2|3|20

UNIT-5

Page No.:
Date:

* Test-1 Chi-Square test:- (M-Imp)

☆ Case (i) :- [Observed (O) - Expected (E) Frequency]

Step-1 Ho: There is no difference between observed & expected frequencies.

H1: There is some difference between observed & expected frequencies.

Step-2 Expected Frequency = Average of Observed Frequency.

Step-3

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

$$\sum \frac{(O-E)^2}{E} = \chi^2 \text{ value}$$

Step-4 $d.f = n-1$

Step-5 χ^2 table, $\chi^2_{d.f, 0.05}$

Page No.:
Date:

Eg:1 A die is thrown for 300 times & the following distribution is obtained. Can the die be regarded unbiased.

Number on the die	1	2	3	4	5	6
Frequency	41	44	49	53	57	56

Soln: **Step-1** H_0 : Die is unbiased
H_1 : Die is biased

Step-2

$$\text{Expected Frequency} = \frac{41+44+49+53+57+56}{6}$$

$$= 50$$

Step-3

O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
41	50	-9	81	1.62
44	50	-6	36	0.72
49	50	-1	1	0.02
53	50	3	9	0.18
57	50	7	49	0.98
56	50	6	36	0.72
				= 4.24

Step-4 $d.f = n-1 = 6-1 = 5$

Step-5 χ^2-table,

$$\chi^2_{5,\,0.05} = 11.07$$

Here, $x^2 = 4.24 < 11.07$

$\Rightarrow$ Accept Ho

$\therefore$ Die is unbiased.

Eg:2 The number of road accidents on a highway during a week is given below. Can it be concluded that the proportion of accidents are equal for all days.

Days	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.	Sun
No. of accidents	14	16	8	12	11	9	14

Solⁿ:- Step-1

Ho: There is no difference between observed & expected frequency of accidents.

H1: There is some difference between observed - expected frequency of accidents.

Step-2 Expected Frequency $= \frac{14+16+8+12+11+9+14}{7}$

$= \frac{84}{7} = 12$

Scanned with CamScanner

Step-3

O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
14	12	2	4	0.33
16	12	4	16	1.33
8	12	-4	16	1.33
12	12	0	0	0
11	12	-1	1	0.08
9	12	-3	9	0.75
14	12	2	4	0.33
				= 4.15

Step-4 $d.f = n-1 = 7-1 = 6$

Step-5 χ^2-table, $\chi^2_{6,0.05} = 12.592$

Here,

$\chi^2 = 4.15 < 12.592$

$\Rightarrow H_0$ is accepted.

Scanned with CamScanner

Page No.:
Date:

☆ Case (ii) :- [Dependency - Independency]

Table Format :-

	1	2	3
A	–	–	–
B	–	–	–
C	–	–	–

Eg:1 In a industry, 200 workers employed for a specific job were classified according to their performance & training received/not received. Test independence of training & performance. The data are summarised as follows :-

	Performance		
	Good	Not Good	Total
Trained	100	50	150
Untrained	20	30	50
Total:-	120	80	200

Soln:

	Good	Not Good	Total
Trained	100 E_{11}	50 E_{12}	150 - 1st row
Untrained	20 E_{21}	30 E_{22}	50 - 2nd row
	120	80	
	1st column	2nd column	

Page No.:
Date:

Step-1 H_0: Performance is Independent of training

H_1: Performance is dependent of training.

Step-2

$$E_{11} = \frac{150 \times 120}{200} = 90$$

$$E_{12} = \frac{150 \times 80}{200} = 60$$

$$E_{21} = \frac{50 \times 120}{200} = 30$$

$$E_{22} = \frac{50 \times 80}{200} = 20$$

Step-3

O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
100	90	10	100	1.11
50	60	-10	100	1.66
20	30	-10	100	3.33
30	20	10	100	5
				11.1

Step-4 $d.f = (r-1)(c-1)$ r : rows

$= (2-1)(2-1)$ c : columns

$= (1)(1)$

$= 1$

Step-5 $\chi^2_{1, 0.05}$
$= 3.84$

$\chi^2 = 11.1 > 3.84$
$\Rightarrow$ Reject Ho

$\therefore$ Performance is dependent on training.

