$ & $p(X > 18.3) = \alpha$
 find (i) x_0 (ii) x_1 (iii) α

$\Rightarrow n = 10 - \text{degree of freedom}$

i) To find x_0 such that $p(X > x_0) = 0.95$
 $\Rightarrow x_0 = \chi^2_{10, 0.95} = \boxed{3.9403}$

ii) to find x_1 such that $p(X \leq x_1) = 0.01$

$$p(X \leq x_1) = 1 - p(X > x_1) \Rightarrow p(X > x_1) = 1 - 0.01 = 0.99$$

$$\Rightarrow p(X > x_1) = 0.99 \Rightarrow x_1 = \chi^2_{10, 0.99} = \boxed{12.5582}$$

iii) $p(X > 18.3) = \alpha \Rightarrow \chi^2_{10, \alpha} = 18.3 \Rightarrow \alpha = \boxed{0.05}$

Q2) A random variable Y follows chi square distribution with S.D. = 4, find y_0 if $p(Y \leq y_0) = 0.05$

$\Rightarrow \therefore S.D = 4 / \therefore \text{var} = 16 \Rightarrow 2n = \boxed{16}$
 $n = \boxed{8}$

$$p(Y \leq y_0) = 0.05 \Rightarrow p(Y \leq y_0) = 1 - p(Y > y_0)$$

$$= 1 - 0.05$$

$$= \boxed{0.95}$$

$$\therefore p(Y > y_0) = 0.95 \Rightarrow y_0 = \chi^2_{8, 0.95}$$

$$y_0 = \boxed{2.7326}$$

Q.3) If a random variable x follows Chi-square distribution with S.D. = 4, find mean and mode.

$$\Rightarrow \text{S.D.} = 4 / \text{Var} = 16 / \chi^2 = 2n \\ = 2(8) = 16 \\ \therefore n = 8$$

∴ Mean of $\chi^2 = n = 8$

∴ Mode of $\chi^2 = n - 2 = 6$

* The Chi-square test for goodness of fit: - It involves a test statistic that can be shown to follow Chi-square distribution under assumptions and the test to is known as Chi-square test for goodness of fit.

Suppose we have an observed frequency distribution with n classes having observed frequencies $O_1, O_2, O_3, \dots, O_n$ with $\sum_{i=1}^{i=n} O_i = N$.

Further, suppose that according to our assumption to be called the null hypothesis H_0 , the expected frequencies are $E_1, E_2, E_3, \dots, E_n$ such that $\sum_{i=1}^{i=n} E_i = N$.

Then the test statistic proposed by Karl Pearson is given by $\chi^2 = \sum_{i=1}^{i=n} \frac{(O_i - E_i)^2}{E_i}$

* Decision Criterion:

While performing a test for goodness of fit, we shall use the following decision criteria:-

Reject H_0 if $\chi^2 = \sum_{i=1}^{10} \frac{(O_i - E_i)^2}{E_i} > \chi^2(n-1), \alpha$

Do not reject H_0 i.e. accept H_0 if $\chi^2 \leq \chi^2(n-1), \alpha$

- Q.1) The following data represents the last digit of the cars passing at a certain traffic signal observed during last 30 minutes for 180 cars.

Last digit	0	1	2	3	4	5	6	7	8	9
frequency	12	20	14	10	21	18	17	26	19	21

Can we retain at 5% level of significance that all the digits are equally likely to occur?

\Rightarrow If all digits are equally likely, $p = 1/10, E_i = Np = 180 \times 1/10 = 18$

H_0 : All digits are equally likely to occur.

H_1 : not H_0 i.e. logical negation to H_0 i.e. all digits are not equally likely

LOS :- 5% i.e. $\alpha = 0.05, N = 180 \geq 50$ and $E_i = 18 \leq 5, n = 10$

$$\chi^2 = \sum_{i=0}^{9} \frac{(O_i - E_i)^2}{E_i}$$

Decision criteria:- Reject H_0 if $\chi^2 > \chi^2_{9, 0.05}$ where $\chi^2_{9, 0.05} = 16.91$

Do not reject H_0 if $\chi^2 \leq \chi^2_{9, 0.05}, \chi^2 = \sum_{i=0}^{9} \frac{(O_i - E_i)^2}{E_i}$

$$1. \chi^2 = 9.777 < 16.9190 = \chi^2_{9, 0.05} \text{ do not reject H}_0$$

Digit	Observed frequency	Exp. freq	$(O_i - E_i)^2 / E_i$
0	12	18	-6
1	20	18	2
2	14	18	-4
3	12	18	-6
4	21	18	3
5	18	18	0
6	17	18	-1
7	26	18	8
8	19	18	1
9	21	18	3
			total (9.778)

Q2) As per Mendel's theory according to the shape and color, certain variety of pea that can be classified into four categories Round and yellow, Round and green, Angular and yellow, Angular and green occur in the proportion of 9:3:3:1. To test this a sample of $N=128$ peas was taken and the foll were the observed frequencies.

