Non-Parametric Tests
(Practical)

(1) A manufacturer of electric bulbs claims that he has developed a new production process which will increase the mean efficiency (in suitable units) from the present value 9.03, the results obtained from an experiment with 15 bulbs are
9.29, 9.76, 8.93, 10.15, 12.05, 9.02, 8.69, 12.38, 10.87, 11.25, 9.08, 10.00, 11.47, 10.25, 11.56

Ans: $H_0: \theta_0 = 9.03$
against
$H_1: \theta_0 > 9.03$

Rejection rule: $S \geq K_\alpha$, S be the sign test statistic

$S \sim Bin(15, \frac{1}{2})$ under H_0,

i.e, $\sum_{S=K_\alpha}^{15} \binom{15}{S} \left(\frac{1}{2}\right)^{15} \leq 0.05\ (\alpha)$

The condition holds for $K_\alpha = 12$ as

$\sum_{S=12}^{15} \binom{15}{12} \left(\frac{1}{2}\right)^{15} = 0.017 < 0.05$ but

$\sum_{S=11}^{15} \binom{15}{11} \left(\frac{1}{2}\right)^{15} = 0.059 > 0.05$

Therefore from trial-error method we had that $K_\alpha = 12$ and we reject H_0 if $S \geq 12$.

(2) Below are given the marks obtained by a group of 20 students in a subject in a college test and in the subsequent college examination. Test at 1% level whether the group has improved its net performance from the college test to the public exam by using

i) Signed test

ii) Signed rank test (critical value - 43)

Serial No.	Marks in College test (X_i)	Marks in public exam (Y_i)
1	183	133
2	175	192
3	134	170
4	170	164
5	183	199
6	167	160
7	120	168
8	175	158
9	126	162
10	187	176
11	123	126
12	121	141
13	175	103
14	133	126
15	144	146
16	109	155
17	165	162
18	144	161
19	164	182
20	125	119

~~S.No $D_i = |X_i - Y_i|$~~ Ans:-

S. No.	$D_i = Y_i - X_i$	$\|D_i\|$	Rank $\|D_i\|$	Z_i
1	−50	50	19	0
2	18	18	12.5	1
3	36	36	15.5	1
4	−6	6	4.5	0
5	16	16	9	1
6	−7	7	6.5	0
7	48	48	18	1
8	−17	17	10.5	0
9	36	36	15.5	1
10	−11	11	8	0
11	3	3	2.5	1
12	20	20	14	1
13	−72	72	20	0
14	−7	7	6.5	0
15	2	2	1	1
16	46	46	17	1
17	−3	3	2.5	0
18	17	17	10.5	1
19	18	18	12.5	1
20	−6	6	4.5	0

Sign test statistic : S = no. of positive terms among D_i
$= 11$

Signed rank test statistic

$W^+ = \sum_{i=1}^{20} Z_i \,\text{rank}|D_i| = 127$, $W^- = \frac{20 \times (20+1)}{2} - W^+$
$= 83$

$\therefore$ The test statistic $= \min(W^+, W^-)$
$= \min(127, 83) = 83$

Critical value is $43 < \min(W^+, W^-)$

We fail to reject H_0 (null hypothesis)

i.e, students have not improved in public examination

③ A firm is advertising that it has been successful in designing a new home automatic clothes washer which is more effective in removing dirt than the most popular washer now in use. And in support of its claim, it is also displaying the following data of the dirt removed by the most popular washer and the new washer for 16 equally sized and equally soiled loads of clothes.

