unit - 4

Non - parametric test :

It is one of the methods of statistical analysis, which does not require any distribution to meet the required assumptions that has to be analyzed.

i) Run Test :

It is used to identify a non-random pattern in a sequence of elements.

$\underbrace{PPPPP}_{1}$ $\underbrace{FF}_{1}$ $\underbrace{PPP}_{1}$ $\underbrace{F}_{1}$ $\underbrace{P}_{1}$ $\underbrace{FFF\,FFFFF}_{1}$

n_1 , n_2 , r

$n_1 \rightarrow 9$ $r \rightarrow$ no. of runs.

$n_2 \rightarrow 11$ $r \rightarrow 6$.

One sample Run test : (Small sample
($n_1 \leq 20$ & $n_2 \leq 20$)

$\therefore \boxed{n_1 = 5, n_2 = 21}$
Large sample

step 1 :

(pattern)
H_0 : The arrangement is random

H_1 : The " " not " (two tailed)

step 2 :
5% (or) 1%.

Find n_1, n_2 and r

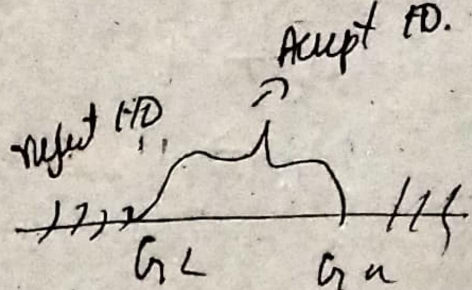
step 3 : conclusion.

Find lower critical value (G_L) & upper critical value using the run table.

If $G_L < r < G_U$ then accept H0.

otherwise reject H0.

7, 18
↓ ↓
G_L G_U



Q) The following is the arrangement of defective (d) and non defective (n) pieces produced in the given order by a certain machine :

$\underline{nnnnn}$ $\underline{dddd}$ $\underline{nnnnnnnnnn}$ $\underline{dd}$ $\underline{nn}$ $\underline{dddd}$ Test for randomness at the 0.01 LOS.

Soln :

$n_1 \rightarrow$ $r \rightarrow$

$n_2 \rightarrow$

step 1 :

H0 : The arrangements are random

H1 : The arrangements are not random

step 2 :

$n_1 \rightarrow$ NO. of symbols of the 1st kind

$n_1 \rightarrow 17$

$n_2 \rightarrow$ No. of symbols of the 2nd kind.

$n_2 \rightarrow 10$.

$r \rightarrow 6$.

step 3 :

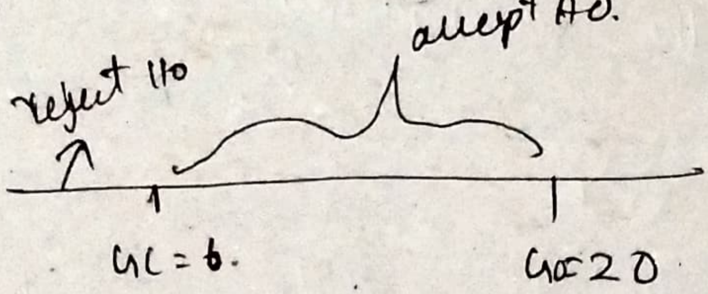
Here $n_1 = 17 < 20$ & $n_2 = 10 < 20$.

(so small sample).

Now find the critical value using run test table at 1% LOS.

from the table, LCV $G_L = 6$

UCV $G_U = 20$



$G_L < r < G_U$

here $G_L = r$ $\therefore$ reject H0. Hence the Arrangement are not random.

2) one sample run test (Small sample)

We have 20 people that enrolled in a drug abuse program. Test the claim that the ages of the people, according to the order in which they enrol occur at random at $\alpha = 0.05$. The Datas are as follows: 18, 36, 19, 22, 25, 44, 23, 27, 27, 35, 19, 43, 37, 32, 28, 43, 46, 19, 20, 22.

soln:

First find the sequence of symbols.

here the median is 27.

[18, 19, 19, 19, 20, 22, 22, 23, 25, 27, 27, 28, 32, 35, 36, 37, 43, 43, 44, 46].

now replace the numbers with 'a' if the observation is > 27 and 'b' if the observation is < 27.

∴ The sequence is.

b a bbb a b ... a b aaaaaa
bbb

step 1 :

H0 : the arrangement is random

H1 : The " " not "

step 2 :

$n_1 \to 9$

$n_2 \to 9$

$r \to 9$

step 3 :

$n_1 = 9 < 20$, & $n_2 = 9 < 20$. so it is small sample.

Now find the critical value using run test table for $\alpha = 0.05$.

here $G_L = 5$ & $G_u = 15$

$G_L < r < G_u$ $\therefore r = 9.$

Hence accept H_0.

3) one sample Run test (large sample) $\begin{bmatrix} n_1 > 20, \\ n_2 > 20 \end{bmatrix}$

Formula :

$$Z = \frac{r - \mu r}{\sigma r},$$ where r = No. of runs (groups) available.

$\mu r \rightarrow$ Mean of r &

$\sigma r \rightarrow$ Standard error of r.

1. H_0, H_1

2. n_1, n_2, r

3. $r = ?$

4. conclusion.

$$\mu r = \frac{2 n_1 n_2}{n_1 + n_2} + 1$$

where n_1 = no. of elements in 1st group

n_2 = " " " " 2nd "

$$\& \ \sigma r = \sqrt{\frac{2 n_1 n_2 (2 n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}$$

Significance value : $z = \cancel{0} 1.5$

$z = 1.5$

1% = 2.58 , 5% = 1.96.

