Degree of freedom

Number of values in the sample that can be choosen freely. For 'n' sample observations, the degree of freedom is (n-1).

↓

for small sample sized test

Identification of one tailed and two taidled test.

Aug 10, 2024

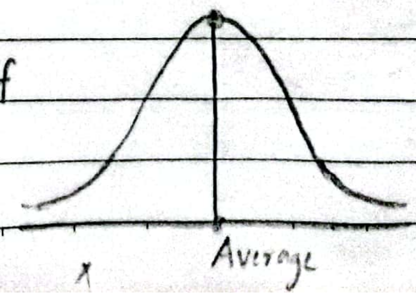
UNIT 6 : Non-parametric test: (NP-Test)

Parametric test
t- test
Z - test
F- test
} Follows normal distribution

* Introduction

Non-parametric tests are defined as those statistical test of hypothesis in which parametric values are not involved.
In other words, those tests of hypothesis are said to be non-parametric test if the hypothesis does not involve any parameter of the population and if measurements are on the nominal or ordinal scale.

As $n \to \infty$,
Z



Scaling
⇓

Nominal
Ordinal
] Qualitative data

Ratio
Interval
} Quantitative data

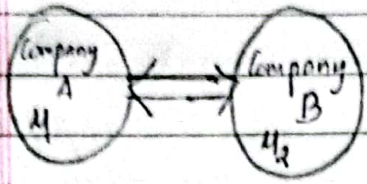
VVI

Difference between parametric and non-parametric test

Parametric	Non-parametric
i) Parametric tests are those tests whose models specify certain conditions about the parameter of the population from which the sample has been drawn.	i) N-P tests are those tests whose models do not specify any conditions about the parameters of the population from which the sample has been drawn.
ii) Parametric tests are mostly applied only to the data which are measured in interval or ratio scale.	ii) N-P tests are applied only to data which are measured in nominal and ordinal scale.
iii) Parametric tests are the most powerful.	iii) N-P tests are less powerful or weaker than the parametric tests.
iv) Parametric tests are designed to test the statistical hypothesis of one or more parameters of the population.	iv) N-P tests are designed only to test statistical hypothesis which does not involve any parameter of the population.

Mann-Whitney U test

Alternative



Small Sample size	Large sample size
n_1 and $n_2 \leq 10$	$n_1 > 10$, $n_2 > 10$
$n_1 + n_2 \leq 20$	$n_1 + n_2 > 20$
	z-test

↳ $Md_1 = Md_2$ ✓

or, $E(X) = E(Y)$

This test is the most powerful non-parametric test for testing hypothesis of difference between two independent location of two independent random samples. This

average

is non-parametric alternative test of t-test for difference of means.

Procedure.

Null hypothesis (H_0): $Md_1 = Md_2$ i.e. there is no significant difference between two sample medians or two popⁿ medians.

Alternative hypothesis (H_1): $Md_1 \neq Md_2$ i.e. there is a significant difference between two sample medians or two population medians. (Two tailed test)

or $H_1: Md_1 > Md_2$ (Right) } one tailed test

or, $H_1: Md_1 < Md_2$ (Left)

Sample size ($n_1 \leq 10$ & $n_2 \leq 10$)

Test-statistic: Under H_0,

$$U_0 = \text{minimum of } \{U_1 \text{ and } U_2\}$$

where $U_1 + U_2 = n_1 n_2$

p-value $\Rightarrow U_0, n_1, n_2$

x_1	x_2
10 (4)	15 (8.5)
18 (11)	12 (6.5)
12 (6.5)	19 (12)
15 (8.5)	25 (13)
17 (10)	6 (2)
5 (1)	9 (3)
	11 (5)
$n_1 = 6$	$n_2 = 7$

for 12, $\frac{6+7}{2} = 6.5$

Common rank

(Take average)

Increasing or ascending order

Rank of x_1 Rank of x_2

Sum of ranks, $R_1 =$ $R_2 =$

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

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

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

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

R_1 = sum of ranks given to first sample data.

R_2 = " " " " " second " ".

P-value > α, then we do not reject H_0 or insignificant

P-value $\leq \alpha$ then we reject H_0 or significant

Critical value: The critical value of the test-statistic is obtained from Mann-Whitney U-test table at α level of significance and $(n_1 \& n_2)$.

Decision: If $U_0 > U_{\alpha,(n_1, n_2)}$ then we do not reject H_0 otherwise reject H_0.

$\alpha = 0.05$

Two tailed

$n_1 = 7, \quad n_2 = 8$

$n_2 \backslash n_1$	7
8	10

$n_2 \backslash n_2$	8
7	10

p-value

$n_1 = 5 \qquad U_0 = 6$

$n_2 = 10$

$\alpha = 0.05$

p-value = 0.0097

Multiply by 2 for two-tailed test.

$2P = \alpha$

The p-value is obtained from Mann-Whitney probability table for (n_1, n_2) and U_0.

For two tailed, P-value = $2P_0$
For one tailed, P-value = P_0

If p-value > α, then we do not reject H_0 otherwise reject H_0.

August 13, 2024 [SMALL SAMPLE SIZE]

Q14.

$n_1 \leq 10$ & $n_2 \leq 10$

$n_1 = 5$
$n_2 = 4$
$\alpha = 0.05$ = level of significance

If level of significance 'α' is not given, then consider 5%.