RY - 66	RG - 28
AY - 29	AG - 5

Perform the chi-square test for goodness of fit.

→ $N=128$, $n=4$. The probability of occurrence and Expected frequencies are given by.

category	P_i	$E_i = N p_i$
RY	9/16	$128 \times 9/16 = 72$
RG	3/16	$128 \times 3/16 = 24$
AY	3/16	$128 \times 3/16 = 24$
AG	1/16	$128 \times 1/16 = 8$

$\therefore H_0$:- The four categories of peas i.e. RY, RG, AY, AG have expected frequencies 72, 24, 24, 8 resp.

$\therefore H_1$:- not H_0

$$L.O.S = 5\%, \alpha = 0.05$$

Precision criteria is given by Reject H_0 if $\chi^2 > \chi^2_{3, 0.05} = 7.8147$

Do not reject H_0 if $\chi^2 \leq 7.8147$, where $\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$

Category of pea	Obs. freq. (O_i)	Exp. freq. (E_i)	$O_i - E_i$	$\frac{(O_i - E_i)^2}{E_i}$
RY	66	72	-6	0.5
RG	28	24	4	0.6667
AY	29	24	5	1.0417
AG	5	8	-3	1.125
			Total	3.3334

$$\chi^2 = 3.3334 < \chi^2_{3, 0.05} = 7.8147$$

\Rightarrow Accept H_0 at 5% L.O.S.

(Q.3) Four identical coins are tossed 100 times and the foll results are obtained:-

no. of heads (x)	0	1	2	3	4
frequency	8	29	40	19	4

Are there sufficient evidences to conclude that the coins are biased at 5% L.O.S.

\rightarrow Let p denote the probability of getting a head with each of the four coins,

X : no. of heads follows binomial distribution with $n=4, p$

$X \sim B(n, p)$ i.e. $X \sim B(4, p)$

H_0 : the coins are unbiased i.e. $p = 1/2$

H_1 : the coins are biased i.e. $p \neq 1/2$

LOS = 5%, $\alpha = 0.05$

$$X \sim B(4, p) \Rightarrow p(x) = \binom{4}{x} p^x 2^{4-x}$$

$$\therefore p(0) = p^0 2^4 = 1/16$$

$$\therefore p(1) = 4p^1 2^3 = 4/16$$

$$\therefore p(2) = 6p^2 2^2 = 6/16$$

$$\therefore p(3) = 4p^3 2^1 = 4/16$$

$$\therefore p(4) = p^4 2^0 = 1/16$$

Decision criteria :- Reject H_0 if $\chi^2 > \chi^2_{4, 0.05} = 9.4877$

Expected frequencies $E_i = N \cdot p(i)$

x	obs. freq. O_i	Exp. freq. E_i	$O_i - E_i$	$\frac{(O_i - E_i)^2}{E_i}$
0	8	100 $p(0) = 6.25$	1.75	0.49
1	29	25	-4	0.64
2	40	37.5	2.5	0.1667
3	19	25	-6	1.44
4	4	6.25	-2.25	0.81
		total		3.5467

$$\chi^2 = 3.5467 < \chi^2_{4, 0.05} = 9.4877$$

Accept H_0 at 5% LOS \Rightarrow coins are unbiased.

O.4} The random variable X denotes the number of street accidents per week.

X	0	1	2	3	4	5	6	7
obs. freq.	15	30	28	14	8	4	0	1
exp. freq.	14	27	27	18	9	4	1	0

Test whether the random variable X follows Poisson distribution with parameter $m=2$ at 1% level of significance.

→ As the expected frequencies is less than 5 for $x=5, 6 \& 7$ we combine them into one class.

