Unit-IV Non-Parametric Test

A population in statistics means a set of object. The population is finite or infinite according to the number of elements of the set is finites or infinite.

Sampling:

A sample is a finite subset of the population. The number of elements in the sample is called size of the sample.

Large and small sample:

The number of elements in a sample is greater than or equal to 30 then the sample is called a large sample and if it is less than 30, then the sample is called a small sample.

Parameters:

Statistical constant like mean μ, variance σ^2 , etc., computed from a population are called parameters of the population.

Statistics:

Statistical constants like $\bar{x}$, variance S^2, etc., computed from a sample are called samlple staticts or statistics.

POPULATION (PARAMETER)	**SAMPLE (STATISTICS)**
Population size=N	Sample size=n
Population mean= μ	Sample mean= x
Population s.d.=σ	Sample s.d.=S
Population proportion= P	Sample proportion= p

<u>Tests of significance or Hypothesis Testing:</u>

Statistical Hypothesis:

In making statistical decision, we make assumption, which may be true or false are called Statistical Hypothesis.

Null Hypothesis(H_0):

For applying the test of significance, we first setup a hypothesis which is a statement about the population parameter. This statement is usually a hypothesis of no true difference between sample statistics and population parameter under consideration and so it is called null hypothesis and is denoted by H_0.

Alternative Hypothesis (H_1):

Suppose the null hypothesis is false, then something else must be true. This is called an alternative hypothesis and is denoted by H_1.

Eg. If H_0 is population mean μ=300, then H_1 is $\mu \neq 300$ $(ie. \mu < 300 \ or \ \mu > 300)$ or $H_1 \ is \ \mu > 300 \ or \ H_1 \ is \ \mu < 300$. So any of these may be taken as alternative hypothesis.

$$t = \frac{15-14}{3.146\sqrt{\frac{1}{10}+\frac{1}{12}}} = 0.742$$

Critical value: The critical value of t at 5% level of significance with degrees of freedom $n_1 + n_2 - 2 = 10 + 12 - 2 = 20$ is 2.086

Conclusion: calculated value < table value

H_0 is Accepted.

ii) F-test to test equality of populations variances:

Null Hypothesis H_0: $\sigma_1^2 = \sigma_2^2$ The population Variances are equal

Alternative Hypothesis H_1: $\sigma_1^2 \neq \sigma_2^2$ The population Variances are not equal

Level of significance: $\alpha = 5\%$

Test Statistics:

$$F = \frac{S_1^2}{S_2^2}$$

Where $S_1^2 = \frac{1}{n_1 - 1}\sum(x-\bar{x})^2 = \frac{1}{10-1}(90) = 10$

$$S_1^2 = \frac{1}{n_1 - 1}\sum(y-\bar{y})^2 = \frac{1}{12-1}(108) = 9.818$$

Here $S_1^2 > S_2^2$ $\therefore F = \frac{S_1^2}{S_2^2} = \frac{10}{9.818} = 1.02$

Critical value: The critical value of F at 5% level of significance with degrees of freedom $(n_1 - 1, n_2 - 1) = (9,11)$ is 2.90

Here calculated value < table value, we accept H_0

Conclusion: Both null hypothesis $\mu_1 \neq \mu_2$ and $\sigma_1^2 = \sigma_2^2$ are accepted.

Hence we may conclude the two samples are drawn from same normal population.

III χ^2 **-test:**

(i). χ^2 -Test for a specified population variance

(ii). χ^2 -test is used to test whether differences between observed and expected frequencies are significant (goodness of fit).

(iii). χ^2 -test is used to test the independence of attributes.

χ^2 -Test for a specified population variance:

The test statistics $\chi^2 = \frac{ns^2}{\sigma^2}$

Which follows χ^2 - distribution with (n – 1) degrees of freedom

<u>Problem:</u>

1. **The lapping process is used to grind certain silicon wafers to the proper thickness is acceptable only σ , the population S.D. of the thickness of dice cut from the wafers, is at most 0.5mil. Use the 0.05 level of significance to test the null hypothesis σ =0.5 against the alternative hypothesis σ >0.5, if the thickness of 15 dice cut from such wafers have S.D of 0.64mil.**

Solution:

Given $n = 15$, s=0.64, σ =0.5

$H_0 : \sigma = 0.5$, $H_1 : \sigma > 0.5$

Under H_0 , The test statistics $\chi^2 = \frac{ns^2}{\sigma^2} = \frac{15\ (0.64)^2}{(0.5)^2} =$ 24.576

From χ^2 table, with degrees of freedom = 14, $\chi^2_{0.05} = 23.625$

$\therefore \chi^2 > \chi^2_{0.05}$ H_0 is rejected. Hence $\sigma > 0.5$

χ^2-test is used to test whether differences between observed and expected frequencies are significant (goodness of fit):

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

Where O_i is observed frequency, and E_i is the expected frequency.