Eg:2 The result in the last examination of a sample of 100 students is given below:-

	1st class	2nd class	3rd class	Total
Boys	10	28	12	50
Girls	20	22	8	50
Total:-	30	50	20	100

Can it be said that the performance in the examination depends upon sex.

Soln: =

	1st class	2nd class	3rd class	Total
Boys	10 (E_{11})	28 (E_{12})	12 (E_{13})	50
Girls	20 (E_{21})	22 (E_{22})	8 (E_{23})	50
Total:-	30	50	20	100

Step-1: H0: Performance independent upon sex.
H1: Performance dependent upon sex.

Step-2: $E_{11} = \frac{50 \times 30}{100} = 15$

$E_{12} = \frac{50 \times 50}{100} = 25$

$E_{13} = \frac{50 \times 20}{100} = 10$

$E_{21} = \frac{50 \times 30}{100} = 15$

$E_{22} = \frac{50 \times 50}{100} = 25$

$E_{23} = \frac{50 \times 20}{100} = 10$

Step-3

O	E	O·E	$(O \cdot E)^2$	$\frac{(O \cdot E)^2}{E}$
10	15	-5	25	1.667
28	25	3	9	0.36
12	10	2	4	0.4
20	15	5	25	1.667
22	25	-3	9	0.36
8	10	-2	4	0.4
				4.854

Step-4: d.f = (r-1)(c-1)
= (2-1)(3-1)
= (1)(2)
= 2

r = 2 (No. of rows)

c = 3 (No. of columns)

Step-5: χ^2 table value at d.f 2,
$= 5.991$

Here, $\chi^2 = 4.854 < 5.991$

$\Rightarrow$ Ho may be accepted.
$\Rightarrow$ Performance is independent upon sex.

* Definition:- Goodness of Fit Test:-

— Suppose we have obtained an observed frequency distribution & we are interested in knowing whether the observed frequency distribution support a particular hypothesis. For this a very powerful test for testing the significance of the discrepancy between observed frequency distribution & expected frequency distribution was given by Karl Pearson in 1900. The test is known as χ^2 test of goodness of fit.
Under the null hypothesis that there is no significant difference between observed & expected frequencies, the value of χ^2 is calculated by the formula:-

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$

If all the observed frequencies & expected frequencies are equal, the value of χ^2 will

be zero. This will signify a perfect agreement of observations with expectations.
More the value of X^2, more is the divergence between the observed & expected frequencies.
The value of X^2 is calculated from the given data & it is compared with the table value of X^2 on n-1 degrees of freedom & at a required level of significance.

NON-PARAMETRIC TESTS :-

- These tests do not require any specific form for the distribution of the population is called non-parametric tests.

- Non-parametric methods provide an alternative series of statistical methods that require no or very limited assumptions to be made about the data.

- There is a wide range of methods that can be used in different circumstances.

Date:

* Advantages Of Non-Parametric Methods/Test:-

- Non-parametric test requires no or very limited assumptions to be made about the format of the data, & they may therefore be prefarable when the assumptions required for parametric Test are not valid.

- Non-parametric test can be useful for dealing with unexpected outlying observations that might be problematic with a parametric approach.

- Non-parametric test are Intuitive & are simple to carryout by hand for small samples at least.

- Non-parametric tests are often useful in the analysis of ordered categorical data in which assignation of scores to a individual categories may be inappropriate.

* Disadvantages Of Non-Parametric Tests:-

- Non-parametric tests/methods may lack power as compared with more traditional approaches. This is a particular concern if the sample size is small or if the assumptions for the corresponding parametric method hold.

Page No.:
Date:

- Tied values can be problematic when these are common & adjustments to the test stastic may be necessary.

- Appropriate computer software for non-parametric methods can be limited although the situation is improving.

- In addition, how a software package deals with tied values or how it obtains appropriate P values may not always be obvious.

- Non-parametric methods are geared toward hypothesis testing rather than estimation of effects. It is often possible to obtain non-parametric estimates & associated confidence intervals, but this is not generally useful everytime.

Page No.:
Date:

* Test 1 : Sign Test for One Sample :-

This test is used for testing on Median value (M).