X	0	1	2	3	4	5-7
obs. freq.	15	30	28	14	8	5
exp. freq.	14	27	27	18	9	5

H_0 :- X follows Poisson distribution with parameter $m=2$

H_1 :- not H_0

L.S. :- 1% $\alpha = 0.01$

Decision criteria is given by Reject H_0 if $\chi^2 > \chi^2_{5, 0.01} = 15.086$

Do not reject H_0 if $\chi^2 \leq \chi^2_{5, 0.01}$, where $\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$

i	X	obs.freq. O_i	exp.freq. E_i	$(O_i - E_i)$	$\frac{(O_i - E_i)^2}{E_i}$
1	0	15	14	1	0.0714
2	1	30	27	3	0.3333
3	2	28	27	1	0.0370
4	3	14	18	-4	0.8889
5	4	8	9	-1	0.1111
6	5-7	5	5	0	0
				total	1.4417

$$\therefore \chi^2 = 1.4417 < \chi^2_{5, 0.01} = 15.086$$

∴ Accept H_0 i.e. at 1% level of significance hypothesis that the variable X follows Poisson distribution with parameter $m=2$ is retainable.

* Test for independence of attributes: On the basis of data regarding two attributes for some units from the population, we shall investigate whether the observed data provide sufficient reasons to reject the claim that the two attributes are independent of each other for the population under consideration. Such a test is called test for independence of attributes.

2×2 contingency table is given as below:-

		Attribute B		total
		B1	B2	
Attribute A	A1	a	b	a+b
	A2	c	d	c+d
		a+c	b+d	a+b+c+d = N

With the help of the test statistic $\chi^2 = \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$ we can perform a test for testing the hypothesis, H_0 : the attributes A and B under the consideration are independent against the logical alternative

H_1 : not the H_0 i.e. the attribute A and B are dependent, subject to the conditions $N \geq 50$ and each of the observed frequencies $a, b, c, d \geq 5$.

The decision criteria at level of significance α is given by
 Reject H_0 : if $\chi^2 > \chi^2_{1,\alpha}$, do not reject H_0 ; if $\chi^2 \leq \chi^2_{1,\alpha}$

O.1) The foll results are obtained at the end of six months of a kind of psychotherapy given to a group of 120 patients and also for another group of 120 patients who were not given the psychotherapy.

		psychotherapy	
		given	not given
condition improved	improved	71	42
	did not improve	49	78

Can we conclude at 5% LOS that the psychotherapy is effective?

$\Rightarrow H_0$: psychotherapy is not effective

H_1 : psychotherapy is effective.

LOS = 5%, $\alpha = 0.05$

Decision criteria, Reject H_0 if $\chi^2 > \chi^2_{1,0.05} = 3.8415$

$N = 240, a = 71, b = 42, c = 49, d = 78$

$$\chi^2 = \frac{N(ad - bc)^2}{(a+b)(a+c)(b+d)(c+d)} = \frac{240(71 \times 78 - 49 \times 42)^2}{113 \cdot 120 \cdot 120 \cdot 127}$$

$$\chi^2 = 240 \times \frac{12110400}{206654400} = 14.0645$$

$$\Rightarrow \chi^2 = 14.0645 > \chi^2_{1,0.05} = 3.8415$$

\Rightarrow Reject H_0 at 5% level of significance, we may say that the Psychotherapy is effective at 5% LOS.

* Yate's correction for continuity: when the cell frequencies a, b, c, d as observed in case of the four classes corresponding to two attributes are small, we cannot use the test Statistic χ^2 as defined in previous section i.e. for when the assumption a, b, c, d are greater than or equal to 5 does not hold, the distribution of $\chi^2 = \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$ cannot be considered to be Chi-square with one degree of freedom.

In such a case i.e. if the cell frequencies a, b, c, d are not all greater than or equal to 5, we make the following adjustment called Yate's correction.

- if $ad < bc$ add $\frac{1}{2}$ to a and b and subtract $\frac{1}{2}$ from both b and c both.
- if $ad > bc$ add $\frac{1}{2}$ to b and c and subtract $\frac{1}{2}$ from a and b both.

The decision criteria at level of significance $= \alpha$ is given by

Reject H_0 : if $\chi^2 > \chi^2_{1,\alpha}$, do not reject H_0 : if $\chi^2 < \chi^2_{1,\alpha}$

$$\text{where } \chi^2 = \frac{N(ad-bc - \frac{N}{2})^2}{(a+b)(a+c)(b+d)(c+d)}$$

(ii) In a experiment on immunization of cattle from tuberculosis the foll results were obtained:

	affected	unaffected
Inoculated	11	31
Not inoculated	14	4

Examine the effect of vaccine in controlling the incidence of the disease at 1% LOS.

$\Rightarrow H_0$: the attributes are independent.

H_1 : the attributes are not independent.