If the data given in a series of n number, then degree of freedom = n - 1 .

Note: In case of binomial distribution d.f = n – 1, poisson distribution d.f = n – 2, normal distribution d.f = n – 3.

Problem:

1. **The following data give the number of aircraft accident that occurred during the various days of a week:**

Days :	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**
No of accidents:	**15**	**19**	**13**	**12**	**16**	**15**

Test the whether the accident are uniformly distributed over the week.

Solution:

The expected number of accident on any day = $\frac{90}{6} = 15$

Let H_0 : Accidents occur uniformly over the week

H_1 : Accidents not occur uniformly over the week

Days	**Observed Freqency (O_i)**	**Expected Frequency (E_i)**	$(O_i - E_i)$	$\frac{(O_i - E_i)^2}{E_i}$
Mon	15	15	0	0
Tue	19	15	4	1.066
Wed	13	15	-2	0.266
Thu	12	15	-3	0.6
Fri	16	15	1	0.066
Sat	15	15	0	0
		90		1.998

Now, $\chi^2 = \sum_i \left[\frac{(O_i - E_i)^2}{O_i} \right] = 1.998$

Here 6 observations are given, degrees of freedom = n – 1= 6 – 1 = 5

From χ^2 table, with degrees of freedom = 5, $\chi^2_{0.05} = 11.07$

$\therefore \chi^2 < \chi^2_{0.05}$ H_0 is accepted.

Conclusion: $\therefore$ Accidents occur uniformly over the week

2. A survey of 320 families with 5 children each revealed the following distribution:

No. of Boys:	**5**	**4**	**3**	**2**	**1**	**0**
No. of Girls:	**0**	**1**	**2**	**3**	**4**	**5**
No. of families:	**14**	**56**	**110**	**88**	**40**	**12**

Is the result consistent with the hypothesis that male and female births are equally probable?

Solution:

Let H_0: Male and female births are equally probable

H_1: Male and female births are not equally probable

Probability of male birth $= p = \frac{1}{2}$, Probability of female birth $= q = \frac{1}{2}$

The probability of x male births in a family of 5 is $p(x) = 5C_x p^x q^{5-x}, x = 0,1,2...5$

Expected number of families with x male births $= 320 \times 5C_x p^x q^{5-x}, x = 0,1,2...5$

$$= 320 \times 5C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x}$$

$$= 320 \times 5C_x \left(\frac{1}{2}\right)^5 = 10 \times 5C_x$$

The χ^2 is calculated using the following table:

No. of Boys	**Observed Freqency** (O_i)	**Expected Frequency** $E_i = 10 \times 5C_x$	$(O_i - E_i)$	$\frac{(O_i - E_i)^2}{E_i}$
5	14	10	4	1.6
4	56	50	6	0.72
3	110	100	10	1
2	88	100	-12	1.44
1	40	50	-10	2
0	12	10	2	0.4
Total	320	320		7.16

$$\therefore \chi^2 = 7.16$$

The tabulated value of χ^2 for n – 1 = 6 – 1 =5 degrees of freedom at 5% level of significance $= \chi^2_{0.05} = 11.07$

Since $\chi^2 < \chi^2_{0.05}$. So we accepted H_0.

Conclusion: $\therefore$ The male and female births are equally probable.

3. Fit a poisson distribution to the following data and test the goodness of fit.

x:	**0**	**1**	**2**	**3**	**4**	**5**	**6**
f(x):	**275**	**72**	**30**	**7**	**5**	**2**	**1**

Solution:

Mean of the given distribution $=\bar{x}=\dfrac{\sum f_i x_i}{\sum f_i}=\dfrac{189}{392}=0.482$

To fit a poisson distribution to the given data:

We take the parameter of the poisson distribution equal to the mean of the given distribution. $=\lambda = \bar{x} = 0.482$

The poisson distribution is given by $P(X=x)=\dfrac{e^{-\lambda}\lambda^x}{x!};\ x=0,1,2...\infty$

and the expected frequencies are obtained by $f(x)=\left(\sum f_i\right)\times\dfrac{e^{-\lambda}\lambda^x}{x!}=392\times\dfrac{e^{-0.482}(0.482)^x}{x!}$

we get $f(0)=392\times\dfrac{e^{-0.482}(0.482)^0}{0!}=242.1,\ f(1)=392\times\dfrac{e^{-0.482}(0.482)^1}{1!}=116.69$

$f(3)=4.518,\ f(4)=0.544,\ f(5)=0.052\approx 0.1,\ f(6)=0.004\approx 0$

x:	0	1	2	3	4	5	6	Total
Expected Frequency:	242.1	116.69	28.12	4.518	0.544	0.052	0.004	392

H_0: The poisson distribution fit well into the data.

H_1: The poisson distribution does not fit well into the data.