Conclusion :

If $z <$ table value, then accept H_0.

" $z >$ " " " reject H_0

Note : In this test assume

H_0 : Arrangement is Random

H_1 : " " not "

The following is an arrangement of 25 Men (M) and 15 women (W) lined up to purchase tickets for a premier picture show. Test for randomness at 5 % significance.

M WW MMM W MM W M W M WWW MMM W
MM WWW MMMMMM WWW MMMMMM

soln :

step 1 :

H0 : The arrangement is random

H1 : " " " not "

step 2 :

$n_1 \to$ Mens , $n_1 \to 25$

$n_2 \to$ Women , $n_2 \to 15$

$r \to$ groups $r \to 17$

step 3 :

$$Z = \frac{r - \mu r}{\sigma r}.$$

$$\mu r = \frac{2 n_1 n_2}{n_1 + n_2} + 1 = \frac{2(25 \times 15)}{25 + 15} + 1 = 19.75.$$

$$\sigma r = \sqrt{\frac{2 n_1 n_2 (2 n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}$$

$$= \sqrt{\frac{2(25 \times 15)(2(25 \times 15) - 25 - 15)}{(40)^2 (40 - 1)}}.$$

$$= \sqrt{\frac{750 \times 710}{1600 \times 39}} = \sqrt{\frac{532500.}{62,400}}$$

$$= \sqrt{8.53\cancel{4}}$$

$$= 2.922 \Rightarrow Z = \frac{17 - 19.75}{2.922.}$$

$$= -2.75$$

$$= -0.94$$

$Z = 0.94$

step 4 :

$Z = 0.94$ Table value $= 1.96$.

$0.94 < 1.96$.

It accepts H0.

4) Test for the randomness of the following

b b b a a a

set of 26 observations : 24, 35, 12, 50, 60, 70

a b a b b b b a a b a

68, 49, 80, 25, 69, 28, 28, 31, 37, 34, 54, 75, 45, 95,

75, 26, 43, 57, 94, 48.

soln : a b b a a b

$48.5 > b$

$48.5 < a$

step 1 :

Arrange in Ascending order to find Median.

12, 24, 25, 26, 28, 28, 31, 34, 35, 37,

43, 45, $\boxed{48, 49}$, 50, 54, 57, 60, 68, 69,

70, 75, 75, 80, 94, 95

Median = 48.5

The sequence a is above the Median and b is below the Median.

bbb aaaaaa b a bbbbb aa b

aa bb aa b

step 1 :

H0 : The Arrangement is random

H1 : the " " Not " .

step 2 :

$n_1 \rightarrow$ No. of symbols 'a', $n_1 \rightarrow 13$

$n_2 \rightarrow$ " " " 'b', $n_2 \rightarrow 13$

$r \rightarrow$ Total no. of groups, $r \rightarrow 11$

step 3 :

here $n_1 = 13 < 20$ & $n_2 = 13 < 20$ (so it is small sample)

from the Table lower critical value

$G_L = 8$ & upper critical "

$G_U = 20$.

$G_L < V < G_U$

$8 < 11 < 20$.

It accepts the H_0.

∴ The arrangement is random.

5) An engineer is concerned about the possibility that too many changes are being made in the settings of an automatic lathe, given the following mean diameters (in inches) of 40 successive shafts turned on the lathe.

a a b a b a

0.261 0.258 0.249 0.251 0.247 0.256

b a b a

0.250 0.247 0.255 0.243 0.252 0.250

a b a b a a

0.253 0.247 0.251 0.243 0.258 0.251

b b a a

0.245 0.250 0.248 0.252 0.254 0.250

b

0.247 0.253 0.251 0.246 0.249 0.252

a a b b a

0.247 0.250 0.253 0.247 0.249 0.253

b a b b a

0.246 0.251 0.249 0.253.

b a b a

use the 0.01 level of significance to test the null hypothesis of randomness against the alternative that there is a frequently alternating pattern.

soln:

Ascending order.

0.243, 0.243, 0.245, 0.246, 0.246,

0.247, 0.247 0.247, 0.247 0.247,

0.247, 0.248, 0.249, 0.249, 0.249

0.249

0.244.

0.250 , 0.250 , 0.250 , 0.250, 0.250 ,

0.251 , 0.251 , 0.251 , 0.251, 0.251 , 0.251

0.252 , 0.252 , 0.253 , 0.253, 0.253 ,

0.253 , 0.253 , 0.254 , 0.255, 0.256 ,

0.258 , 0.258 , ~~0.178~~

0.261 , ~~0.1769~~.

Median = 0.250.

The sequence 'a' is greater than median and 'b' is less than Median.

aa b a b a b a b aa b

a b aa bb aa b aa

bb ab a bb a b a b a

step 1 :

H0 : the Arrangement is random

H1 : " " " not ".

step 2 :

n1 → The sequence of 'a' , n1 → 19

n2 → " " " 'b' , n2 → 16

r → 27

step 3 :

Here n1 = 19 < 20 , n2 = 16 < 20 so small test.

$\alpha = 0.05$.