H_0: $Md_1 = Md_2$ i.e. There is no difference between two average numbers of correction of trained and untrained managers.
H_1: $Md_1 \neq Md_2$ i.e. there is difference between two average numbers of correction of trained and untrained managers (Two tailed test)

Test statistic: Under H_0,
U_0 = minimum of $\{U_1$ and $U_2\}$

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

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

Calculation:

Trained	Untrained	Ranks of trained (Combined)	Ranks of untrained
78	110	7	9
64	70	4	5
75	53	6	3
45	57	1	2
92		8	
		$R_1 = 26$	$R_2 = 19$

Now,

$$U_1 = 5 \times 4 + \frac{5(5+1)}{2} - 26$$

$$= 20 + 15 - 26$$

$$= 9$$

$$U_2 = 5 \times 4 + \frac{4 \times 5}{2} - 19$$

$$= 20 + 10 - 19$$

$$= 11$$

$\therefore U_0 = 9$

Critical value : We have, $\alpha = 0.05$ and $(n_1, n_2) = (5, 4)$

$\therefore U_{(5,4), 0.05}$ (two-tailed test) $= 1$

Decision : Since $U_0 > U_{(5,4), 0.05}$, so we do not reject H_0. Hence, there is no difference between the average number of correction of trained and untrained managers.

p-value = 0.4524
$\therefore$ p-value = 2(0.4524)
= 0.9048
$\alpha = 0.05$
$\because$ p-value > α, so we do not reject H_0.

Assignment
Q15 and 17

Q. The heart beating rate of 5 vegetarians and 5 non vegetarians are recorded below:

Vegetarians:	56	67	82	60	75
Non-vegetarians:	53	42	75	58	65

Is the mean heart beating rate of non vegetarians significantly high. Use Mann Whitney U test at $\alpha = 5\%$.

Solution. $n_1 = 5$, $n_2 = 5$
$\alpha = 0.05$

$H_0: Md_1 = Md_2$ i.e. the mean heart beating rate of non-vegetarian is not significantly high than that of vegetarian.
$H_1: Md_1 < Md_2$ i.e. the mean heart beating rate of non-vegetarian is significantly higher than vegetarian (one tailed test)

Test statistic: Under H_0
U_0 = minimum of $\{U_1, U_2\}$

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

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

Combined

Calculation

Vegetarian	Non-vegetarian	Ranks of vegetarian	Ranks of n
56	53	3	
67	42	7	
82	75	10	
60	58	5	
75	65	8.5	
		$R_1 = 33.5$	

Now,

$$U_1 = 5\times5 + \frac{5\times6}{2} - 33.5$$
$$= 25 + 15 - 33.5$$
$$= 6.5$$

$$U_2 = 5\times5 + \frac{5\times6}{2} - 21.5$$
$$= 25 + 15 - 21.5$$
$$= 18.5$$

$\therefore U_0 = 6.5$

Critical value ; We have $\alpha = 0.05$ and $(n_1, n_2) = (5,5)$
$\therefore U_{(5,5), 0.05}$ (one-tailed test) $= 4$

Decision : Since $U_0 > U_{(5,5), 0.05}$. Hence, ~~there~~ the mean heart beating rate of non-vegetarian is not significantly high than that of vegetarian.

.1111
.1548
.13335 > 0.0

vegetarian

2

1

8.5

4

6

$R_2 = 21.5$

[LARGE SAMPLE SIZE]

For large sample size

$U_0 \sim N(\mu_{U_0}, \sigma^2_{U_0})$

Large sample size i.e. $n_1 > 10$ and $n_2 > 10$

For large sample size, the distribution of U_0 is approximately by normal distribution with mean $= \frac{n_1 n_2}{2}$

and variance $= \frac{n_1 n_2 (n_1+n_2+1)}{12}$

Test statistic: Under H_0

$$Z = \frac{U_0 - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}}$$

In case of tied observation, the corrected value of s.d. is calculated as

$$\sigma_U = \sqrt{\frac{n_1 n_2}{n(n-1)}\left[\frac{n^3-n}{12} - \frac{\Sigma(t_i^3 - t_i)}{12}\right]}$$

where, $n = n_1 + n_2$

t_i = number of times i^{th} rank is repeated

$X \sim N(\mu, \sigma^2)$

$$Z = \frac{X - \mu}{\sigma^2}$$

$U_0 \sim N\left(\frac{n_1 n_2}{2}, \frac{n_1 n_2 (n_1+n_2+1)}{12}\right)$

$$Z = \frac{U_0 - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}}$$

P 123

Q16.

Solution

$n_1 = 11$

$n_2 = 9$

H_0: $Md_1 = Md_2$ i.e. yields of wheats per unit area is not different

H_1: $Md_1 \neq Md_2$ i.e. yields of wheats per unit area is different (Two-tailed test)

H_0: $Md_1 = Md_2$ i.e. there is no significant difference in the average yields between two varieties of wheat I and wheat II.

H_1: $Md_1 \neq Md_2$ i.e. there is significant difference in the average yields between two varieties of wheat I and wheat II (two-tailed test).

Calculation :

Wheat I	Wheat II	Combined Ranks of wheat I	Ranks of
~~15.9~~	~~16.4~~	9	12.5
~~15.3~~	~~16.8~~	5	15
~~16.4~~	~~17.1~~	12.5	17
~~14.9~~	~~16.9~~	3	16
~~15.3~~	~~18.0~~	5	19
~~16.0~~	~~15.6~~	10.5	8
14.6	~~18.1~~	2	20
~~15.3~~	~~17.2~~	5	18
~~14.5~~	15.4	1	7
~~16.6~~		14	
~~16.0~~		10.5	
		$R_1 = 77.5$	$R_2 =$

$t_1 = 3$

$t_2 = 2$

Now,

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

$$= 11\times 9 + \frac{11\times 12}{2} - 77.5$$

$$= 87.5$$

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

$$= 11\times 9 + \frac{9\times 10}{2} - 132.5$$

$$= 99 + 45 - 132.5$$

$$= 11.5$$

$\therefore$ $U_0 =$ minimum of $\{U_1 \;\&\; U_2\}$

$$= 11.5$$

Test-statistic: Since, the problem is large sample size so

$$Z = \frac{U_0 - E(U_0)}{\sqrt{Var(U_0)}}$$

Now, $E(U_0) = \frac{n_1 n_2}{2} = \frac{11(9)}{2} = 49.5$

and $V(U_0) = \frac{n_1 n_2}{n(n-1)}\left[\frac{n^3-n}{12} - \frac{(3^3-3)+(2^3-2)+(2^3-2)}{12}\right]$

$$= 172.46$$

$\therefore$ $Z = \frac{11.5 - 49.5}{\sqrt{172.46}} = -2.89 \Rightarrow |Z\text{-}cal| = 2.89$

at II

132.5

Critical value: We have, $\alpha = 0.01$

$$Z_{tab} = Z_{0.01} \text{ (two-tailed test) } = 2.575$$

Decision: Since $|Z_{cal}| > Z_{tab}$, so we reject H_0.

