Non-parametric test:→

— Many of the hypothesis tests require that popⁿ is normally distributed. ~~or some~~
But If, for a given test, such requirements cannot be met then what to do?
For these cases, there are hypothesis tests that are distribution free. Such tests are called non-parametric tests.

(1) **Sign Test** :→ One of the simplest test of the entire non-parametric tests.

→ This test is based on the sign (Plus (+) or minus (−)) of the deviations rather than the exact magnitude of the variable values.

→ It is used to test the hypothesis concerning the median for one popⁿ.

Suppose we want to test the hypothesis that

$H_0: \eta = \eta_0$ where η – popⁿ median
η_0 – some specified value.

$H_1: \eta \neq \eta_0$ (Two tailed)
or $H_1: \eta > \eta_0$ (right tailed)
or $H_1: \eta < \eta_0$ (left tailed)

Method:→ (1) Let $X_1, \ldots, X_n$ be a random sample of size n from the given popⁿ with median $\eta = \eta_0$

(2) Subtract η_0 from each X_i's and write
(a) Plus (+) sign if deviation is positive
(b) Minus (−) sign if deviation is negative
(c) Zero (0) if the deviation is zero.

By defⁿ of Median, $P(x > \text{median}) = P(x < \text{median}) = \frac{1}{2}$

ie. $P(x > \eta_0) = P(x < \eta_0) = \frac{1}{2}$ under H_0.

$\therefore$ If H_0 is true, then the no. of (+) signs should be appr. equal to the (−) signs.

After discarding zeros T^+ = no. of + signs T^- = no. of − signs $T = \min(T^+, T^-)$

— Notations to be used.

In Short

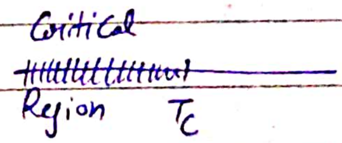
Step I Set the null and Alternative hypothesis
ie. $H_0: \eta = \eta_0$ and $H_1: \eta < \eta_0$ or $\eta > \eta_0$ or $\eta \neq \eta_0$.

Step II Compute T^+ and T^-

Step III Test Statistic $T = \min\{T^+, T^-\}$

Step IV Critical region $T \leq T_c$
where T_c is the critical value of T at the given level of significance for one-tailed or two-tailed.

$\Rightarrow$ If $T \leq T_c$ then Reject H_0
Otherwise donot reject H_0.



For small samples ($n \le 25$)

Ex For $\eta = 5$; Compute the value of T^+, T^- and T for the following data:-

8 9 3 5 4 11

Sol^n- Subtract 5 from each observations and writing the sign as

$+ \quad + \quad - \quad 0 \quad - \quad +$

T^+ = no. of Positive Signs = 3

T^- = no. of negative signs = 2

$T = \min\{T^-, T^+\} = 2$

Que-(1) Following is the data showing - "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 sign test to test the hypothesis at 5% level of significance that the median no. of hours a student studies before a test is 3. Given that the critical value of sign test for $n = 11$ at 5% level of significance for two tailed test is 1.

Sol^n

Step I $H_0: \eta = 3$

$H_1: \eta \neq 3$

Step II Subtract 3 from each observative and writing the sign

$+ \quad + \quad - \quad - \quad - \quad + \quad + \quad + \quad 0 \quad + \quad + \quad +$

~~Step~~ $T^+ = 8$

$T^- = 3$

Step III $T = \min\{3, 8\} = 3$

$n = T^- + T^+ = 8 + 3 = 11$

Critical region is $T \le 1$.

Critical region

$T_c = 1$ (given)

$\because$ $T_{\text{Calculated}} = 3 > 1$

$\Rightarrow$ we donot reject H_0.

Que (2) A teacher claims that the median time to do a particular type of stats Problem is atmost 3 minutes. But here students Claim that the median time is more than 3 minutes. A random sample of 10 students Completed the problem in following times (in minutes)

2.5   2   4   4.5   4   2.5   4.5   3   3.5   5.

Use the sign test with 5% level of significance to test the claim. Given that the critical value of sign test for n=9 at 5% l.o.s for one tailed test is 1.

Sol^n

Step I $H_0: \eta \leq 3$

$H_1: \eta > 3$

Step II $-$ $-$ $+$ $+$ $+$ $-$ $+$ 0 $+$ $+$

$T^+ = 6$

$T^- = 3$

Step III $T = \min\{3, 6\} = 3$

$n = T^+ + T^- = 6+3 = 9$

$T_c = 1$

Critical region

T_c

So critical region is $T \leq 1$

$T_{cal} = 3 > 1$

$\Rightarrow$ do not reject H_0.

ie. teachers claim may be regarded as True.