(19, 16) ± 10 , 27

$G_L = 10$, $G_u = 27$

$G_L < r < G_u$ ∴ It ~~Accept~~ ~~Accept~~ H0 reject sign Test:

It is used to compere the continuous outcome in the paired samples or the two matches samples.

small sample $\to n \le 25$.

large " $\to n > 25$

$n \to$ no. of positive and -ive sign

step 1 : Write H_0 & H_1

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

$\mu > \mu_0$

$\mu < \mu_0$

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

$\mu_1 > \mu_2$

$\mu_1 < \mu_2$

step 2 : Find n

step 3 : Find p'

$$p' = \begin{cases} P(x \leq k), & \text{if } k \text{ is no. of -ve deviatio} \\ P(x \geq k), & \text{" " " " " +ve "} \end{cases}$$

$$p(x) = nC_x \underset{1/2}{p}^x q^{n-x}$$

$$P(x > \mu) = p(x < \mu)$$

$$\left(\tfrac{1}{2}\right) \qquad \left(\tfrac{1}{2}\right)$$

$$+ \qquad -$$

$$\therefore p(x) = nC_x \left(\tfrac{1}{2}\right)^n$$

step 4: conclusion.

1. If $p' \leq 0.05$ for 5% LOS & $p' \leq 0.01$ " 1% " } Reject H0.

2) If $p' > 0.05$ for 5% LOS & $p' > 0.01$ " 1% " } Accept H0.

Q)

1) The following data constitute a random sample of 15 measurements of the octane rating of a certain kind of gasoline.

99.0 102.3 99.8 100.5 99.7 96.2

99.1 102.5 103.3 97.4 100.4 98.0

101.6

Test the null hypothesis $\mu = 98.0$ against the alternative hypothesis $\mu > 98.0$ at the 0.01 los, conducting a sign test to test the hypothesis.

Soln:

step 1: $H_0 : \mu = 98$

$H_1 : \mu > 98$

step 2: Now replacing each value greater than 98 with '+' sign and less than 98 with '–' ive sign.

$+ + + + + - + + + - + + + 0 +$

n = no. of + and - sign = 14

$(p(x) = nC_x (\frac{1}{2})^n$

step 3 :

$p' = ?$

Let $x = 2$ (No. of - sign)

$\therefore p' = P[x \leq 2]$

$= p(x=0) + p(x=1) + p(x=2)$

$= 14C_0 (\frac{1}{2})^{14} + 14C_1 (\frac{1}{2})^{14} + 14C_2 (\frac{1}{2})^{14}$.

$= \frac{1}{2^{14}} [1 + 14 + 91]$

$p' = 0.0065$.

step 4 : conclusion :

given LOS is 0.01

here $p' = 0.0065 < 0.01$

$\therefore$ Reject H0.

Sign Test for paired Data

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 0.05 LOS to test whether the safety program is effective.

soln :

step 1 :

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$

step 2 : Find n value.

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

$\therefore n = 10$

step 3 : Find p' value.

$$p' = \begin{cases} p(x \leq k) & (\alpha \text{ -ve}) \\ p(x \geq k) & (\alpha \text{ +ve}). \end{cases}$$

$$p' = p(x \leq k)$$

$$= p(x \leq 1)$$

$$= p(x=0) + p(x=1)$$

$$= 10C_0 \left(\tfrac{1}{2}\right)^{10} + 10C_1 \left(\tfrac{1}{2}\right)^{10}$$

$$= \left(\frac{1}{2}\right)^{10} \left[1 + 10\right]$$

$$= \frac{11}{2^{10}}$$

$$= 0.0107$$

step 4 : conclusion :

hence LOS = 0.05.

$p' < 0.05$.

Reject H0.

Sign Test - Large sample ($n > 25$)

step 1 : Write H0 & H1

H0 : $\mu = \mu_0$ | H0 : $\mu_1 = \mu_2$

H1 : $\mu \neq \mu_0$, | H1 : $\mu_1 \neq \mu_2$, $\mu_1 > \mu_2$,

$\mu > \mu_0$, | $\mu_1 < \mu_2$

$\mu < \mu_0$

Mean = np

step 2 : Find n

$\underset{\sigma^2}{\text{var}} = npq$

step 3 : Find z value.

$\because z = \frac{x - \mu}{\sigma}$

$$z = \frac{(x - 0.5) - \frac{n}{2}}{\sqrt{\frac{n}{4}}}, \text{ if } x > \frac{n}{2},$$

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

(or)

$$Z = \frac{(x + 0.5) - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \quad ; \text{ if } x < \frac{n}{2}$$

step 4 : conclusion :

1. If calculated value of z < table value then accept H_0.

2. If calculated value of z > table value then reject H_0.

Table value :

	5%	1%
One tail	1.65	2.33
Two "	1.96	2.58

$\neq$ two

$\gtrless$ one

3) The quality control department of a large manufacturer obtained the following sample data (in pounds) on the breaking strength of a certain kind of 2-inch cotton ribbon:

153, 159, 144, 160, 158, 153, 171, 162, 159, 137, 159, 159, 148, 162, 164, 169, 160, 157, 140, 168, 163, 148, 151, 153, 157, 155, 148, 168, 162, 149.

use the sign test at the 0.01 LOS to test the null hypothesis $\mu = 150$ against the alternative hypothesis $\mu > 150$.

soln:

step 1 : $H_0 : \mu = 150$

$H_1 : \mu > 150$ (one tail)

step 2 : Find n value.

given the median value is 150.