(X) Popular washer ; 13 10 9 12 11 10 8

(Y) New washer ; 10 11 12 13 9 11 14 12 13

Do you have reasons to believe that the firm's claim is genuine? (Use Mann-Whitney test)

Ans:- The sample size for X and Y $n_1 = 7$, $n_2 = 9$

$n_1 + n_2 = N = 16$

test statistic $U = 1+1+2+4+6 = 14$ (from the combined sample)

$U' = n_1 n_2 - U = 63 - 14 = 49$

and $U^* = \dfrac{14 - \frac{63}{2}}{\sqrt{\frac{63\cdot(7+9+1)}{12}}} = -1.852$

From table VIII of appendix C, we find that for $n_2 = 9$, $n_1 = 7$, for a two tail test at level 0.02, the critical value is 9. Since 49 is greater than 9, we have no reason to believe that the samples are drawn from identical distribution ⇒ Null hypothesis is failed to reject.

(4) The following are the marks secured by two batches of salesman in the final test taken after the completion of training.

Use U-test and Run test for the null hypothesis that the samples are drawn from identical distributions against that the distn differ. in location only.

Batch (A) [X] : 26 27 31 26 19 21 20 25 30

Batch (B) [Y] : 29 28 26 24 22 19

Ans:- $n_1 = 9$ and $n_2 = 6$

$\therefore N = n_1 + n_2 = 15$

Run table

19	19	20	21	22	23	24	25	26	26	26	27	28	30	31	R
X	Y	X	X	Y	Y	Y	X	Y	X	X	X	Y	X	X	9
Y	X	X	X	Y	Y	Y	X	Y	X	X	X	Y	X	X	9
X	Y	X	X	Y	Y	Y	X	X	Y	X	X	Y	X	X	9
Y	X	X	X	Y	Y	Y	X	X	Y	X	X	Y	X	X	8
X	Y	X	X	Y	Y	Y	X	X	X	Y	X	Y	X	X	9
Y	X	X	X	Y	Y	Y	X	X	X	Y	X	Y	X	X	8

max (R) = 9, min (R) = 8

critical value at $n_1 = 9$, $n_2 = 6$ is 4

$\therefore$ Null hypothesis will be accepted.

U-statistic from the combined sample is

$U = 1+1+4+4+4+5+6+6 = 31$

and $U' = n_1 n_2 - U = 23$

as critical value $U'_{9,6} = 7$ and $U'_{cal} > U_{tab}$

Ho will be failed to reject.

⑤ Median test for problem ④

$n_1 = 9$ and $N = n_1 + n_2 = 15$
$n_2 = 6$

19 19 20 21 22 23 24 25 26 26 26 26 27 28 30 31

Median of the combined sample is $\boxed{25}$

No. of Y_i's exceeds 25 is $\boxed{2}$

$\therefore T = 2$

$$P(T=2) = \frac{\binom{8}{4}\binom{7}{2}}{\binom{15}{6}} = 0.293$$

$$P(T=1) = \frac{\binom{8}{5}\binom{7}{1}}{\binom{15}{6}} = 0.0783$$

$$P(T=0) = \frac{\binom{8}{6}\binom{7}{0}}{\binom{15}{6}} = 0.00559$$

$\therefore P(T \leq 2) = 0.3769 = P_1$

$P(T \geq 2) = 0.693 = P_1$

$P(T \neq 2) = 0.3769 \times 2 = 2 \min\{P_1, P_2\}$
$= 0.7538$

P-value for the test is $0.7538 \not< 0.05$

$\therefore$ The null hypothesis is not rejected.

6) The 20 obs: below were chosen randomly from $U(0,1)$ distribution recorded to 4 decimal places. Test the null hypothesis that the square root of observations also have $U(0,1)$ distⁿ.