Step-1 : $H_0: M =$ Given Number
$H_1: M \neq$ given Number

For e.g., ($H_0: M = 22$, $H_1: M \neq 22$)

Step-2 : Let level of significance
$\alpha = 0.01$ or $\alpha = 0.05$
i.e. $\alpha = 1\%$ or $\alpha = 5\%$

either (Sometimes given, otherwise use 0.05)

Step-3 : Assign the signs to given numbers

i.e., +ve sign → if number is greater than given number.

-ve sign → if number is lesser than given number.

Zero → if number is same.

Step-4 Count Number of Signs & +ve signs :-

i.e., n = Total Number of Signs (excluding 0).
x = Total Number of +ve signs.

Step-5 Find the Probability by using binomial distribution for $P(x) \geq x$ upto n.

where,

$P(x) = {}^nC_x \, p^x q^{n-x}$, p = success probability
q = failure probability
i.e., $q = 1-P$
Usually,
$P = 0.5$ & $q = 1-0.5$
$= 0.5$

&

$$ {}^nC_x = \frac{n!}{x!(n-x)!} $$

if $P(x) < 0.01$ or 0.05
$\Rightarrow$ Ho may be rejected.

&

if $P(x) > 0.05$ or 0.01
$\Rightarrow$ Ho may be Accepted.

Note:- In sign test only, if P(x) is less than 0.01 or 0.05 so we will reject usually for less than we are accepting but here we have to remember this.

Page No.:
Date:

Eg:1 The following data constitute a random sample of 15 measurement of the octane rating of a certain kind of gasoline:

99, 102·3, 99·8, 100·5, 99·7, 96·2, 99·1, 102·5, 103·3, 97·4, 100·4, 98·9, 98·3, 98·0, 101·6.

Test the hypothesis that population median is 98 or not?

Soln: Here, we have to test on Median. So we will use sign test.

Step-1 $H_0: \mu = 98$
$H_1: \mu \neq 98$

Step-2 Here, level of significance not given so we will take $\alpha = 0.05$

Step-3 Now, we will assign the signs to above numbers as:-
+ve sign → if number is greater than 98.
−ve sign → if number is less than 98
0 for same number.

$$\boxed{\begin{array}{l} +ve,\ +ve,\ +ve,\ +ve,\ +ve,\ -ve,\ +ve \\ +ve,\ +ve,\ -ve,\ +ve,\ +ve,\ +ve,\ 0 \\ +ve \end{array}}$$

Page No.:
Date:

Step-4

n = Total number of signs (excluding 0).

$\boxed{n = 14}$

x = Total number of positive (+ve) signs.

$\boxed{x = 12}$

Step-5 Using Binomial distribution, we have to find,

$P(x) \geq x$ upto n

i.e., $P(x) \geq 12$ upto 14

i.e., $P(x)$ for $x = 12, 13, 14$

$\therefore P(x) = {}^{n}C_{x} p^{x} q^{n-x}$

Now,
Probability value $= P(12) + P(13) + P(14)$

$= {}^{14}C_{12} p^{12} q^{14-12} + {}^{14}C_{13} p^{13} q^{14-13} + {}^{14}C_{14} p^{14} q^{14-14}$

$= {}^{14}C_{12} (0.5)^{12} (0.5)^{2} + {}^{14}C_{13} (0.5)^{13} (0.5)^{1} + {}^{14}C_{14} (0.5)^{14} (0.5)^{0}$

$= {}^{14}C_{12} (0.5)^{12+2} + {}^{14}C_{13} (0.5)^{13+1} + {}^{14}C_{14} (0.5)^{14+0}$

Date:

$$= {}^{14}C_{12}(0.5)^{14} + {}^{14}C_{13}(0.5)^{14} + {}^{14}C_{14}(0.5)^{14}$$

$$= (0.5)^{14}\left[{}^{14}C_{12} + {}^{14}C_{13} + {}^{14}C_{14}\right]$$

$$= (0.5)^{14}\left[\frac{14!}{12!(14-12)!} + \frac{14!}{13!(14-13)!} + 1\right]$$

$(\because {}^{n}C_{n} = 1)$

$$= (0.5)^{14}\left[\frac{14 \times 13 \times \cancel{12!}}{\cancel{(12!)}(2!)} + \frac{14 \times \cancel{13!}}{\cancel{13!}(1)!} + 1\right]$$

$$= (0.5)^{14}\left[\frac{\overset{7}{\cancel{14}} \times 13}{\cancel{2} \times 1} + \frac{14}{1} + 1\right]$$

$$= (0.5)^{14}\left[91 + 14 + 1\right]$$

$$= (0.5)^{14}(106)$$

$$= 0.0000610352 \times 106$$

$$P(x) = 0.00647$$

$\Rightarrow$ Ho may be rejected.