The χ^2 is calculated using the following table:

x	**Observed Freqency** (O_i)		**Expected Frequency** (E_i)		$\dfrac{(O_i-E_i)^2}{E_i}$
0	275		242.1		4.471
1	72		116.7		17.122
2	30		28.1		0.128
3	**7**		**4.5**		
4	**5**	**15**	**0.5**	**5.1**	**19.218**
5	**2**		**0.1**		
6	**1**		**0**		
Total	392		392		40.939

$\therefore \chi^2 = 40.939$

The tabulated value of χ^2 for = 7 – 1 –1– 3 =2 degrees of freedom at 5% level of significance $=\chi^2_{0.05}=5.991$

Since $\chi^2 > \chi^2_{0.05}$. So we rejected H_0.

Conclusion: $\therefore$

The Poisson distribution is not a good fit to the given data.

χ^2-test is used to test the independence of attributes:

An attributes means a equality or characteristic. χ^2- test is used to test whether the two attributes are associated or independent. Let us consider two attributes A and B. A is divided into three classes and B is divided into three classes.

		Attribute B			
		B_1	B_2	B_3	Total
Attribute A	A_1	a_{11}	a_{12}	a_{13}	R_1
	A_2	a_{21}	a_{22}	a_{23}	R_2
	A_3	a_{31}	a_{32}	a_{33}	R_3
	Total	C_1	C_2	C_3	**N**

Now, under the null hypothesis H_0: The attributes A and B are independent and we calculate the expected frequency E_{ij} for varies cells using the following formula.

$$E_{ij} = \frac{R_i \times C_j}{N}, \; i = 1,2,...r, \; j = 1,2,...s$$

$E(a_{11}) = \frac{R_1 \times C_1}{N}$	$E(a_{12}) = \frac{R_1 \times C_2}{N}$	$E(a_{13}) = \frac{R_1 \times C_3}{N}$	R_1
$E(a_{21}) = \frac{R_2 \times C_1}{N}$	$E(a_{22}) = \frac{R_2 \times C_2}{N}$	$E(a_{23}) = \frac{R_2 \times C_3}{N}$	R_2
$E(a_{31}) = \frac{R_3 \times C_1}{N}$	$E(a_{32}) = \frac{R_3 \times C_2}{N}$	$E(a_{33}) = \frac{R_3 \times C_3}{N}$	R_3
C_1	C_2	C_3	**N**

and we compute $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

Which follows χ^2 distribution with n = (r-1) (s-1) degrees of freedom at 5% or 1% level of significance.

1. **Calculate the expected frequencies for the following data presuming two attributes viz., conditions of home and condition of child as independent.**

Condition of Child	**Condition of home**		
		Clean	**Dirty**
	Clean	**70**	**50**
	Fair	**80**	**20**
	Dirty	**35**	**45**

Use Chi-Square test at 5% level of significance to state whether the two attributes are independent.

Solution:

Null hypothesis H_0 : Conditions of home and conditions of child are independent.

Alternate hypothesis H_1 : Conditions of home and conditions of child are not independent.

Level of significance: $\alpha = 0.05$

The test statistics: $\chi^2 = \sum_{i=1}^{r} \sum_{i=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

Analysis:

	Condition of home			Total
		Clean	Dirty	
Condition of Child	Clean	70	50	120
	Fair	80	20	100
	Dirty	35	45	80
Total		185	115	300

Expected Frequency $= \frac{\text{Corresponding row total} \times \text{Column total}}{\text{Grand Total}}$

Expected Frequency for 70 $= \frac{120 \times 185}{300} = 74$, Expected Frequency for 80 $= \frac{100 \times 185}{300} = 61.67$,

Expected Frequency for 35 $= \frac{80 \times 185}{300} = 49.33$, Expected Frequency for 50 $= \frac{120 \times 115}{300} = 46$,

Expected Frequency for 20 $= \frac{100 \times 115}{300} = 38.33$, Expected Frequency for 45 $= \frac{80 \times 115}{300} = 30.67$

O_{ij}	E_{ij}	O_{ij} - E_{ij}	$(O_{ij} - E_{ij})^2$	$\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
70	74	-4	16	$\frac{16}{74} = 0.216$
50	46	4	16	0.348
80	61.67	18.33	335.99	5.448
20	38.33	-18.33	335.99	8.766
35	49.33	-14.33	205.35	4.163
45	30.67	14.33	205.35	6.695
Total				25.636

$\therefore \chi^2 = 25.636$

$\alpha = 0.05$ Degrees of freedom $= (r-1)(c-1) = (3-1)(2-1) = 2 \quad \therefore \chi^2_\alpha = 5.991$

Conclusion:

Since $\chi^2 > \chi^2_\alpha$, we Reject our Null Hypothesis H_0. Hence, Conditions of home and conditions of child are not independent.