Now replace greater than 150 with '+' and less " " with '−'

$+ \; + \; - \; + \; + \; + \; + \; + \; + \; - \; + \; + \; -$

$+ \; + \; + \; + \; + \; - \; + \; + \; - \; + \; + \; +$

$+ \; - \; + \; + \; -$

$n = 30.$

step 3 : Now find z value.

here $x = \min\{23, 7\}$

$= 7$

we have
$$z = \begin{cases} \dfrac{(x+0.5) - \frac{n}{2}}{\sqrt{\frac{n}{4}}} & \text{if } x < \frac{n}{2} \\[2ex] \dfrac{(x-0.5-\frac{n}{2}}{\sqrt{\frac{n}{4}}} & \text{if } x > \frac{n}{2} \end{cases}$$

$$z = \frac{(x+0.5) - \frac{n}{2}}{\sqrt{\frac{n}{4}}}$$

$$= \frac{(7+0.5)-15}{\sqrt{\frac{30}{4}}}$$

$$z = -2.7386$$

$$\therefore |z| = 2.7386$$

step 4 :

The Table value of z at 0.01 LOS for one tailed test is 2.33

calculated value > Table value

$\therefore$ Reject H0

Hence $\mu > 150$

Rank - Sum Tests

U-Test (given 2 sample) H-Test (more than 2 sample).

Mann - whitney U Test : (small sample)

step 1 :

H0 : $\mu_1 = \mu_2$

H1 : $\mu_1 \neq \mu_2$, $\mu_1 > \mu_2$, $\mu_1 < \mu_2$

step 2 : Find R1 & R2.

R1 → sum of ranks of 1st sample

R2 → " " " " 2nd "

step 3 : Find U1, U2 and U

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

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

where n_1 & n_2 are the no. of elements in 1st & 2nd sample.

$$U = \min\{U_1, U_2\}$$

step 4: write conclusion:

1. If $U <$ table value, then accept H_0.
2. If $U >$ table value, then Reject H_0.

Q) 1) For the following data, test the hypothesis that the median measure in the population x is less than the median measure in the population y, using Mann - whitney U-Test at 5% L.O.S.

X : 60 45 23 32

Y : 10 25 20 54 32 65 8

soln:

step 1 :

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 < \mu_2$

step 2 : Find R_1 & R_2

Data : 8 10 20 23 25 32 ~~40~~ 32 45

54 60 65.

Rank : 1 2 3 4 5 6.5 6.5 8

9 10 11

$\therefore R_1 = 4 + 6.5 + 8 + 10 = 28.5$

$R_2 = 1 + 2 + 3 + 5 + 6.5 + 9 + 11 = 37.5$

step 3 : Find U_1 , U_2 & U.

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

$$= 28.5 - \frac{4(4+1)}{2} = 18.5$$

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

$$= 28.5 - \frac{7(7+1)}{2} = 9.5$$

$\therefore$ $U = \min \{R_1, R_2\}$

$= \min \{18.5, 9.5\}$

$U = 9.5$

step 4 : conclusion :

here $n_1 = 4$, $n_2 = 7$ & LOS = 5%

Table value at 5% LOS for one tailed test is 4 here $\mu >$ table

$9.5 > 4$

$\therefore$ Reject H0.

Mann whitney U-Test (Large sample)

$n_1 > 10$ & $n_2 > 10$

Procedure:

step 1: write H0 & H1

H0: $\mu_1 = \mu_2$

H1: $\mu_1 \neq \mu_2$, $\mu_1 > \mu_2$, $\mu_1 < \mu_2$

step 2: Find R1 & R2

step 3: Find U_1, U_2, U

$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$, $U_2 = R_2 - \frac{n_2(n_2+1)}{2}$

$U = \min\{U_1, U_2\}$

step 4 : Find value of z

$$z = \frac{U - \mu}{\sigma}$$

$$\mu = \frac{n_1 n_2}{2}$$

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

step 5 : conclusion

table value.

	5%	1%
One Tail	1.65	2.33
Two "	1.96	2.58.

1. If C.V of z < T.V then Accept HO

2. " " " z > " " Reject HO.

Q) suppose that in a study of sedimentary rocks, the following diameters (in millimeters) were obtained for two kinds of sand.

Sand I :	0.63	0.47	0.17	1.36	
	0.35	0.51	0.49	0.45	
	0.18	0.84	0.43	0.32	
	0.12	0.40	0.20		
Sand II :	1.13	1.01	0.54	0.48	0.96
	0.89	0.26	1.07	0.39	1.11
	0.88	0.58	0.92	0.53	

use the U test at the 0.01 LOS to test the null hypothesis that the two samples comes from identical populations against the alternative hypothesis that the populations are not identical.

soln :

step 1 :

HO : $\mu_1 = \mu_2$

H1 : $\mu_1 \neq \mu_2$

step 2 : Find R1 & R2

Data : (Ascending)

Data : (Ascending)	0.12	0.17	0.18	0.20	0.26	0.32
	0.35	0.39	0.40	0.43	0.45	0.47
	0.48	0.49	0.51	0.53	0.54	0.58
	0.63	0.84	0.88	0.89	0.92	0.96
	1.01	1.07	1.11	1.13	1.36	
Rank :	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.	

here $n_1 = 15, \quad n_2 = 14$

$R_1 = 1+2+3+4+6+7+9+10+11+12+14+15+19+20+29$

$= 162$

$R_2 = 5+8+13+16+17+18+21+22+23+24+25+26+27+28.$

$= 273$

step 3 : Find U_1, U_2 & U.