$\therefore$ Population Median can't be 98.

Eg: 2 The following are the average weekly losses of worker-hours due to accidents in 10 industrial plants before & after a certain safety program was put into operation.

Before	45	73	46	124	33	57	83	34	26	17
After	36	60	44	119	35	51	77	29	24	11

Use the sign test at the 0.05 level of significance to test whether the safety program is effective.

Soln: Step-1 H_0: There is no effect on safety program.
H_1: There is some effect on safety program.

Step-2 Here, $\alpha = 0.05$ which is given.

Step-3 Now, we will take the difference between before & after & accordingly we will get signs.

Page No.:
Date:

Before (x)	After (y)	$d = x - y$
45	36	+9
73	60	+13
46	44	+2
124	119	+5
33	35	−2
57	51	+6
83	77	+6
34	29	+5
26	24	+2
17	11	+6

Step-4

n = Total Number of Signs

$\boxed{n = 10}$

x = Total number of positive signs.

$\boxed{x = 9}$

Step-5 Using Binomial Distribution we have to find,

$P(x) \geq x$ upto n

i.e., $P(x) \geq 9$ upto 10

i.e., $P(x)$ for $x = 9, 10$

Now,

Probability value $= P(9) + P(10)$

$= {}^{10}C_9 \, p^9 q^{10-9} + {}^{10}C_{10} \, p^{10} q^{10-10}$

$$= {}^{10}C_9 (0.5)^9 (0.5)^1 + {}^{10}C_{10} (0.5)^{10} (0.5)^0$$

$$= \left[{}^{10}C_9 + {}^{10}C_{10}\right] (0.5)^{10}$$

$$= \left[\frac{10!}{9!\,(10-9)!} + 1\right] (0.5)^{10}$$

$$= \left[\frac{10 \times \cancel{9}}{\cancel{9}! \times 1!} + 1\right] (0.5)^{10}$$

$$= [11] (0.5)^{10}$$

$$= [11] [0.0009765625]$$

Probability value $= 0.01074$

Here, Probability value $= 0.01074 < 0.05$

$\Rightarrow$ Ho may be rejected.

$\Rightarrow$ Safety program is effective.

Page No.:
Date:

* Test-2 Mann Whitney U-Test :-

This test is used when,

1.) Two groups are to be compared
2.) The scale is ordinal (for data types of ranks & scores).
3.) Data are not normally distributed.
4.) The distribution is not known or doubtful.

The U-test is also called rank-sum test because the test depends on the ranks of the sample observations.

Step-1 H_0: There is no difference between the groups.
H_1: There is some difference between the groups.

Step-2
$\alpha = 0.01$ or 0.05
i.e., 1% or 5%
z-value ⇒ 2.58 or 1.96

Step-3 Assign the ~~&~~ ranks on given data A & B say R_1 & R_2.

Step-4 $U_1 = \underset{(R_1)}{\text{Rank Sum}} - \dfrac{n_1(n_1+1)}{2}$

Page No.:
Date:

U_2 = Rank Sum $(R_2) - \frac{n_2(n_2+1)}{2}$

Step-5

U = smaller value of U_1 & U_2.

Step-6

Find $\mu = \frac{n_1 n_2}{2}$, $\sigma = \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{2}}$

& $Z = \frac{U-\mu}{\sigma}$

if $|Z| > 1.96 \Rightarrow$ Reject Ho.
$|Z| < 1.96 \Rightarrow$ Accept Ho//.

Eg:1 Twenty-three applications for a position are interviewed by three administrators & rated on a scale of 5 as to suitability for a "stability" score which is the sum of the three numbers. Use the Mann-Whitney U-test to determine whether there was a difference in the scores of the two groups.

Group A	7	11	9	4	8	6	12	11	9	10	11	11
Group B	8	9	13	14	11	10	12	14	13	9	10	8

Soln:

Step-1 H_0: There is no difference in both the groups A & B.

H_1: There is some difference in both the groups A & B.

