Two sample Tests: Two sample Run, Kolmogorov-Smirnov two sample, Wilcoxon, mann whitney U test.

1) Two sample test to check if they come from the same distribution.

Hypothesis:

H_0: Two samples come from same distribution $F(x)=G(x)$

vs

H_1: Two samples come from different / not same distribution

$$F(x) \neq G(x)$$

Let $F_n(x)$ & $G_n(y)$ be emprical distribution x and y respectively where.

$$m=n=10$$

$F_n(x)$ is defined by.

$$F_n(x) = \begin{cases} 1 & \text{if } x < x_{(1)} \\ i/n & \text{if } x_{(i)} \leq x \leq x_{i+1} \\ 0 & \text{if } x < x_{(n)} \end{cases}$$

Similarly, emprical dist^n function $G_n(y)$ can be determined.

KS test statistic is given by

$$D_{m,n} = \sup_x \left| f_n(x) - G_n(x) \right|$$

$D_{m,n} > D_{m,n,\alpha}$, we reject H_0.

X	cum(xi)	Fn(xi)	Y	cum(yi)	Fn(yi)
0.6	0.6	0.029	2.1	2.1	0.058
1.2	1.8	0.086	2.3	4.4	0.22
1.6	3.4	0.163	3.0	7.4	0.205
1.7	5.1	0.245	3.1	10.5	0.292
1.7	6.8	0.327	3.2	13.7	0.381
2.1	8.9	0.408	3.2	16.9	0.469
2.8	11.7	0.565	3.5	20.4	0.567
2.9	14.6	0.702	3.8	24.2	0.672
3.0	17.6	0.846	4.6	29.8	0.8
3.2	20.8	1	7.2	36	1

Fn(xi) - Fn(yi)	|Fn(xi) - Fn(yi)|
-0.029	0.029
-0.086	0.086
-0.042	0.042
-0.047	0.047
-0.054	0.054
-0.041	0.041
-0.002	0.002
0.03	0.03
0.046	0.046
0	0.

K-S Test Statistic is 0.054

Table value $T_{m,n,\alpha} = 1.36\sqrt{\frac{m+n}{mn}} \Rightarrow 1.36\sqrt{\frac{20}{100}} = 0.608.$

Conclusion: T (0.054) is less than $T_{m,n,\alpha}$ (0.608)

$0.054 < 0.608$

Accept H_0. Two samples are drawn from same distribution.

2) Two different brands of electric bulbs: brand A & B. To test 2 samples come from same dist^n using K-S two sample test.

Hypothesis:

H_0: Two samples come from same distribution

vs

H_1: Two samples come from different distribution.

Brand A (x) : 80 100 90 101 125 135 70, m = 7

Brand B (y) : 100 120 80 140 130 160 115 120, n = 8.

Items	$B_1(x)$	cu(x)	$B_2(y)$	cu(y)	Fn(xi)	Gn(yi)	|Fn(xi) - Gn(yi)|
70	1	1	0	0	1/7	0/8	0.143
80	1	2	1	1	2/7	1/8	0.161
90	1	3	0	1	3/7	1/8	0.304
100	1	4	1	2	4/7	2/8	0.32
101	1	5	0	2	5/7	2/8	0.464
115	0	5	1	3	5/7	3/8	0.339
120	0	5	2	5	5/7	5/8	0.089
125	1	6	0	5	6/7	5/8	0.232
130	1	7	1	6	1	6/8	0.25
140	0	7	1	7	1	7/8	0.125
160	0	7	1	8	1	1	0.
	7		8				

KS two sample test statistic 0.464.

$D_{m,n,\alpha} = 0.338$ (Table value).

Conclusion:

$D(0.464)$ is greater than $T_{16,16,\alpha}$ (0.388). We reject H_0. we conclude assume of 2 sample come from different distribution.

8)

Hypothesis: $H_0: F(x) = G(x)$ Vs $H_1: F(x) \neq G(x)$

on combining the samples and arranging in ascending order, we get

8	14	14	14	15	18	19	19	19	20	21	21
Y	Y	Y	Y	X	X	X	X	X	Y	Y	X
①				②					③		

22	23	24	26	26	26	26	27	28	28	29	30
Y	Y	Y	X	X	X	Y	Y	X	Y	X	X
④	⑤		⑥			⑦		⑧	⑨	⑩	

30	30	31	31	33	34	34	35	36
Y	Y	X	Y	X	Y	Y	Y	X
⑪				⑫		⑬		⑭

No of Runs = $R = 14$.

$R_{18,16,\alpha} = 11$.

Here Runs (14) is greater than $R_{18,16,\alpha}$ (11)

i.e. $14 > 11$

Hence, we accept H_0.