LOS = 1%, $\alpha = 0.01$

Decision criteria :- Reject H_0 iff $\chi^2 > \chi^2_{1,0.01} = 6.6349$

$$N = 60, a = 11, b = 31, c = 14, d = 4$$

$$\begin{aligned} \chi^2 &= \frac{N(ad - bc)^2}{(a+b)(a+c)(b+d)(c+d)} = \frac{60(11 \times 4 - 31 \times 14)^2}{42 \cdot 18 \cdot 25 \cdot 35} \\ &= \frac{60(390 - 30)^2}{661500} = \frac{60 \times 360 \times 360}{661500} = 11.7551 \end{aligned}$$

$$\therefore \chi^2 = 11.7551 > \chi^2_{1,0.01} = 6.6349$$

\therefore Reject H_0 at 1% LOS.

We can say at 1% LOS that the Inoculation and affection due to disease are dependent.

* Test In $r \times c$ contingency table.

Test statistic is $\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, where $E_{ij} = \frac{a_i b_j}{N}$

are the expected frequencies, we can perform a test for testing the hypothesis H_0 : the attributes A and B under the consideration are independent against the logical alternative H_1 : not H_0 i.e. the attributes A and B are dependent subject to the conditions $N > 50$ and each of the observed frequencies ≥ 5 ($O_{ij} \geq 5$).

The decision criterion at level of significance $= \alpha$ is given by

Reject H_0 : if $\chi^2 > \chi^2_{(r-1)(c-1), \alpha}$,

do not reject H_0 : if $\chi^2 \leq \chi^2_{(r-1)(c-1), \alpha}$,

Q.1 Using the data given in the following table decide whether we can conclude that standard of clothing of a salesman has significant effect on his performance in field selling at 5% LOS.

		Performance in Field Selling				
		Disappointing	Satisfactory	Excellent	Total	
		21	15	6	42	
Poorly dressed		21	15	6	42	
Well dressed		24	35	26	85	
Very well dressed		35	80	58	173	
Total		80	130	90	300	

$\Rightarrow H_0$: attributes are independent

H_1 : attributes are not independent

LOS : 5%, $\alpha = 0.05$

Decision criteria:- Reject H_0 iff $\chi^2 > \chi^2_{4, 0.05} = 9.4877$,

where $\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$.

$$E_{11} = \frac{a_1 b_1}{N} = \frac{42.80}{300} = 11.2$$

$$E_{12} = \frac{a_1 b_2}{N} = \frac{42.130}{300} = 18.2$$

$$E_{13} = \frac{a_1 b_3}{N} = \frac{42.90}{300} = 12.6$$

$$E_{21} = \frac{a_2 b_1}{N} = \frac{85.80}{300} = 22.87 \quad / E_{22} = 36.83 \quad / E_{23} = 25.5$$

$$E_{31} = \frac{a_3 b_1}{N} = \frac{173.80}{300} = 46.13$$

$$E_{32} = \frac{a_3 b_2}{N} = \frac{173.130}{300} = 54.97$$

$$E_{33} = \frac{a_3 b_3}{N} = \frac{173.90}{300} = 51.9$$

O_{ij}	E_{ij}	$O_{ij} - E_{ij}$	$\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
21	11.2	9.8	8.575
15	18.2	+3.2	0.5626
8	12.6	-6.6	3.4571
24	22.87	1.33	0.078
35	36.83	-1.83	0.0909
26	25.5	0.5	0.0097
35	46.13	-11.13	2.6854
80	54.97	5.03	0.3375
58	51.9	6.1	0.7170
—	—	Total	16.5133

$$\chi^2 = 16.5133 > \chi^2_{4, 0.05} = 9.488$$

∴ Reject H_0 at 5% LOS.

∴ We decide to reject H_0 at 5% LOS and conclude that the standard of clothing of a salesman has significant effect on his performance in field selling.

- * Co-efficient of contingency: A measure of the degree of relationship, association, or dependence of the classification in a contingency table is given by :-

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

which is called co-efficient of contingency

Q.1) Previous question same as it is :-

$$\rightarrow C = \sqrt{\frac{\chi^2}{\chi^2 + N}} = \sqrt{\frac{16.5133}{16.5133 + 300}} = 0.2284$$

- * Correlation of Attributes:- The classification in a contingency table often describe characteristics of individuals or objects, they are often referred to as attributes, and the degree of dependence, association, or relationship is called the correlation of attributes.

For $k \times k$ tables, we define :- $r = \frac{\chi^2}{N(k-1)}$ as the correlation

co-efficient between attributes (or classification) - This coefficient lies between 0 and 1. For 2×2 tables in which $k=2$, the correlation is often called tetrachoric correlation.