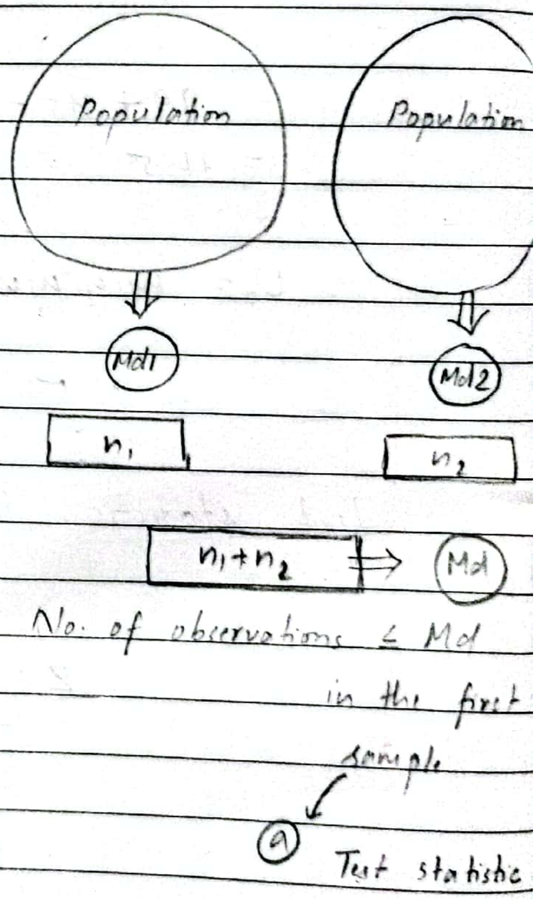
Assignment Q18

August 15, 2024

Median Test

Median Test:

Median test is non-parametric test used to test the difference in medians of two independent distributions. In other words, this test is applied to test whether two independent random samples have been drawn from two populations with same median or not. Also, it is used to test whether the two treatments applied in an experiment are equally effective or not.



Case I:

Small sample size i.e. $n_1 \le 10$ and $n_2 \le 10$.

Null Hypothesis (H_0): $Md_1 = Md_2$
i.e. there is no significant difference between two samples medians.

Alternative Hypothesis (H_1): $Md_1 \neq Md_2$ i.e. there is a significant difference between two sample medians. (Two Tailed test)

or $H_1: Md_1 > Md_2$ } One Tailed test.
or $H_1: Md_1 < Md_2$ }

<u>Test statistic</u>: Under H_0, the test-statistic is obtained as follows:

(i) Combined ~~with~~ both the samples and arrange them in ascending order of magnitude such that $n = n_1 + n_2$.

(ii) Calculate the median of the combined sample and count the number of observations less or equal to median in the first sample which is called the test statistic and is denoted by 'a'.

Now,

$$P_0 = P(A=a)$$

$$= \frac{{}^{n_1}C_a \; {}^{n_2}C_{K-a}}{{}^{n_1+n_2}C_K}$$

where, $a = 0, 1, 2, \ldots,$ ~~min~~ minimum of (n_1, K)

$$K = \frac{n_1+n_2}{2}$$ ← round up when in decimal

<u>Critical value</u>:

We can obtain P-value (P_0), the probability associated with the ~~extreme~~ value as extreme as observed 'a' for n_1 and n_2.

$\therefore P_0 = P(A \geq a)$

<u>Decision</u>

For one-tailed test, P-value $= P_0 \geq \alpha$ then we do not reject H_0 otherwise reject H_0. For two-tailed test, P-value $= 2P_0 > \alpha$, then we do not reject H_0 otherwise reject H_0.

Q7.

$n_1 = 10$ and $n_2 = 10$

H_0 : $Md_1 = Md_2$ i.e. the two training programs are equally effective.

H_1 : $Md_1 \neq Md_2$ i.e. the two training programs are not equally effective. (Two-tailed test)

<u>Test-statistic</u> : Under H_0, the test statistic is 'a'.

<u>Calculation:</u>

Arranging observations of both samples in ascending order of magnitude.

30, 32, 34, 39, 40, 42, 43, 44, 45, 46, 47, 47, 48, 49, 50, 51, [illegible], 55, 55, 59, 71

stem (10)	leaf (1)
3	~~2 9 0 4~~
4	~~6 8 2 8 9 4~~
5	~~1 [illegible] 8 9 7~~
6	
7	~~1~~

Md = Value of $\left(\frac{n+1}{2}\right)^{th}$ term ; $n = n_1 + n_2$

$= \frac{20+1}{2}$

$= 10.5^{th}$ term

$\therefore Md = \frac{46+47}{2} = 46.5$

Now, a = no. of observations less or equal to Md (46.5) in the first sample.

$= 7$

Calculation of P-value

$$P_0 = P(A \geq a) \; ; \; K = \frac{n_1 + n_2}{2} = \frac{10+10}{2} = 10$$

$$= \sum_{a=7}^{10} \frac{{}^{n_1}C_a \; {}^{n_2}C_{K-a}}{{}^{n_1+n_2}C_K}$$

$$= \frac{1}{{}^{20}C_{10}} \left[{}^{10}C_7 \; {}^{10}C_{10-7} + {}^{10}C_8 \; {}^{10}C_{10-8} + {}^{10}C_9 \; {}^{10}C_{10-9} + {}^{10}C_{10} \; {}^{10}C_{10-10} \right]$$

$$= \frac{16526}{184756}$$

$$= 0.0894$$

Decision : $P\text{-value} = 2P_0 = 2 \times 0.089$
$= 0.178$ [$\because$ The problem is two-tailed]

and $\alpha = 0.05$

Since, $P\text{-value} > \alpha$, so we do not reject H_0.
Hence, both the training programs are equally effective.