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

$$= 162 - \frac{15(15+1)}{2} = 42$$

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

$$= 273 - \frac{14(14+1)}{2} = 168$$

$U = \min\{U_1, U_2\} = \min\{42, 168\}$

$U = 42$

step 4 : Find Z

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

$$\mu = \frac{n_1 n_2}{2} = \frac{15 \times 14}{2} = 105$$

$$\sigma = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{15 \times 14 (15 + 14 + 1)}{12}}$$

$$= 22.9129.$$

$$Z = \frac{42 - 105}{22.9129}$$

$$= -2.7495$$

$\therefore |Z| = 2.7495.$

step 5 : Conclusion,

given LOS = 0.01, Table value of Z at 0.01 LOS for two tailed test is 2.58

$\therefore$ CN > TV

$\therefore$ Reject H0, Hence $\mu_1 \neq \mu_2$

Q) Test whether the following two samples have been draw from the same population using Rank sum test :

sample I : 134, 146, 104, 119, 124, 101, 107, 113, 94

sample II : 70, 101, 118, 85, 107, 132, 94, 97.

soln :

step1 :

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$.

step 2 : Find R1 & R2

Data : 70, 85, 94, 94, 97, 101, 104, 107, 107, 113, 118, 119, 124, 132, 134, 146, 161.

Rank : =. 1, 2, 3.5, 3.5, 5, 6, 7, 8.5, 8.5, 10, 11, 12, 13, 14, 15, 16, 17.

$R1 = 3.5 + 7 + 8.5 + 10 + 12 + 13 + 15 + 16 + 17$

$= 102.$

$R2 = 1 + 2 + 3.5 + 5 + 6 + 8.5 + 11 + 14$

$= 51$

step 3 : Find U1, U2 & U.

$$U1 = 102 - \frac{9(9+1)}{2} = 57$$

$$U2 = 51 - \frac{8(8+1)}{2} = 15$$

$u = \min\{U1, U2\} = \min\{57, 15\} = 15$

step 4 :

$n1 = 9$, $n2 = 8$ LOS = 0.05 & Two Tailed

Table value. = 15

$\therefore$ Accept H0.

C.V $\leq$ T.V..

spearmans Rank correlation coefficient

(It is used to find the rank correlation coeff b/w 2 variables)

formula : $r_s = 1 - \left[\dfrac{6\sum d^2}{n(n^2-1)}\right]$,

where $d \rightarrow$ difference between the ranks of two items. & $n \rightarrow$ no. of observations.

Note :

1. $-1 \leq r_s \leq 1$

2. when $r_s = 1 \Rightarrow$ +ve correlation

3. " " $= -1 \Rightarrow$ -ve "

4. " " $= 0 \Rightarrow$ NO "

5. If there is any ranks repeated, then,

$$r_s = 1 - \frac{6(\sum d^2 + C.F)}{n(n^2-1)}$$, where

C.F $\rightarrow$ correction Factor.

$$C.F = \frac{\sum mi(mi^2-1)}{12}$$

mi → No. of times that the rank repeated.

Q) 1. The following are the average weekly saves of worker hours due to accidents in 10 industrial plants before and 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

calculate Rank correlation:

soln :

here $n = 10$

x	y	Rank of x	Rank of y	$d^2 (x_i - y_j)^2$
45	36	5	5	0
73	60	8	8	0
46	44	6	6	0
124	119	10	10	0
33	35	3	4	1
57	51	7	7	0
83	77	9	9	0
34	29	4	3	1
26	24	2	2	0
17	11	1	1	0
				2

$$r_s = 1 - \frac{6 \Sigma d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(2)}{10(99)}$$

$$= 0.988$$

2) compute the rank correlation coefficient for the following grades of 12 students selected at random.

Maths grade : 85 83 87 84 88 82 90 86 88 87 89 90

Economic " : 87 86 88 86 90 86 85 92 86 88 91

soln : $n = 12$

x	y	Rank of x	Rank of y	$d^2 = (x_i - y_i)^2$
85	88	4	7.5	12.25
83	86	2	3.5	2.25
87	88	6.5	7.5	1
84	86	3	3.5	0.25
88	90	8.5	10	2.25
82	86	1	3.5	6.25
90	88	11.5	7.5	16
86	85	5	1	16
88	92	8.5	12	12.25
87	86	6.5	3.5	9
89	88	10	7.5	6.25
90	91	11.5	11	0.25

$\Sigma d^2 = 84$

In x values,

87 repeated 2 times, so, $CF = \frac{2(2^2-1)}{12} = 0.5$

88 " " " " $CF = \frac{2(2^2-1)}{12} = 0.5$

90 " " " " $CF =$ " = "

In y values,

86 repeated 4 times, so $C.F = \frac{4(4^2-1)}{12} = \frac{15}{3} = 5$

88 " " " $CF = 5.$

$$rs = 1 - \left[\frac{6(\Sigma d^2 + C.F)}{n(n^2-1)}\right]$$

$$= 1 - \left[\frac{6(84 + 0.5 + 0.5 + 0.5 + 5 + 5)}{12(12^2-1)}\right]$$

$$= 1 - \left[\frac{6(95.5)}{12 \times 143}\right]$$

$$rs = 1 - \left[\frac{573}{1716}\right]$$

Kruskal Wallis Test (H - Test)

It is a generalization of the U - Test. It is used to test the null hypothesis that k independent random samples come from identical populations.

Procedure :

step 1 : H0 : The populations are identical.

(i.e $\mu_1 = \mu_2 = \mu_3 = \dots$)

H1 : the populations are not identical.

step 2 : Find R_i

step 3 : Find H

$$H = \frac{12}{n(n+1)}\left[\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \frac{R_3^2}{n_3} + \dots + \frac{R_k^2}{n_k}\right] - 3(n+1)$$

step 4 : conclusion :

First find table value (use χ^2 table)

1. If C V of H $<$ T. V , then Accept H_0.
2. " " " " $>$ " , " Reject "

Note : d.o.f $\gamma = K-1$, $K \rightarrow$ No. of samples.