0.0123, 0.1039, 0.1954, 0.2621, 0.2802, 0.3217, 0.36, 0.3919, 0.4240, 0.4814, 0.5139, 0.5846, 0.6275, 0.6541, 0.6889, 0.7621, 0.8320, 0.8871, 0.9249, 0.9634

Ans: $D_n = \sup_x \left| F_n(x) - F_0(x) \right|$

$$= \max_x \left[\left| F_n(x) - F_0(x) \right|, \left| F_n(x-\epsilon) - F_0(x) \right| \right]$$

$$F_n(x) = \frac{\text{no. of obs.} \leq x}{\text{Total no. of obs.}}$$

H_0: samples are from $U(0,1)$

i.e. $F_0(x) = x$, $0<x<1$

x	$F_n(x)$	$F_0(x)$	$\lvert F_n(x)-F_0(x)\rvert$	$\lvert F_n(x-\epsilon)-F_0(x)\rvert$
0.1109	0.05	0.1109	0.0609	0.1109
0.3223	0.1	0.3223	0.2223	0.2723
0.4420	0.15	0.4420	0.2920	0.3420
0.5119	0.2	0.5119	0.3119	$\boxed{0.3619}$
0.5293	0.25	0.5293	0.2793	0.3293
0.5672	0.3	0.5672	0.2672	0.3172
0.6037	0.35	0.6037	0.2537	0.3037
0.6260	0.4	0.6260	0.2260	0.2760
0.6512	0.45	0.6512	0.2012	0.2512
0.6938	0.5	0.6938	0.1938	0.2438
0.7169	0.55	0.7169	0.1669	0.2169
0.7646	0.6	0.7646	0.1646	0.2146
0.7921	0.65	0.7921	0.1421	0.1921
0.8088	0.7	0.8088	0.1088	0.1588
0.8300	0.75	0.8300	0.0800	0.1300
0.8729	0.8	0.8729	0.0729	0.1229
0.9121	0.85	0.9121	0.0621	0.1121
0.9419	0.9	0.9419	0.0419	0.0919
0.9617	0.95	0.9617	0.0117	0.0617
0.9815	1	0.9815	0.0184	0.0315

$n = 20$, α (level) $= 0.05$ and $D_n = 0.3619$

As observed $D_n > D_n(\text{tab}) = 0.294_{(\alpha = 0.05)}$

$\Rightarrow$ We reject H_0

i.e. samples with root sign are not from Uniform(0,1) distribution at all.

⑦ 1.5 2.3 4.2 7.1 10.4 8.4 9.3 6.5 2.5 4.6

Test whether the data comes from exponential distribution

Ans: For $\exp(\theta)$ distribution

$\hat{\theta} = \frac{1}{\bar{x}}$ and $F_0(x) = 1 - e^{-\hat{\theta} x}$

[since parameter θ is not specified, we take the MLE of θ as an estimate for calculation]

from the data

$\bar{x} = 5.68$

$\therefore \hat{\theta} = \frac{1}{5.68} = 0.1761$

x	$F_n(x)$	$F_0(x)$	$\lvert F_n(x) - F_0(x) \rvert$	$\lvert F_n(x-\epsilon) - F_0(x) \rvert$
1.5	0.1	0.2321	0.1321	0.2321
2.3	0.2	0.3330	0.1330	0.2330
2.5	0.3	0.3561	0.0561	0.1561
4.2	0.4	0.5227	0.1227	0.2227
4.6	0.5	0.5552	0.0552	0.1552
6.5	0.6	0.6817	0.0817	0.1817
7.1	0.7	0.7136	0.0136	0.1136
8.4	0.8	0.7722	0.0278	0.0722
9.3	0.9	0.8056	0.0944	0.0056
10.4	1	0.8398	0.1602	0.0602

$\therefore D_n = \max_x \left[\lvert F_n(x) - F_0(x) \rvert, \lvert F_n(x-\epsilon) - F_0(x) \rvert \right] = 0.2330$

under α (0.05), $D_{10}(\text{tab}) = 0.409 \not< 0.2330$

$\therefore H_0$ is accepted $\Rightarrow$ observations are from exp. distⁿ.

(8) Normal (X): -1.91 -1.22 -0.96 -0.72 0.14 0.82 1.45 1.86

χ^2 (Y): 4.90 7.25 8.04 14.10 18.3 21.21 23.1 28.12

Two mutually independent random samples, each of size 8 were generated [X is from N(0,1), Y is from χ^2_{18}]. Based on the sample, investigate whether $\frac{\chi^2-n}{\sqrt{2n}}$ approaches a N(0,1) dist^n even for moderate d.f.