Assignment 10

Large Sample size ($n_1 > 10$ and $n_2 > 10$)

	No. of obs. $\leq Md$	No. of obs. $> Md$	Total
I^{st} Sample	a	b	a+b
II^{nd} Sample	c	d	c+d
Total	a+c	b+d	a+b+c+d = n

In this case, the median test is equivalent to χ^2-test.
Therefore, the test-statistic is

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi^2(1)$$

Critical value: The critical value of the test statistic is obtained from χ^2-table at α level of significance and 1 dof.

$$\chi^2_{tab} = \chi^2_{\alpha, 1}$$

Decision: If $\chi^2_{cal} \geq \chi^2_{tab}$, then we reject H_0

otherwise we do not reject H_0.

Q9.

$n_1 = 12, \quad n_2 = 12$

$n = n_1 + n_2 = 12 + 12 = 24$

$H_0: Md_1 = Md_2$ i.e. the marks distbn of two teachers do not differ significantly.

$H_1: Md_1 \neq Md_2$ i.e. the marks distbn of two teachers differ significantly (Two-tailed test).

Test statistic: Under H_0,

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Calculation:

Arranging observations of both samples in ascending order of magnitude:

31, 36, 43, 45, 48, 60, 62, 66,
73, 73, 74, 76, 77, 77, 77, 78, 78, 80, 82,
85, 88, 89, 91, 93

S	L
3	6 1
4	3 5 8
5	
6	6 0 2
7	3 3 7 8 4 7 8 6 7
8	9 2 0 8 5

classmate

Assignment 8,10

Md. value of $\left(\frac{n+1}{2}\right)^{th}$ item ; $n = n_1 + n_2$
$= 12 + 12 = 24$

$= \left(\frac{24+1}{2}\right)^{th}$ item

$= 12.5^{th}$ item

$\therefore Md = \frac{76+77}{2}$

$= 76.5$

2x2 Contingency table

	$x_i \le Md$	$x_i > Md$	Total
Sample I	$a = 7$	$b = 5$	12
Sample II	$c = 5$	$d = 7$	12
Total	$(a+c) = 12$	$(b+d) = 12$	24

$$\therefore \chi^2 = \frac{24(7\times 7 - 5\times 5)^2}{12\times 12\times 12\times 12}$$

$= 0.666$

Critical value:

We have, $\alpha = 0.05$ and degree of freedom (df) = 1

$\therefore \chi^2_{tab} = \chi^2_{1, 0.05} = 3.841$

Decision: Since $\chi^2_{\cancel{tab}\ cal} < \chi^2_{\cancel{cal}\ tab}$, so we do not reject H_0. Hence, the marks distbn of two teachers do not differ significantly.

August 17, 20..

Kolmogorov Smirnov Test (K-S Test)
[One sample, two samples]

Kolmogorov-Smirnov (K-S) test is a test of goodness-of-fit. It is alternative to Chi-square test for goodness-of-fit when sample size is small.

χ^2 - test for the goodness-of-fit.

Expected = $x \cdot P(x)$

$f(x) =$

$P(x) =$

$F(x)$ = Cumulative probability distn function.

$-\infty$ ——— a ——— $+\infty$

$F(a) = P(X \leq a)$

$f(x) = \frac{dF(x)}{dx}$

One-Sample

Procedure:

H_0: $F(x) = F_0(x)$ i.e. there is no significant difference between observed and expected frequency. or frequencies are uniformly distributed.

H_1: $F(x) \neq F_0(x)$ i.e. there is significant difference between observed and expected frequency or frequencies are not uniformly distributed (Two Tailed test)

or H_1: $F(x) > F_0(x)$
or H_1: $F(x) < F_0(x)$ } (One Tailed Test).

Test-statistic: Under H_0,

D_0 = Maximum $|F_e(x) - F_0(x)|$

where, $F_e(x) = \frac{cfe}{n}$; cfe = expected cumulative frequency

$n = \Sigma f$ = Total frequency / sample size.

fe = expected frequency = $n p_i$; p_i = probability of each category

$= \frac{\Sigma f}{\text{no. of categories}}$

$n \le 40$ Small
$n > 40$ large

$F_0(x) = \frac{cf_0}{n}$; cf_0 = observed cumulative frequency

Critical value: The critical value of the test-statistic is obtained from K-S table at α level of significance and size of sample 'n'.

$$\therefore D_{tab} = D_{n,\alpha}$$

Decision: If $D_0 \geq D_{n,\alpha}$ then we reject H_0 otherwise do not reject H_0.

Q17.

H_0: The computers of five hard disk are uniformly infected.
H_1: The computers of five hard disk are not uniformly infected. (Two Tailed test).

Test-statistic: Under H_0,

$$D_0 = \text{Maximum} \left| F_e(x) - F_0(x) \right|$$

Calculation

Capacity of hard disk	No. of virus infected (O)	cfo	$F_0(x)$	Expected (E)
500	11	11	11/50	
320	15	26	26/50	
1000	20	46	46/50	
2000	3	49	49/50	
4000	1	50	50/50	

$\longrightarrow$

Calculation :

Capacity of hard disk	No. of virus infected (O)	Cfo	Fo(x)	Expected (E) = np	Cfe	Fe(x)
500	11	11	11/50	10	10	10/50
320	15	26	26/50	10	20	20/50
1000	20	46	46/50	10	30	30/50
2000	3	49	49/50	10	40	40/50
4000	1	50	50/50	10	50	50/50

$\therefore D_0 = 16/50 = 0.32$

Critical value: We have, $\alpha = 0.05$ and $n = 50$

$\therefore D_{n,\alpha}$ (Two Tailed Test) $= D_{50, 0.05}$

(100)

$= \frac{1.36}{\sqrt{50}}$

$= 0.192$

Decision: Since $D_0 > D_{n,\alpha}$, we reject H_0. Hence, the computers of five hard disk are not uniformly infected.