Date ____
Page ____

For large samples $(n > 25)$

Under H_0: $P(T > \eta_0) = P(T < \eta_0) = 0.5$

where p = prob. of + signs = 0.5

which is Constant for each trial.

$\therefore$ under H_0, the variable $T \sim B(n, p)$

where $p = 0.5$

$\therefore \quad E(T) = np = n \times 0.5 = \frac{n}{2}$

$$\sigma(T) = \sqrt{V(T)} = \sqrt{npq} = \sqrt{\frac{n}{4}} = \frac{\sqrt{n}}{2}.$$

For large samples, Bin $\xrightarrow{\text{appr.}}$ normal distn.

$\therefore$ The test statistics is

If $T^+ < \frac{n}{2}$ then

$$Z_{Stat} = \frac{(T^+ + 0.5) - \frac{n}{2}}{\frac{1}{2}\sqrt{n}} \sim N(0,1)$$

If $T^+ > \frac{n}{2}$ then

$$Z_{Stat} = \frac{(T^+ - 0.5) - \frac{n}{2}}{\frac{1}{2}\sqrt{n}} \sim N(0,1)$$

→ Compare Z_{stat} with Critical value of z at given level of significance.

Que (3) To test the claim that the median age of a professor in State College is atleast 42 years. The results from a random sample of 32 professors gave the following ages (in yrs)

56 62 61 54 52 32 24 35 50 42 52 49
26 31 31 54 38 36 45 53 37 40 38 31 29
25 45 32 49 39 36 38.

Use the sign test at 5% level of significance to test the claim.

Date ______
Page ______

Sol^n Step I $H_0: n \geq 42$

$H_1: n < 42$ (left-tailed)

Step II Subtract 42 from each obs^n and writing the signs, we get

$T^+ = 13$

$T^- = 18$

~~T = min{T~~ $n = T^+ + T^- = 18 + 13 = 31$

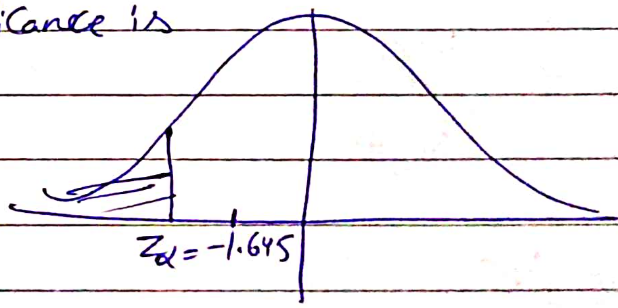
$n = 31 > 25$ (Large Sample)

Step III $T^+ = 13 < \frac{31}{2}$

So Test Statistic is $Z_{stat} = \dfrac{(T^+ + 0.5) - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}$

$= \dfrac{13 + 0.5 - \frac{31}{2}}{\frac{1}{2}\sqrt{31}} = -0.7184$

Z_α at 5% level of Significance is



$Z_{stat} > Z_{value}$

i.e. $-0.7184 > -1.645$

$\Rightarrow$ we don't reject H_0.

H.W

Que (1) The results of a random sample of 35 long distance Calls give the following duration (in minutes)

12 18 23 28 15 10 13 3 6 16 25 35 23

25 28 23 27 28 5 9 37 27 32 19 13 19 18

23 35 4 18 25 31 24 21.

Use the Sign test at 1% level of significance to test the median length of long distance calls is atmost 15 minutes.

$(Z_{stat} = 2.57)$
$(Z_{value} = 2.33)$

Que(5) A random sample of 32 Checking accounts at First State bank gives the following monthly balances (in \$)

185	210	394	150	165	134	165	195	245
164	155	320	175	146	148	152	126	159
164	158	211	215	249	168	146	164	157
231	194	182	168	154.				

Use the sign test at 5% level of significance and test the hypo. that median monthly balance is atleast \$200.

$Z_{stat} = -3.0052$
$Z_{value} = -1.645$

Date ______
Page ______

(2) Wilcoxon Signed-Rank Test:→

- The sign test is not very sensitive to the amount of variation of a data value from the Conjectured value of the mean.
- It uses only the Information whether the value is above or below (+ or −) below the median value.

→ To overcome this, a new test 'Wilcoxon Signed-rank test', in addition to signs, also uses the amount of variation of data values from the Conjectured values.
→ Its better than the Sign test.

Signed Ranks :– (1) Subtract η_0 from each observation.
(2) If difference is zero then omit that observation and reduce the sample size.
(3) Rank the remaining no.s based on their absolute values.
(4) To each rank, we attach the sign (+ or −) and these ranks are Called signed ranks.