Ans: X : $-1.91, -1.22, -0.96, -0.72, -0.14, -0.82, 1.45, 1.86$

$\left(\frac{\chi^2-n}{\sqrt{2n}}\right)$ Y$'$: $-2.183, -1.791, -1.66, -0.65, 0.05, 0.535, 0.85, 1.687$

Table

Combined ordered obs. (t)	$\#X \le t$	$\#Y \le t$	$F_{n_1}(t)$	$G_{n_2}(t)$	$\lvert F_{n_1}(t)-G_{n_2}(t)\rvert$
-2.183 (Y)	0	1	0	1/8	1/8
-1.91 (X)	1	1	1/8	1/8	0
-1.791 (Y)	1	2	1/8	1/4	1/8
-1.66 (Y)	1	3	1/8	3/8	1/4
-1.22 (X)	2	3	1/4	3/8	1/8
-0.96 (X)	3	3	3/8	3/8	0
-0.72 (X)	4	3	1/2	3/8	1/8
-0.65 (Y)	4	4	1/2	1/2	0
0.05 (Y)	4	5	1/2	5/8	1/8
0.14 (X)	5	5	5/8	5/8	0
0.54 (Y)	5	6	5/8	3/4	1/8
0.82 (X)	6	6	3/4	3/4	0
0.85 (Y)	6	7	3/4	7/8	1/8
1.45 (X)	7	7	7/8	7/8	0
1.69 (Y)	7	8	7/8	1	1/8
1.86 (X)	8	8	1	1	0

$$\max_{t}\left[F_{n_1}(t) - G_{n_2}(t)\right] = \frac{1}{4}$$

$\therefore$ $P_{H_0}\left(D_{n_1,n_2} \geq \frac{1}{4}\right)$ is the rejection region

$$P_{H_0}\left(n_1 n_2 D_{n_1,n_2} \geq \frac{1}{4} \times 8 \times 8\right) = P_{H_0}\left(n_1 n_2 D_{n_1,n_2} \geq 16\right)$$

$$\geq P_{H_0}\left(n_1 n_2 D_{n_1,n_2} \geq 32\right) = 0.283$$

As the p-value $0.283 > 0.05$

we fail to reject H_0

i.e, with moderate d.f, standardized chi-square converges to $N(0,1)$ in distribution.

(9) In a comparision of the cleaning of four detergents, 20 pieces of white cloth were 1st soiled with ink, the cloths were then washed under control condition with 5 pieces washed by each of the detergents. Whiteness readings are shown below.

Detergent

A	B	C	D
77	74	73	76
86	66	78	85
61	58	57	77
76	63	69	64
69	61	63	80

Test whether hypothesis of no difference between four brands of detergents regarding the average whiteness reading after washing it.

Ans: Combined observation in ascending order with [ranks]

57 (C) [1]
58 (B) [2]
61 (A) [3.5]
61 (B) [3.5]
63 (B) [5.5]
63 (C) [5.5]
64 (D) [7]
66 (B) [8]
69 (C) [9.5]
69 (A) [9.5]
73 (C) [11]
74 (B) [12]
76 (A) [13.5]
76 (D) [13.5]
77 (A) [15.5]
77 (D) [15.5]
78 (C) [17]
80 (D) [18]
81 (A) [19]
85 (D) [20]

$\therefore R_A = 3.5 + 9.5 + 13.5 + 15.5 + 19$
$= 61$

$R_B = 2 + 3.5 + 5.5 + 8 + 12$
$= 31$

$R_C = 44$, $R_D = 74$

Now, $H = \frac{12}{N(N+1)} \sum_{i=A}^{D} \frac{R_i^2}{n_i} - 3(N+1)$

$N = 20$

and reject H_0 if $H > \chi^2_{3,0.05}$

$\therefore H = 6.109$ (from calculations)
$< \chi^2_{3,0.05} = 7.815$

i.e. we accept H_0

(10) The following table shows the life time in hours in excess of thousand hours of the samples of 60W electric light bulbs of 3 different brands. Test at 5% level of significance the hypothesis that there is no difference between the three brands with respect to avg. lifetime.