Assignment

Q. A random sample of 20 volume based internet connected have following speed of internet connection in mps:

2.7, 2.9, 3.0, 3.1, 2.8, 3.0, 2.9, 3.0, 2.6, 3.1, 3.2, 3.1, 3.0, 2.9, 3.3, 3.0, 2.8, 2.9, 3.0, 2.9.

Apply Kolmogorov Smirnov test for testing that the internet speed are equally distributed. Use α = ~~0.05~~ 5%.

2.6 1
2.7 1
2.8 2
2.9 5

[cut off] 1
1/50
6/50
16/50
9/50
0

Two Samples Test:

[Small Sample Size], $n_1 = n_2 \leq 40$ or $n_1 \neq n_2 \leq 20$]

[Large Sample size, $n_1 = n_2 > 40$ or $n_1 \neq n_2 > 20$]

Null Hypothesis:

H_0: $F(x) = F(y)$ i.e. the two samples are drawn from same population.

H_1: $F(x) \neq F(y)$ i.e. the two samples are not drawn from same population. (Two Tailed test)

or H_1: $F(x) > F(y)$
or, H_1: $F(x) < F(y)$ } (One Tailed Test)

Small Sample Size

Test-statistic: Under H_0,

$$D_0 = \text{maximum} \left| F(x) - F(y) \right| ; \quad F(x) = \frac{Cfx}{n_1}, \quad F(y) = \frac{Cfy}{n_2}$$

Critical value: The critical values of the test-statistic is obtained from K-S table at α level of significance for sample sizes n_1 and n_2.

$$D_{tab} = D_{(n_1, n_2), \alpha}$$

Different population

Different — No difference ⇒ Same population

Q11. H_0: There is no significant difference between two types of motors.

H_1: There is significant difference between two types of motors.
(Two Tailed Test)

Test-statistic: Under H_0,

$$D_0 = \text{Max}\, |F(x) - F(y)|$$

dif in years	fx	fy	cfx	F(x)	cfy	F(y)	$\lvert F(x)$
2	0	1	0	0/9	1	1/9	1/9
3	3	2	3	3/9	3	3/9	0
4	2	2	5	5/9	5	5/9	0
5	2	1	7	7/9	6	6/9	1/9
6	1	2	8	8/9	8	8/9	0
7	1	1	9	9/9	9	9/9	0
	$n_1 = 9$	$n_2 = 9$					

$$\therefore D_0 = \frac{1}{9}$$

$\therefore$ Critical value: We have, $\alpha = 0.05$, $n_1 = 9$, $n_2 = 9$

$$\therefore D_{(n_1, n_2), \alpha} = D_{(9,9), 0.05} \quad \text{(two-tailed test)}$$

$$= 5/9$$

Decision: Since $D_0 < D_{n_1, n_2, \alpha}$, we don't reject H_0. Thus, the two samples are drawn from same population. (Or there is no significant difference between two motors).

Aug 22, 2024

Large sample size

(i) $n_1, n_2 > 40$

$(n_1 = n_2)$

(ii) $n_1, n_2 > 20$

when, $n_1 \neq n_2$

Large Sample Size:

Test-Statistic:

$D_0 =$ maximum of $|F(x) - F(y)|$ [Two-tailed test]

$\chi^2 = 4D_0^2 \left(\frac{n_1 n_2}{n_1 + n_2}\right)$ [For one-tailed test]

Critical value:

For one tailed test, the critical value of the test-statistic is obtained by χ^2-table at α level of significance and 2 degree of freedom.

Decision

If $D_0 > D_{\alpha, (n_1, n_2)}$, then we reject H_0 Otherwise we do not reject H_0.

If $\chi^2_{cal} \geq \chi^2_{(\alpha, 2)}$, then we reject H_0 otherwise we do not reject H_0.

Q13. $n_1 = 26$, $n_2 = 25$ (Large Sample size)

H_0: There is no significant difference between junior and senior programmers. $F(x) = F(y)$

H_1: There is a significant difference between junior and senior programmers. (Two Tailed Test) $F(x) \neq F(y)$

Test - Statistic : Under H_0,

$$D_0 = Max \mid F(x) - F(y) \mid$$

Calculation

No. of cigarettes	no. of junior prog. (f_x)	no. of senior prog. (f_y)	cf_x	F(x)	F(y)	$\lvert F(x) - F(y)\rvert$
0 - 2	7	5	7	7/26	5/25	0.069
2 - 4	6	4	13	13/26	9/25	0.14 ←
4 - 6	4	6	17	17/26	15/25	0.053
6 - 8	2	3	19	19/26	18/25	0.010
8 - 10	2	4	21	21/26	22/25	0.072
10 - 12	3	2	24	24/26	24/25	0.036
12 - 14	2	1	26	26/26	25/25	0
	$n_1 = 26$	$n_2 = 25$				

$\therefore D_0 = 0.14$

Critical value:

We have $\alpha = 0.05$, $n_1 = 26$, $n_2 = 25$ and the problem is two-tailed,

$$\therefore D_{0.05, (26, 25)} = 1.36 \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

$$= 1.36 \sqrt{\frac{26 + 25}{26 \times 25}}$$

$$= \underline{0.38}$$

Decision

S

we do

Hen

junior

Q. Given

employ

company

Income

20 - 3

30 - 4

40 - 5

50 - 6

60 - 7

70 - 8

80 - 9

classmate
Date
Page

Decision:

Since $D_0 = 0.14 < D_{0.05, (26, 25)} = 0.30$ so, we do not reject H_0.

Hence, there is no significant difference between junior and senior programmers.