Since $n > 30$

$24 > 30$.

we proceed z test

Here $n_1 = 18$, $n_2 = 16$

$$E(R) = \frac{2n_1n_2}{n_1+n_2} + 1$$

$$= \frac{2(18)(16)}{34} + 1 = 17.941$$

$$V(R) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)} = 8.667$$

$$z = \frac{R - E(R)}{\sqrt{V(R)}} \sim N(0,1)$$

$$= \frac{14 - 17.941}{\sqrt{8.667}} = -1.33.$$

Reject H_0 if $|z| > z_{\alpha/2}$

$1.33 < 1.96$.

So we accept H_0, Hence 2 population are comes from same distribution.

4) Hypothesis:

H_0: Two random sample are drawn from identical distribution

Vs.

H_1: Two random sample are not drawn from identical distribution.

To combine samples and sort in ascending order,

10	15	17	21	20	22	22	24	25	27	29	30	80
Y	Y	Y	Y	Y	Y	Y	X	X	X	X	X	Y
①							②					

26 28 38 41 45 49 49 50 55 55 58 60
x x x x y y x x
① ② ③ ④ ⑤ ⑥ ⑦ ⑧

60 61 69 70 72
x y x y y
⑨ ⑩ ⑪ ⑫

Here no of runs (R) = 10.

$n_1 = n_2 = 15$

Calculate $E(R) = \frac{2n_1n_2}{n_1+n_2} + 1 \Rightarrow \frac{2(15)(15)}{30} + 1 = 16.$

$$V(R) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)} = \frac{2(15)(15)(2(15\times15) - 15 - 15)}{(15+15)^2(15+15-1)}$$

$= 7.75$

We proceed Z test

$$Z = \frac{R - E(R)}{\sqrt{V(R)}} \Rightarrow \frac{10-16}{\sqrt{7.75}} = -2.1540$$

$|Z| = 2.1540.$

Conclusion:

$|Z|$ is greater than $Z_{\alpha/2}$

$2.154 > 1.96.$

Reject H_0, we conclude 2 samples are not drawn from same distribution.

5) H_0: There is a significance among the scores of two categories (x)

H_1: There is a no significance among scores of two categories.

By using mann-whitney-u test

$n_1 = n_2 = 5.$

We define $\phi(x_i, y_i)$

$$\phi(x_i, y_i) = \begin{cases} 1, & \text{if } x_i < y_i \\ 0, & \text{if } x_i \geq y_i \end{cases}$$

$$u_k = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \phi(x_i, y_i) \quad ; k = 1, 2$$

Thus, u_1 measures the number of times x_i precedes y_j amongst all possible pairs of (x_i, y_i) and u_2 measure no of time y_j precedes x_i amongst all possible pairs of x_i, y_i.

combine both & write in ascending order

10	28	33	48	52	60	89	90	110	125
B	B	B	A	B	B	A	A	A	A

Here $u_1 = 2$

for u_2 = Total combination $- u_1$

$= 25 - 2 = 23$

$\min(u_1, u_2) = \min(2, 23) \Rightarrow 2$

Decision rule

$u > u_{n_1, \alpha}$, we accept H_0

$u \leq u_{n_1, \alpha}$ we reject H_0

$U_{(8,9)}(0.05) = 9$.

$2 > 2$

We reject H_0

Hence, there is no significance amongst the scores in two categories.

6) Hypothesis:

H_0: The scores of english group doesn't differ significantly with that of mother tongue group

vs

H_1: The scores of english group differ significantly with that of mother tongue group

Sort the groups together in ascending order.

Groups	Scores	Ranks
E	10	1.5
M	10	1.5
E	11	3.5
M	11	3.5
M	12	4.5
M	12	4.5
E	13	8
M	13	8
M	13	8
E	15	10.5
M	15	10.5
E	16	12.5
M	16	12.5
E	17	15
E	17	15
M	17	15
E	18	17.

T_1 Sum of ranks of english $\Rightarrow$ 83

T_2 Sum of ranks given to mother tongue $\Rightarrow$ 70

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - T_1$$

$$= 8 \times 9 + \frac{8(8+1)}{2} - 83$$

$$U_1 = 25$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - T_2$$

$$= 8 \times 9 + \frac{9(9+1)}{2} - 70 \Rightarrow 47.$$

KS test statistic is given by

$$U = \min(25, 47) \Rightarrow 25.$$

$U_{(8,9)}(0.05) = 15$.

Conclusion:

$U(25)$ is greater than $U_{n_1 n_2}(0.05)(15)$

$25 > 15$

Accept H_0 (or) fail to reject H_0.

Hence, score of english group differ significantly with that of mother tongue groups.

10/10/25 (10)