Q) 1. An experiment designed to compare three methods for preventing corrosion, yielded the following maximum depths of pits (in thousands of an inch) in pieces of wire subjected to the respective treatments :

Method A : 77 54 67 74 71 66

" B : 60 41 59 65 62 64 52

" C : 49 52 69 47 56

use 0.05 LOS to test the null hypothesis that the 3 samples come from the identical population. conducting H test to compare the 3 methods.

soln $\therefore n = n_1 + n_2 + n_3 = 18.$

step 1 :

H0 : population are identical.

H1 : " " not "

step 2 : Find R_1, R_2, R_3.

Values in asc order... } : 41, 47, 49, 52, 52, 54, 56, 59, 60, 62, 64, 65, 66, 67, 69, 71, 74, 77

Rank : 1, 2, 3, 4.5, 4.5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18.

$R_1 = 6 + 13 + 14 + 16 + 17 + 18 = 84$

$R_2 = 1 + 4.5 + 8 + 9 + 10 + 11 + 12 = 55.5$

$R_3 = 2 + 3 + 4.5 + 7 + 15 = 31.5$

step 3 : find H.

$$H = \frac{12}{n(n+1)} \left[\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \frac{R_3^2}{n_3} \right] - 3(n+1).$$

$$= \frac{12}{18(18+1)} \left[\frac{84^2}{6} + \frac{55.5^2}{7} + \frac{31.5^2}{5} \right] - 3(18+1)$$

$$= 0.0351 [1814.486] - 57$$

$$H = 6.688$$

step 4 :

dof = $\gamma = k-1$

$= 3-1$

$\gamma = 2$

LOS = 0.05

$\therefore$ Table value at 0.05 LOS with dof 2 is 5.991

CV > T.V $\therefore$ Reject H0.

Hence, the population are not identical.

Kolmogorov - Smirnov test :

K-S Test For one sample :

procedure :

step 1 :

- H_0 : $F_o(x) = F_n(x)$.

H_1 : $F_o(x) = F_n(x)$.,

where $F_o(x) \rightarrow$ observed cdf

$F_n(x) \rightarrow$ theoretical "

step 2 : Find D value , where

$D = \max | F_o(x) - F_n(x) |$

step 3 : Write conclusion :

1. If $D <$ T.V , then Accept H_0.
2. " $D >$ " " Reject "

Q) 1) It is desired to check whether pinholes in electrolytic tin plate are uniformly distributed across a plated coil on the basis of the following distances in inches of 10 pin holes from one edge of a long strip of tin plate 30 inches wide :

4.8 14.8 28.2 23.1 4.4 28.7 19.5

2.4 25.0 6.2

Test the null hypothesis at the 0.05 LOS using the Kolmogorov - smirnov test for uniformity.

$a < x < b$

$f(x) = \frac{1}{b-a}$

$= \frac{1}{30}$

soln :

step 1 : $H_0 : F_0(x) = F_n(x)$

$H_1 : F_0(x) \neq F_n(x)$

step 2 : Find the statistic D

The pdf of an uniform distribution in the intervals $(0, 30)$ is.

$$f(x) = \frac{1}{30}, \quad 0 < x < 30$$

The cumulative distribution function

$$F_0(x) = p(X \leq x)$$

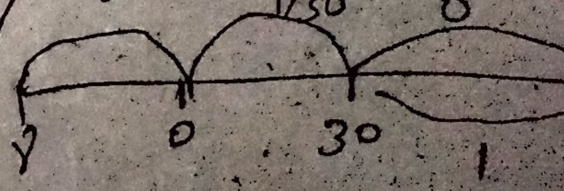
$$= \int_{-\infty}^{0} + \int_{0}^{x} = \int_{-\infty}^{x} f(x)\,dx = \int_{0}^{x} \frac{1}{30}\,dx.$$

$$= \frac{1}{30}(x)_0^x = \frac{x}{30}$$

$$\therefore F_0(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{x}{30} & \text{"} \ 0 < x < 30 \\ 1 & \text{"} \ x \geq 30 \end{cases}$$

and

$$F_n(x) = \left\{ \frac{K}{10}, K = 1, 2, \ldots 10 \right\} \quad f_0(x) = \frac{1}{30}, \quad 0 < x < 3$$



K	x	Fo	Fn	D = |Fo − Fn|
1	2.4	2.4/30 = 0.08	1/10 = 0.1	0.02
2	4.4	0.1467	0.2	0.0533
3	4.8	0.16	0.3	0.14
4	6.2	0.2067	0.4	0.1933
5	14.8	0.4933	0.5	0.0067
6	19.5	0.65	0.6	0.05
7	23.1	0.77	0.7	0.07
8	25	0.8333	0.8	0.0333
9	28.2	0.94	0.9	0.04
10	28.1	0.9557	1	0.0433

$\therefore$ D = max |Fo − Fn| = 0.1933

step 3: n = 10

The TV of D at 0.05 LOS with n = 10, is 0.409.

here D value < Table Value

$\therefore$ Accept H0, Hence Fo(x) = Fn(x)

(i.e) the pinholes are uniformly distributed.