$F(y)$	$\lvert F(x) - F(y) \rvert$
5/25	0.069
9/25	0.14 ←
15/25	0.053
19/25	0.010
22/25	0.072
24/25	0.036
25/25	0

Q. Given below represents monthly income distribution of employees in a hardware company and software company.

Income ('000 Rs)	No. of employees in hardware company	No. of employees in software company
20 - 30	6	12
30 - 40	10	18
40 - 50	11	16
50 - 60	13	12
60 - 70	25	10
70 - 80	15	12
80 - 90	10	10

classmate
Date
Page

Assignment

DQ ~~is~~ the income distⁿ support that income of employees in Hardware company is more than income of employees in software company? Use K-S Test at $\alpha = 0.05$.

$$\chi^2_{(2, 0.05)} = 5.991$$

Alternative test, when population is non-normal

Wilcoxon Test

Paired t-test
⇓
Treatment → Object
↓
Effectiveness

Students	1	2	3	4	5	6
Before treatment → Score I	✓	✓	✓	✓	✓	✓
After treatment → Score II	✓	✓	✓	✓	✓	✓

$\mu_1 > \mu_2$

Direction: One Tailed Test. effective

Large sample: $n > 25$ $n_e = n - t \rightarrow$ (difference 0 count)

Individual	X	Y	d	Rank of |d|	Rank with (+)	(−)
1	10	8	2	2.5	2.5	—
2	15	17	−2	2.5	—	2.5
3	12	18	−6	6	—	6
4	13	14	−1	1	—	1
5	16	19	−3	4	—	4
6	20	16	+4	5	5	—
					S(+) =	S(−) =

Test Statistic:

$$T = \min\{S(+), S(-)\}$$

~~Wrot-~~

Wilcoxon Matched Pair Signed Rank Test:

It is used to test the effectiveness of a treatment applied in an experiment.

It is non-parametric test used to compare two populations for which observations are paired.*

Procedure

H_0: $Md_1 = Md_2$ i.e. There is no significant difference in the scores before and after the treatment.

H_1: $Md_1 \neq Md_2$ i.e. There is significant difference in the scores before and after the treatment.

(Two Tailed Test)

$$\left.\begin{array}{l} H_1: Md_1 > Md_2 \\ \text{or } H_1: Md_1 < Md_2 \end{array}\right\} \text{(One Tailed Test)}$$

where, Md_1 = Median score before treatment.

Md_2 = Median score after treatment.

Test-Statistic: Under H_0,

$$T = \text{Minimum of } \{S(+), S(-)\}$$

where,

$S(+)$ = Sum of the ranks of difference with '+' sign.

$S(-)$ = Sum of the ranks of difference with '−' sign.

NOTE: Ranks of $|d|$

where,

d = difference.

classmate

Date

Page

Critical value:

The critical value of the test-statistic is obtained from Wilcoxon Matched Paired Signed Rank Test table at α level of significance and effective sample size (n_e).

$$T_{tab} = T_{\alpha, n_e}$$

where, n_e = effective sample size
$= n - t$

t = no. of differences with zero.

August 24 2024

P140

Q8.

Let, X = Score before course

Y = Score after course

H_0: $Md_1 = Md_2$ i.e. the course is not beneficial

H_1: $Md_1 < Md_2$ i.e. the course is beneficial (One Tailed Test)

Test-statistic: Under H_0,

$$T = \text{minimum of } \{S(+), S(-)\}$$

Calculation

X	Y	$d = X - Y$	Rank of $\lvert d \rvert$	Rank with sign '+'	Rank with sign '-'
57	60	-3	4	–	4
80	90	-10	8	–	8
64	62	+2	3	3	—
70	70	0		—	—
90	95	-5	6.5	—	6.5
59	58	+1	1.5	1.5	—
76	80	-4	5	–	5
98	99	-1	1.5	–	1.5
70	75	-5	6.5	–	6.5
83	94	-11	9	–	9
				$S(+) = 4.5$	$S(-) = 40.5$

24.2024

$\therefore T = 4.5$

(Wilcoxon) This + MW test

If calculated $\le$ Tabulated then we reject H_0

Critical value

We have, $\alpha = 0.05$ and $n_e = n - t$

$= 10 - 1$

$= 9$

$\therefore T_{n_e, \alpha}$ (One Tailed Test) $= T_{9, 0.05}$ (One Tailed Test)

$= 8$

Decision : Since $T = 4.5 < T_{9, 0.05} = 8$, so we reject H_0.

Hence, the course is beneficial.

Q9 (IMP) Let, X: First attempt
Y: Second attempt

H_0: $Md_1 = Md_2$ i.e. The difference between scores is not significant.
H_1: $Md_1 \neq Md_2$ i.e. The difference between scores is significant (Two tailed test)

Test-statistic: Under H_0,
$T =$ minimum of $\{S(+), S(-)\}$

Calculation

X	Y	$d = X - Y$	Rank of $\lvert d \rvert$	Rank with sign +	Rank with sign −
470	510	−40	6		6
530	550	−20	4		4
610	600	+10	1	1	
440	490	−50	7		7
600	585	+15	2	2	
590	620	−30	5		5
580	598	−18	3		3
				$S(+) = 3$	$S(-) = 25$

$\therefore T = 3$

Critical value: We have $\alpha = 0.05$ & $n_e = 7$
$\therefore T_{n_e, \alpha}$ (Two tailed Test) $= T_{7, 0.05}$ (Two tailed Test)

$= 2$

Decision: Since $T = 3 > T_{7, 0.05} = 2$, we accept H_0.
Thus, there is no difference in scores between attempts.

(1 sure Q)

Assign ranks based on overall [illegible]

Kruskal Wallis H-Test (K-W One way ANOVA test)

It is used to test the difference between the locations (Median) of three or more samples.

K = no. of samples ≥ 3

Procedure

	T_1	T_2	T_3	T_4	T_5	
S_1	11	12	13	14	–	R_1
S_2	15	12	13	13	19	R_2
S_3	10	12	14		[illegible]	R_3
S_4	15	14	8	9	18	R_4

$n_1 = 4$, $n_2 = 5$, $n_3 = 3$, $n_4 = 5$, $(K=4)$

$H_0: Md_1 = Md_2 = \dots = Md_K$ i.e. there is no significant difference between k-samples or k-populations.

$H_1: Md_1 \neq Md_2 \neq \dots \neq Md_K$ i.e. there is a significant difference between k-samples or k-populations.

Test-statistic: Under H_0,

$$H = \frac{12}{n(n+1)} \left(\sum_{i=1}^{K} \frac{R_i^2}{n_i} \right) - 3(n+1)$$

where, $n = n_1 + n_2 + \dots + n_K$

n_1 = size of first sample.

n_2 = size of second sample.