2. The following contingency table presents the reactions of legislators to a tax plan according to party affiliation. Test whether party affiliation influences the reaction to the tax plan at 0.01 level of signification.

Reaction				
Party	**In favour**	**Neutral**	**Opposed**	**Total**
Party A	**120**	**20**	**20**	**160**
Party B	**50**	**30**	**60**	**140**
Party C	**50**	**10**	**40**	**100**
Total	**220**	**60**	**120**	**400**

Solution:

Null hypothesis H_0 : Party affiliation and tax plan are independent.

Alternate hypothesis H_1 : Party affiliation and tax plan are not independent.

Level of significance: $\alpha = 0.05$

The test statistic: $\chi^2 = \sum_{i=1}^{r} \sum_{i=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

Analysis:

Reaction				
Party	**Infavour**	**Neutral**	**Opposed**	**Total**
Party A	120	20	20	160
Party B	50	30	60	140
Party C	50	10	40	100
Total	220	60	120	**400**

E(120)= $\frac{160 \times 220}{400} = 88$; E(20)= $\frac{160 \times 60}{400} = 24$; E(20)= $\frac{160 \times 120}{400} = 48$

E(50)= $\frac{140 \times 220}{400} = 77$; E(30)= $\frac{140 \times 60}{400} = 21$; E(60)= $\frac{140 \times 120}{400} = 42$

E(50)= $\frac{100 \times 220}{400} = 55$; E(10)= $\frac{100 \times 60}{400} = 15$; E(40)= $\frac{120 \times 100}{400} = 30$

O_{ij}	E_{ij}	O_{ij} - E_{ij}	$(O_{ij} - E_{ij})^2$	$\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
120	88	32	1024	11.64
20	24	-4	16	0.67
20	48	-28	784	16.33
50	77	-27	729	9.47
30	21	9	81	3.86
60	42	18	324	7.71
50	55	-5	25	0.45
10	15	-5	25	1.67
40	30	10	100	3.33
Total				55.13

$\therefore \chi^2 = 55.13$

$\alpha = 0.05$ Degrees of freedom = $(r-1)(s-1) = (3-1)(3-1) = 4 \quad \therefore \chi^2_{0.05} = 13.28$

Conclusion: Since $\chi^2 > \chi^2_\alpha$, we Reject our Null Hypothesis H_0

Hence, the Party Affiliation and tax plan are dependent.

3. From a poll of 800 television viewers, the following data have been accumulated as to, their levels of education and their preference of television stations. We are interested in determining if the selection of a TV station is independent of the level of education

Educational Level				
Public	**High School**	**Bachelor**	**Graduate**	**Total**
Broadcasting	**50**	**150**	**80**	**280**
Commercial Stations	**150**	**250**	**120**	**520**
Total	**200**	**400**	**200**	**800**

(i) State the null and alternative hypotheses.

(ii) Show the contingency table of the expected frequencies. (iii) Compute the test statistic.

(iv) The null hypothesis is to be tested at 95% confidence. Determine the critical value for this test.

Solution:

(i) **Null Hypothesis**: Selection of TV station is independent of level of education

Alternative Hypothesis: Selection of TV station is not independent of level of education

(ii) **Level of significance**: $\alpha = 0.05$

Educational Level				
Public	**High School**	**Bachelor**	**Graduate**	**Total**
Broadcasting	**50**	**150**	**80**	**280**
Commercial Stations	**150**	**250**	**120**	**520**
Total	**200**	**400**	**200**	**800**

To Find Expected frequency:

$$\text{Expected Frequency} = \frac{\text{Corresponding row total} \times \text{Column total}}{\text{Grand Total}}$$

$$\text{Expected Frequency for } 50 = \frac{280 \times 200}{800} = 70, \text{ Expected Frequency for } 150 = \frac{280 \times 400}{800} = 140$$

$$\text{Expected Frequency for } 80 = \frac{280 \times 200}{800} = 70, \text{ Expected Frequency for } 150 = \frac{520 \times 200}{800} = 130$$

$$\text{Expected Frequency for } 250 = \frac{520 \times 400}{800} = 260, \text{ Expected Frequency for } 120 = \frac{520 \times 200}{800} = 130$$

The test statistic: $\chi^2 = \sum_{i=1}^{r} \sum_{i=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

Analysis:

O_{ij}	E_{ij}	O_{ij} - E_{ij}	$(O_{ij} - E_{ij})^2$	$\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
50	70	-20	400	5.714
150	140	10	100	0.174
80	70	10	100	1.428
150	130	20	400	3.076
250	260	-10	100	0.385
120	130	-10	100	0.769
TOTAL				11.546

(iii) Test statistic = 11.546

(iv) Critical Chi-Square = 5.991,

Conclusion: Calculated value > table value

Hence, we reject Null Hypothesis.