Step-2 Let $\alpha = 0.05$ i.e, 5%.

$Z = 1.96$

Step-3 Assigning the ranks on given groups A & B.

First arrange the numbers smaller to larger, if there are repeated numbers then take the average of ranks & give the equal number to repeated numbers.

Scores | | Ranks

A	B	Ranks	Common Ranks	Ranks A	Ranks B
		1	1	1	
4		2	2	2	
6		3	3	3	
7	8, 8	4, 5, 6	5	5	10
8	9, 9	7, 8, 9, 10	8.5	17	17
9, 9	10, 10	11, 12, 13	1.2	12	24
10	11	14, 15, 16, 17, 18	16	64	16
11, 11, 11, 11	12	19, 20	18.5	18.5	18.5
12	13, 13	21, 22	20.5		41
	14, 14	23, 24	23.5		47
				122.5	173.5

Step-4 $U_1 = \underset{(R1)}{\text{Rank Sum}} - \frac{n_1(n_1+1)}{2}$

$= 122.5 - \frac{\cancel{12}^{6}(12+1)}{\cancel{2}}$, n_1 = Total numbers in group A = 12

$= 122.5 - 78$

$U_1 = 44.5$

&

$U_2 = \underset{(R_2)}{\text{Rank Sum}} - \frac{n_2(n_2+1)}{2}$

$= 173.5 - \frac{12(12+1)}{2}$

$= 173.5 - 78$

$= 95.5$

Date:

Step-5 $U = \text{Smaller}(U_1, U_2)$
$= \text{Smaller}(44.5, 95.5)$
$= 44.5$

Step-6 $\mu = \frac{n_1 n_2}{2} = \frac{(12)(12)}{2} = 72$

$$\sigma = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{2}} = \sqrt{\frac{(12)(12)(12+12+1)}{2}}$$
$$= \sqrt{1800}$$
$$= 42.43$$

$$Z = \frac{U - \mu}{\sigma} = \frac{44.5 - 72}{42.43}$$

$$Z = -0.6481$$

$$|Z| = 0.6481 < 1.96$$

⇒ Accept Ho.
⇒ There is no difference among the groups.

Page No.:
Date:

Eg: 2 Use Mann-Whitney U-test to test the no significance difference between the average productivities per hour by male & female in a certain company.

Male	31	44	25	30	70	63	54	42	36	22	25	50
Female	38	34	33	47	58	83	18	36	41	37	24	48

Sol^n:- Step-1 Ho: There is no difference between Males & Females.

H_1: There is some difference between Males & Females.

Step-2 Let $\alpha = 0.05$ or 5%.

$z = 1.96$

Step-3 Assigning the ranks. ($n_1 = 12$, $n_2 = 12$)

Scores			Common	Ranks	
Males	Females	Ranks	Ranks	R_1	R_2
	18	1	1		1
22		2	2	2	~~2~~
	24	3	3		3
25, 25		4, 5	4.5	9	
30		6	6	6	
31		7	7	7	
	33	8	8		8
	34	9	9		9
36	36	10, 11	10.5	10.5	10.5
	37	12	12		12
	38	13	13		13
	41	14	14		14
42		15	15	15	
44		16	16	16	
	47	17	17		17
	48	18	18		18
50		19	19	19	
54		20	20	20	
	58	21	21		21
63		22	22	22	
70		23	23	23	
	83	24	24		24
				149.5	150.5

Step-4

$$U_1 = \text{Rank Sum } (R_1) - \frac{n_1(n_1+1)}{2}$$

$$= 149.5 - \frac{12(12+1)}{2}$$

$$= 149.5 - 78 = 71.5$$

Step-5 U_2 = Smaller (U_1, U_2)

$= \text{Smaller}(71.5, 72.5)$

$= 71.5$

Step-6

$$\mu = \frac{n_1 n_2}{2} = \frac{(12)(12)}{2} = 72$$

$$\sigma = \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{2}} = \sqrt{\frac{12(12)(12+12+1)}{2}} = 42.43$$

$$Z = \frac{U-\mu}{\sigma} = \frac{71.5-72}{42.43} = -0.01178$$

$|Z| = 0.01178 < 1.96$

$\Rightarrow$ Accept H_0.

$\Rightarrow$ No difference among males & females.

(Unit-5 Ends)