. . .

. . .

. . .

n_K = size of k^{th} sample.

R_1 = Sum of the ranks given to first sample

⋮

R_K = Sum of the ranks given to k^{th} sample.

Rank ⇒ Combined rank (i.e. using all the observations).

If there is a tie in the observations, then the corrected value of the test-statistic is obtained as follows:

$$H_{corrected} = \frac{H}{C.F.}$$

$$C.F. = \text{Correction Factor} = 1 - \frac{\sum(t_i^3 - t_i)}{n^3 - n}$$

t_i = no. of times i^{th} rank is repeated.

$$\therefore H_{corr} = \frac{\frac{12}{n(n+1)}\left(\sum_{i=1}^{k} \frac{R_i^2}{n_i}\right) - 3(n+1)}{1 - \frac{\sum(t_i^3 - t_i)}{n^3 - n}}$$

Critical value:

Small sample size:

$n_i \le 5$ and $k = 3$ and the value of test-statistic, we obtain the p-value.

Decision: If P-value > α, then we do not reject H_0. Otherwise, reject H_0.

Large sample size: ($n_i > 5$ and $k > 3$)

The critical value of the test-statistic is obtained from χ^2-table at α level of significance and $(k-1)$ degree of freedom.

$$\therefore \chi^2_{tab} = \chi^2_{\alpha, (K-1)}$$

Decision:

If $H < \chi^2_{tab}$ then we do not reject H_0 otherwise reject H_0.

August 27, 2024

Q14.

Kruskal Wallis H-Test

$K = 3, \quad n_i = 4, \quad \forall i = 1, 2, 3$

H_0: $Md_1 = Md_2 = Md_3$ i.e. All three fertilizers are equally effective.

H_1: $Md_1 \neq Md_2 \neq Md_3$ i.e. All three fertilizers are not equally effective. (or at least one of the fertilizer is different).

Small sample
$n_i \leq 5$ & $K = 3$
p-value

Large sample
$n_i > 5$ and $K > 3$
↓
$\chi^2_{\alpha, (K-1)}$

Calculation

Combined Rank

N	P	K	R_1	R_2	R_3
122 (10)	81 (8)	80 (6)	10	8	6
80 (6)	80 (6)	82 (9)	6	6	9
138 (12)	79 (4)	65 (2.5)	12	4	2.5
121 (11)	65 (2.5)	58 (1)	11	2.5	1
			$\Sigma R_1 = 39$	$\Sigma R_2 = 20.5$	$\Sigma R_3 = 18.5$
			$\Sigma R_1^2/n = \frac{39^2}{4}$		
			$= 380.25$	105.06	85.56

$n = 12$

$t_1 = 2, \quad t_2 = 3$

Test-statistic:

$$H_{corr} = \frac{H}{CF}$$

$$H = \frac{12}{n(n+1)} \left(\sum_{i=1}^{n} \frac{R_i^2}{n} \right) - 3(n+1)$$

$$= \frac{12}{12 \times 13} (380.25 + 105.06 + 85.56) - 3 \times 13$$

$$= 4.91$$

$$C.F. = 1 - \left[\frac{(8-2)+(27-3)}{12^3-12}\right]$$

$$= 1 - \left(\frac{6+24}{12^3-12}\right)$$

$$= 1 - \left(\frac{30}{12^3-12}\right)$$

$$= 0.982$$

$$\therefore H_{corrected} = \frac{4.91}{0.98} = \underline{5.01}$$

Critical value:

For $n_1 = 4$, $n_2 = 4$, $n_3 = 4$ and $H = 5.01$,

$P = 0.097$

Decision: We have, $\alpha = 0.05$ and P-value $= 0.097$. Since, P-value $> \alpha$ so we do not reject H_0. Hence, all three fertilizers are equally effective.

If H_0 is true $\Rightarrow$ insignificant

If H_0 is true $\Rightarrow$ significant

Q13.

H_0: $Md_1 = Md_2 = Md_3$

H_1: $Md_1 \neq Md_2 \neq Md_3$

Calculation.

S(10)	L(1)
6	1 7 9
7	4 9 2 2 6
8	8 7 5 2 4 0 9
9	4 1 7

Method I	94 (17)	88 (14)	91 (16)	74 (6)	87 (13)	97 (18)		$\Sigma R_1 = 84$
Method II	85 (11)	82 (10)	79 (8)	84 (9)	61 (1)	72 (4.5)	80 (8)	$\Sigma R_2 = 55.5$
Method III	89 (15)	67 (2)	72 (4.5)	76 (7)	69 (3)			$\Sigma R_3 = 31.5$

$t_j = 2$

$n = 18$

Test-statistic : Under H_0

$$H_{corrected} = \frac{H}{CF}$$

$$H = \frac{12}{n(n+1)} \left(\sum_{i=1}^{3} \frac{R_i^2}{n_i} \right) - 3(n+1)$$

$$= \frac{12}{18 \times 19} \left(\frac{84^2}{6} + \frac{55.5^2}{7} + \frac{31.5^2}{5} \right) - 3 \times 19$$

$$= \cancel{11.971} \quad 6.666$$

$$C.F. = 1 - \frac{(2^3 - 2)}{18^3 - 18}$$

$$= 1 - \left(\frac{6}{18^3 - 18} \right)$$

$$= 0.9989$$

$$\therefore H_{corr} = \cancel{\frac{11.971}{0.9989}} \cancel{= 11.984} \quad \frac{6.667}{0.998} = 6.68$$

Critical value : We have $K = 3$ and $\alpha = 0.05$

$$\therefore \chi^2_{(K-1), \alpha} = \chi^2_{2, 0.05} = 5.991$$

Decision : Since, $H_{corrected} = 6.68 > \chi^2_{2, 0.05} = 5.991$, we reject H_0. Hence, all three methods are not equally effective.