2). In a study done from various streams of a college 60 students with equal no. of students drawn from each stream, the intension of the students to join the Adventure club of college noted after the interviewed are listed bellow:

	B.SC	B.A	B.com	M.A	M.com
No. in each class :	5	9	11	16	19

It was expected that 12 students from each class would join the adventure club. Using the K-S test to find if there is any difference among student classes with regard to their intension of joining the club.

soln :

step 1 :

$H_0 : F_0(x) = F_n(x)$

$H_1 : F_0(x) \neq F_n(x)$

step 2 : Find the ~~value~~ D.

observed frequency	FO [cdf]	Expected frequency	Fn [cdf]	|Fo-Fn|
5	$5/60$	12	$\frac{12}{60}$	$\frac{7}{60}$
9	$\frac{5+9}{60} = \frac{14}{60}$	12	$\frac{24}{60}$	$10/60$
11	$\frac{5+9+11}{60} = \frac{25}{60}$	12	$36/60$	$11/60$
16	$\frac{5+9+11+16}{60} = \frac{41}{60}$	12	$48/60$	$7/60$
19	$\frac{60}{60} = 1$	12	$\frac{60}{60} = 1$	0

$D = \text{Max}\{|F_0 - F_n|\}$

$= \frac{11}{60}$.

$D = 0.1833$

step 3 : conclusion

here $n = 60$, los $= 0.05$

$\therefore$ TV when $n=60$ & $\alpha = 0.05$ is

$$\frac{1.36}{\sqrt{n}} = \frac{1.36}{\sqrt{60}} = 0.1756$$

hence $D >$ table

$\therefore$ Reject H_0.

$\therefore$ $F_0(x) \neq F_n(x)$

K-S Test for Two sample :

[It is used to test whether the 2 samples came from the same distribution or not] :

procedure :

oney distribution la na → K-S Test

orey population 〃 〃 → U or h test

Q) the following Information summarized related to the classification of the men and women with respect to the age group. Determine whether the ture samples come from the same distribution?

Age	21-22	23-24	25-26	27-28	29-30	31-32	33-34	35-36	37-38	39-48
Men :	4	11	5	7	0	5	9	13	20	6
Women :	7	4	1	11	12	4	2	4	8	9

soln :

step 1 : H_0 : $F_1(x) = F_2(x)$

H_1 : 〃 $\neq$ 〃

step 2 : Find D :

Men	Women	F_1	F_2	$\lvert F_1 - F_2 \rvert$
4	7	$4/80 = 0.05$	$7/62 = 0.1129$	0.0629
11	4	$15/80 = 0.1875$	$11/62 = 0.1774$	0.0101
5	1	$20/80 = 0.25$	$12/62 = 0.1935$	0.0565
7	11	$21/80 = 0.3375$	$23/62 = 0.371$	0.0335
0	12	$27/80 = 0.3375$	$35/62 = 0.5645$	0.2270
5	4	$32/80 = 0.4$	$39/62 = 0.629$	0.2290
9	2	$41/80 = 0.5125$	$41/62 = 0.6613$	0.1488
13	4	$54/80 = 0.675$	$45/62 = 0.7258$	0.0508
20	8	$74/80 = 0.925$	$53/62 = 0.8548$	0.0702
6	9	$80/80 = 1$	$62/62 = 1$	0
80	62			

$\therefore D = \max\{\lvert F_1 - F_2 \rvert\} = 0.2290$

step 3 :

here $n_1 = 80$, $n_2 = 62$.

Let the LOS $\alpha = 5\% = 0.05$

$\therefore$ the Table Value $= 1.36 \sqrt{\dfrac{n_1 + n_2}{n_1 n_2}}$

$= 1.36 \sqrt{\dfrac{80+62}{80 \times 62}}$

$= 0.23011$

here $D <$ Table value,

$\therefore$ Accept H_0.

Hence $F_1(x) = F_2(x)$.

The signed - Rank Test

Wilcoxon signed - Rank Test.

Small sample $(n \leq 30)$ ($n \rightarrow$ no. of samples after excluding 0 difference)

procedure:

step 1 : $H0 : \mu = \mu 0$.

$H1 : \mu = \mu 0 \; ; \; \mu > \mu 0 , \; \mu < \mu 0$

step 2 : Find n and R

$R = \min \{ R^- ; R^+ \}$

step 3 : conclusion :

1. If R Value $\leq$ Table value, the reject H0.

2. If R Value > Table ", then accept H0.

One sample :

Following are the responses to the question,

" How many hours do you study before a major statistics test ?

6 5 1 2 2 5 7 5 3 7 4 7

Use the wilcoxon signed rank test to test the hypothesis at 5 % LOS that the median number of hours a student studies before a test is 3.

soln :

step 1 : $H0 : \mu = 3$

$H1 : \mu \neq 3$ (two tailed)

step 2 : n & R

x	$d = x - \text{median}$ $(x-3)$	$\lvert d \rvert$	Rank of $\lvert d \rvert$
6	3	3	8
5	2	2	5.5
1	-2	2	5.5
2	-1	1	2
2	-1	1	2
5	2	2	5.5
7	4	4	10
5	2	2	5.5
3	0	0	—
7	4	4	10
4	1	1	2
7	4	4	10

1 2 3 .

$\frac{4\ 5\ 6\ 7}{4} = 5.5$

$\frac{9\ 10\ 11}{3} = \frac{30}{3}$

here $n = 11$

$R^- = 5.5 + 2 + 2 = 9.5$

$R^+ = 8 + 5.5 + 5.5 + 10 + 5.5 +$
$10 + 2 + 10$
$= 56.5$

$\therefore R = \min\{R^-, R^+\} = 9.5$

step 3 :

here $n = 11$, LOS $= 5\% = 0.05$

$\therefore$ The Table Value for 2 failed test is 10 .

Hence $R <$ Table Value

$\therefore$ Reject H0

$\therefore \mu \neq 3$

Two sample :

A drug is given to 11 patients and difference in their BP were recorded to be :

Before drug A : 112 113 118 120 119 113 110 122 126 115 119

After drug A : 116 120 117 125 126 111 111 117 126 112 129 ~~126 112 129~~

use the wilcoxon signed rank test to test the hypothesis that the drug has no effect on change of BP?

soln : step 1 :

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$ (Two tailed)

step 2 : Find n & R.