T^+ = Sum of Positive number ranks
T^- = Sum of negative number ranks.
$T = \min\{T^+, T^-\}$

Ex For median = 5, Compute T^+, T^-, T for the following data:–
8 9 3 5 4 11

Solⁿ Subtract 5 from each obs., we have
+3 4 −2 0 −1 6

→ Discard '0' and remaining no.s are denoted as Y as

Y	$\lvert Y\rvert$	Rank of $\lvert Y\rvert$
3	3	3
4	4	4
-2	2	2
-1	1	1
6	6	5

T^+ = Sum of +ive no. ranks

$= 3+4+5 = 12$

T^- = Sum of neg. no. ranks

$= 2+1 = 3$

$T = \min\{T^+, T^-\} = 3$

Que $\eta = 8$, Compute T^+, T^- and T for the foll. data.

11 3 8 10 2 5 9 15 16 13

Sol^n Subtract 8 from the given data, we have

3 -5 0 2 -6 -3 1 7 8 5

Discard 0 and rem. no.s are denoted as Y as

Y	$\lvert Y\rvert$	Rank of $\lvert Y\rvert$
3	3	3.5
-5	5	5.5
2	2	2
-6	6	7
-3	3	3.5
1	1	1
7	7	8
8	8	9
5	5	5.5

$T^+ = 29$; $T^- = 16$ and $\boxed{T = 16}$

Date
Page

Wilcoxon test (for small samples $n \le 30$)

Steps I Set H_0 and H_1

II Compute T^+ and T^-

III Test Statistic, $T = \min\{T^-, T^+\}$

IV Critical region $T \le T_c$.

Critical region

T_c

→ In Que 1, use Wilcoxon-Signed rank test.
Critical value of T for n=11 at 5% level is 11.

Sol^n → $H_0: \eta = 3$

$H_1: \eta \ne 3$ (two-tailed)

→ Subtract 3 from each of the observation.

3 2 -2 -1 -1 2 4 2 0 4 1 4

→ Discard 0 and rem. no.s are

Y	|Y|	Rank
3	3	8
2	2	5.5
-2	2	5.5
-1	1	2
-1	1	2
2	2	5.5
4	4	10
2	2	5.5
4	4	10
1	1	2
4	4	10

$T^+ = 56.5$

$T^- = 9.5$

$T = \min\{56.5, 9.5\}$
$= 9.5$

→ $T_c = 11$.

Critical Reg.

11

$\because T < 11$ or $9.5 < 11$

$\Rightarrow$ Reject H_0.

H.W

Que Similary try Que(2) to test the hypothesis when it is given that the critical value of T for $n=9$ at 5% level for single tail test is 8.

For large samples ($n \geq 30$) — normal test

Step (1) Set the hyp. H_0 and H_1.

(2) Compute T^+ and T^-

(3) Compute $Z = \frac{T - \mu_T}{\sigma_T} \sim N(0,1)$

where $T = \min\{T^+, T^-\}$

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

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

(4) If find critical value from the table at given level of sign.

Que Use the Wilcoxon Signed rank test at 5% level to test the claim in Que(3).

Solⁿ → $H_0: n \geq 42$

$H_1: n < 42$ (left tail)

→ Sub. 42 from each value, we get

14 20 19 12 10 −10 −18 −7 8 0 10

7 −16 −11 −11 12 −4 −6 3 11 −5 −2

−4 −11 −13 −17 3 10 6 −3 −12 −4

→ Discard 0 and take their abs. values

|−2| |−3| 3 3 |−4| |−4| |−4| |−5| |−6| 6 |−7|

7 8 |−10| 10 10 10 |−11| |−11| |−11| 11 |−12|

12 12 |−13| 14 |−16| |−17| |−18| 19 20.

→ Rank them and get

1 3 3 3 6 6 6 8 9.5 9.5 11.5
11.5 13 15.5 15.5 15.5 15.5 19.5 19.5 19.5 19.5 23
23 23 25 26 27 28 29 30 31

→ $T^{+} = 239$

& $T^{-} = 257$

n = no. after discarding '0' = 31 > 30

$$Z = \frac{T - \mu_T}{\sigma_T} \sim N(0,1)$$

$$T = \min\{T^{+}, T^{-}\} = 239$$

$$\mu_T = \frac{n(n+1)}{4} = \frac{31(32)}{4} = 248$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = 51.03$$

$$Z = \frac{239 - 248}{51.03} = -0.18$$

$Z > Z_\alpha$

or $-0.18 > -1.645$

⇒ We donot reject H_0.