Brand I	Brand II	Brand III
16	18	26
15	22	31
13	20	24
21	16	30
15	24	24

Combined sample in ascending order with ranks

13 (I)	1	24 (II)	11
15 (I)	2.5	24 (III)	11
15 (I)	2.5	24 (III)	11
16 (I)	4.5	26 (III)	13
16 (II)	4.5	30 (III)	14
18 (II)	6	31 (III)	15
20 (II)	7		
21 (I)	8		
22 (II)	9		

$\therefore R_1 = 1 + 2.5 + 2.5 + 4.5 + 8$
$= 18.5$

$R_2 = 4.5 + 6 + 7 + 9 + 11$
$= 37.5$

$R_3 = 120 - 18.5 - 37.5 = 64$

Now, $H = \frac{12}{N(N+1)} \sum_{i=1}^{3} \frac{R_i^2}{n_i} - 3(N+1)$, $n_i = 5 \; \forall i = 1(1)3$
$N = 15$

$= 10.445 > \chi^2_{2,0.05} = 5.780$

$\therefore$ Reject H_0

(11) An institute of microbiology is interested in purchasing slides of uniform thickness and needs to choose b/w two different suppliers. Both have the same specifications for the median thickness but they may differ in variability. The institute measures the thickness of random sample of 10 slides from each supplier and reports the data shown below as the deviation from specified median thickness. Which supplier makes slides with a smaller variability in thickness?

Supplier-I : 0.028, 0.029, 0.011, -0.030, 0.017, 0.012, -0.027, -0.018, 0.027, -0.023

Supplier-2 : -0.002, 0.016, 0.005, -0.001, 0, 0.008, -0.005, -0.009, 0.005, -0.019

$\Sigma D_{1i} = 0$ $\quad \Sigma D_{6i} = 1$

$\Sigma D_{2i} = 0$ $\quad \Sigma D_{7i} = 0$ $\quad \forall i = 1(1)10$

$\Sigma D_{3i} = 1$ $\quad \Sigma D_{8i} = 1$

$\Sigma D_{4i} = 0$ $\quad \Sigma D_{9i} = 0$ $\quad \therefore \langle T = 3 \rangle$

$\Sigma D_{5i} = 0$ $\quad \Sigma D_{10i} = 0$

$\therefore$ Test statistic

$$\frac{4\sqrt{3}\left(3 - \frac{10 \times 10}{4}\right)}{\sqrt{100 \times 27}} = -2.933 < -\tau_\alpha \quad [\alpha = 0.05, \tau_\alpha = 1.96]$$

As from the critical region rule, Null hypothesis H_0 is rejected at $\alpha = 0.05$.

(12) Two potential suppliers of street lighting equipments A & B presented their bids to the city manager along with the following data as a random sample of life length in months.

A: 35, 66, 58, 83, 71

B: 46, 56, 60, 49

Test whether the life length of suppliers A and B have equal variability assuming their locations to be equal.

Ans: $\sum D_{1i} = 4$ $\quad \forall i = 1(1)4$ $\quad \therefore T = 5$

$\sum D_{2i} = 0$

$\sum D_{3i} = 1$ $\quad \therefore$ The test statistic

$\sum D_{4i} = 0$

$\sum D_{5i} = 0$

$$\frac{4\sqrt{3}\left(T - \frac{n_1 n_2}{4}\right)}{\sqrt{n_1 n_2 (N+7)}}$$

$$= 0 < |\tau_{\alpha/2}|$$

As for rejection it should be

$$|T_{obv}| > |\tau_{\alpha/2}|$$

Hence, the null hypothesis is accepted.