Before x	After y	difference $d = x - y$	$\lvert d \rvert$	Rank of $\lvert d \rvert$
112	116	-4	4	5
113	120	-7	7	8.5
118	117	1	1	1.5
120	125	-5	5	6.5
119	126	-7	7	8.5
113	111	2	2	3
110	111	-1	1	1.5
122	117	5	5	6.5
126	126	0	0	—
115	112	3	3	4
119	129	-10	10	10

1 2 3

$\frac{1+2}{2} = 1.5$, 6 7 1 9 9

3 = 2, 10 = 10

3 = 4

4 = 5 6.5

$\frac{6+7}{2} =$

$\frac{8+9}{2} = 8.5$

hence $n = 10$ (small sample

$R^- = 5 + 8.5 + 6.5 + 8.5 + 1.5 +$
10
$= 40$

$R^+ = 1.5 + 3 + 6.5 + 4 = 15$

$R = \min \{40, 15\} = 15$

step 3 : conclusion :

here $n = 10$

Let the LOS is 5% $= 0.05$

The TV for Two tail Test is 8

here k value > Table value

$\therefore$ Accept H_0

$\therefore \mu_1 = \mu_2$

Wilcoxon Signed Rank test (Large sample).

$n > 30$

procedure : (one sample) . (2 sample)

step 1 :	$H_0 : \mu = \mu_0$	$H_0 : \mu_1 = \mu_2$
	$H_1 : \mu_1 \neq \mu_0$	$H_1 : \mu_1 \neq \mu_2$
	$\mu_1 > \mu_0$	" > "
	$\mu_1 < \mu_0$	" < "
step 2 :	find n CR	

step 3 : Find z value

$$Z = \frac{R - \mu}{\sigma}$$

$$\mu = \frac{n(n+1)}{4}$$

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

step 4 :

1. If C. |z| value < T.V , the accept HO

2 " " " " > " " Reject HO.

table value

	5%	1%
one tail	1.65	2.33
Two tail	1.96	2.58

Q)1. A pharmaceutical company is testing to see if there is a significant difference in the pain relief for two new pain medications. They randomly assign the two different pain medications for 34 patients with chronic pain and record the pain rating for each patient one hour after each dose. The results are listed below, use the wilcoscon signed-rank test to see if there is a significant difference at $\alpha = 0.05$

patient:	1	2	3	4	5	6	7	8	9	10
Drug 1:	2.4	4.7	1.2	5.9	4.5	4	2.5	3	5	5.8
Drug 2:	2.5	3.3	5.3	5.6	5	5.3	4.6	2.5	3.4	5.4

patient:

Drug 1:

Drug 2:

soln:

step 1: $H0: \mu1 = \mu2$

$H1: \mu1 \neq \mu2$ (Two Tailed).

step 2: Find n & R.

Drug 1 (x)	Drug 2 (y)	Difference ($d = x - y$)	$\|d\|$
2.4	2.5	-0.1	0.1
4.7	3.3	-1.4	1.4
1.2	5.3	-4.1	4.1
5.9	5.6	0.3	0.3
4.5	5	-0.5	0.5
4	5.3	-1.3	1.3
2.5	4.6	-2.1	2.1
3	2.5	0.5	0.5
5	3.4	1.6	1.6
5.8	5.4	0.4	0.4
1.9	5.1	-3.2	3.2
3.2	4.3	-1.1	1.1
4	6.1	-2.1	2.1
2.2	2.9	-0.7	0.7
2.7	4.3	-1.6	1.6
2.9	3.3	-0.4	0.4
5	5	0	

3.1	5.1	-2	2	23
3.3	3.3	.0	-	-
3	5.9	-2.9	2.9	27
5.4	3.2	2.2	2.2	26
4.2	5.9	-1.7	1.7	21
3.6	5.9	-2.3	2.3	27
2.2	5.6	-3.4	3.4	30
4	5.1	-1.1	1.1	13
5.5	4.4	1.1	1.1	13
3.6	3.6	0		
3.8	3.5	0.3	0.3	25
5.4	4.8	0.6	0.6	8
2.4	3.2	-0.8	0.8	10
4.1	2.6	1.5	1.5	18
4.5	5.7	-1.2	1.2	15
4	5.8	-1.8	1.8	22
6	5	1	1	11

here $n = 31$.

$R^+ = 11 + 2.5 + 6.5 + 19.5 + 4.5 + 26 + 13 +$
$2.5 + 18 + 11 = 128.5$

$R^- = 1 + 31 + 6.5 + 16 + 24.5 + 29 + 13 + 24.5 +$
$9 + 19.5 + 4.5 + 23 + 28 + 21 + 27 + 30 +$
$13 + 10 + 15 + 22 = 367.5$

$\therefore \quad R = \min \{R^+, R^-\} = 128.5$

step 3 :

$$\mu = \frac{n(n+1)}{4} = \frac{31(31+1)}{4} = 248$$

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{31(31+1)(2\times 31+1)}{24}}$$

$$= \sqrt{2604} = 51.0294$$

$$z = \frac{R-\mu}{\sigma} = \frac{128.5-248}{51.0294} = -2.3418$$

$|z| = 2.3418.$

step 4 : conclusion :

T V of z at 5% LOS = 1.96.

C V > T V , reject H0

$\therefore \quad \mu_1 \neq \mu_2$