Assignment

(15), (16), (17)

Friedman Two-Way ANOVA Test

Friedman F-test (at least 3 samples, K≥3)

K = no. of samples

n = size of each sample

	b_1	b_2	b_3	b_4	Sum
K_1					
K_2					
K_3					

BLOCKS (columns), SAMPLE (rows)

We have to find rank of each block separately. (Ascending order)

Then, sum for each sample.

<u>Procedure:</u> $H_0: Md_1 = Md_2 = \ldots = Md_K$ i.e. there is no significant difference between K-sample medians or K-samples are identical

$H_1: Md_1 \neq Md_2 \neq \ldots \neq Md_K$ i.e. there is significant difference between K-sample medians or K-samples are not identical. (or at least one of the sample is different).

<u>Test-statistic</u>: Under H_0,

$$F_{r\,corr.} = \frac{12}{nK(K+1)}\left(\sum_{i=1}^{K} R_i^2\right) - 3n(K+1)$$

If tied case occurs then the corrected value of the test-statistic is

$$F_{corr} = \frac{F_r}{CF} \quad ; \quad C.F. = 1 - \frac{\sum(t_i^3 - t_i)}{n(K^3-K)}$$

t_i = no. of times i^{th} rank is repeated.

<u>Critical value:</u>

[Small] For $[2 \leq n \leq 9$ and $K=3]$ and $[2 \leq n \leq 4$ and $K=4]$

the P-value (P) is obtained from Friedman probability table.

[Large] For $n \geq 5$ and $K > 3$, the critical value is $\chi^2_{\alpha,(K-1)}$

Sum

August 29, 2024

[Q 18, 19, 20]

Q18. $K = 3$, $n = 4$

H_0: $Md_1 = Md_2 = Md_3$ i.e. there is no significant difference in the performances of schools with respect to pass percentage

H_1: $Md_1 \neq Md_2 \neq Md_3$ i.e. there is significant difference in the performances of schools with respect to pass percentage.

Calculation

	Grade				R_i	R_i^2
	I	II	III	IV		
Alpha	89 (3)	98 (3)	70 (3)	80 (3)	$R_1 = 12$	$R_1^2 = 144$
Sigma	45 (2)	76 (2)	40 (2)	55 (1)	$R_2 = 7$	$R_2^2 = 49$
Gamma	20 (1)	58 (1)	35 (1)	67 (2)	$R_3 = 5$	$R_3^2 = 25$
						$\sum R_i^2 = 218$

Test-Statistic: Under H_0,

$$F_r = \frac{12}{nK(K+1)} \left(\sum_{i=1}^{K} R_i^2 \right) - 3n(K+1)$$

$$= \frac{12}{4 \times 3 (3+1)} * (218) - 3 \times 4 \times 4$$

$$= \frac{12}{12 \times 4} * 218 - 48$$

$$= 6.5$$

Q20 Assignment

classmate
Date
Page

P-value : for n=40, K=3 and $F_r = 6.5$,
P-value = 0.042

Decision : We have $\alpha = 0.05$ and P-value = 0.042.
Since, P-value < α so we reject H_0.
Hence, there is significant difference.

Q19

$k=4$, $n=4$

H_0: $Md_1 = Md_2 = Md_3 = Md_4$

H_1: $Md_1 \neq Md_2 \neq Md_3 \neq Md_4$

Calculation

	Winter	Spring	Summer	Fall
A	92 (2)	112 (4)	94 (3)	77 (1)
B	9 (1)	11 (3)	10 (2)	15 (4)
C	58 (2)	71 (4)	57 (1)	62 (3)
D	19 (2.5)	26 (4)	19 (2.5)	18 (1)
R_i	$R_1 = 7.5$	$R_2 = 15$	$R_3 = 8.5$	$R_4 = 9$
R_i^2	$R_1^2 = \cancel{72.25}$ 56.25	$R_2^2 = 225$	$R_3^2 = 72.25$	$R_4^2 = 81$

$t_1 = 2$

$\Sigma R_i^2 = 434.5$

Test-statistic : Under H_0,

$$F_r = \frac{12}{4 \times 4 \times (4+1)} * 434.5 - 3 \times 4 \times 5$$

$$= 5.175$$

$$F_{r\,corr} = \frac{F_r}{CF}$$

$$C.F. = 1 - \frac{2^3 - 2}{4 \cdot (4^3 - 4)} = \cancel{10}\ 0.975$$

$\therefore F_{corr} = \frac{F_r}{CF}$

$= \frac{5.175}{0.975}$

$= 5.307.$

Critical value : for $n = 4$, $K = 4$ and $F_{corr} = 5.307$,
P-value = 0.158

Decision : Since P-value (0.158) > α(0.05), we accept H_0. Thus, there is no significant difference.

Q. Three different advertising media TV, Radio and News paper are being compared to study their effectiveness in promoting sales of Wai-Wai noodles. Each advertising media is exposed for specified period of time and sales ('000 packets) from 10 stores located at different areas are recorded.

	Stores									
Advertising Media	A	B	C	D	E	F	G	H	I	J
TV	20	21	15	12	14	17	21	16	20	18
Radio	7	9	11	12	10	10	14	12	8	7
Newspaper	8	6	11	12	9	6	8	10	8	6

Are three advertising media equally effective, use Friedman two way ANOVA test!

Solution

H_0: $Md_1 = Md_2 = Md_3$ i.e. all three advertising media are equally effective

H_1: $Md_1 \neq Md_2 \neq Md_3$ i.e. all three media are not equally effective.

Calculation

TV	20	21	15	12	14	17	21	16	20	18
Radio	7	9	11	12	10	10	14	12	8	7
Newspaper	8	6	11	12	9	6	8	10	8	6

$R_1^2 = 841$

$R_2^2 = 324$

$R_3^2 = 169$

$\Sigma R_i^2 = 1334$

$t_1 = 2$

$t_2 = 3$

$t_3 = 2$

$$F_r = \frac{12}{10 \times 3 \times 4} * 1334 - 3 \times 10 \times 4$$

$$= 13.4$$

$$C.F = 1 - \frac{(2^3-2) + (3^3-3) + (2^3-2)}{10\,(3^3-3)}$$

$$= 0.85$$