Lab no : 3 **Date : 2081/ /**

Title: Nonparametric Testing of Hypothesis

PROJECT 3.1 (Run Test):

In 30 toss of a coin the following sequence of heads (H) and tail (T) is obtained.

H T T H T H H H T H H T T H T H T H H T H T T H T H H T H T

Test at 0.05 level of significance level whether the sequence is random.

PROCEDURE:

1. Enter the data
2. Select Analyze => Nonparametric Test => Legacy Dialogs => Runs
3. Click Options => Select Descriptive.
4. Click Continue => Ok.

SOLUTION:

STEP I: Null Hypothesis (H_0)
The given sequence of sample observations is in random order.

STEP II: Alternative Hypothesis (H_1)
The given sequence of sample observations is not in random order.

STEP III: Test Statistics

$$Z = \frac{R - \mu_r}{\sigma_r}$$

where,

$$\mu_r = \frac{2n_1n_2}{n_1 + n_2} + 1$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}$$

Calculation (FROM SPSS):

Descriptive Statistics				
	N	***Mean***	***Std. Deviation***	***Std. Error Mean***
Coin Toss	30	.53	.507	0

Runs Test	
	Coin Toss
Test Value[a]	1
Cases < Test Value	14
Cases >= Test Value	16
Total Cases	30
Number of Runs	22
Z	2.078
Asymp. Sig. (2-tailed)	.038

So, P Value = 0.038

STEP IV : Critical Value

$$\alpha = 5\% = 0.05$$

STEP V : Decision

Since P < α , accept H_1 and reject H_0

CONCLUSION:

Hence, the given sequence of sample observations of coin toss is not in random order.

PROJECT 3.2 (Binomial Test):

Test whether the coin is unbiased from following observations.

Tail Tail Head Head Tail Head Tail Head Head Tail Head Head Head Tail Head Head Head Head Head Head Tail Tail Tail Tail Head Tail Tail Tail Tail Tail Tail Tail Head Tail Tail Tail Head Tail Tail Tail Tail Head Tail Tail Head Tail Head Tail Tail Tail

PROCEDURE:

1. Enter the data

2. Select Analyze => Nonparametric Test => Legacy Dialogs => Binomial

3. Click Options => Select Descriptive.

4. Click Continue => Ok.

SOLUTION:

STEP I: Null Hypothesis (H_0)

P = $\frac{1}{2}$

i.e. The coin is unbiased.

STEP II: Alternative Hypothesis (H_1)

P $\neq \frac{1}{2}$

i.e. The coin is biased.

STEP III: Test Statistics

$$Z = \frac{X_0 - \mu_r}{\sigma_r}$$

where,

$$\mu_r = np$$
$$\sigma_r = \sqrt{npq}$$

Since X_0 is discrete, so that continuity correction is made as

$$Z = \frac{(X_0 \pm 0.5) - \mu_r}{\sigma_r}$$

Use +0.5 if X_0 < np else use -0.5

Calculation (From SPSS):

Descriptive Statistics					
	N	***Mean***	***Std. Deviation***	***Minimum***	***Maximum***
Coin	50	0.40	0.495	0	1

Binomial Test						
		Category	***N***	***Observed Prop.***	***Test Prop.***	***Exact Sig. (2-tailed)***
Coin	***Group 1***	***Tail***	30	.60	.50	.203
	Group 2	***Head***	20	.40		
	Total		50	1.00		

So, P Value = 0.203

STEP IV : Critical Value

$$\alpha = 5\% = 0.05$$

STEP V : Decision

Since P > α , accept $\boldsymbol{H_0}$ and reject $\boldsymbol{H_1}$

CONCLUSION:

Hence, the coin is unbiased.

PROJECT 3.3 (One Sample Kolmogorov Smirnov Test):

The number of disease infected tomato plants in 10 different plots of equal size are given below. Test whether the disease infected plants are uniformly distributed over the entire area. Use Kolmogrov Smirnov Test.

Plot no.	1	2	3	4	5	6	7	8	9	10
No. of infected plants	8	10	9	12	15	7	5	12	13	9

PROCEDURE:

1. Enter the data
2. Select Analyze => Nonparametric Test => Legacy Dialogs => 1 Sample K-S
3. Click Options => Select Descriptive.
4. Click Continue => Ok.

SOLUTION:

STEP I: Null Hypothesis (H_0)
The disease infected plants are uniformly distributed over the entire area.

STEP II: Alternative Hypothesis (H_1)
The disease infected plants are not uniformly distributed over the entire area.

STEP III: Test Statistics

$$D_0 = Max\,|F_e - F_o|$$

Calculation (From SPSS):

Descriptive Statistics					
	N	***Mean***	***Std. Deviation***	***Minimum***	***Maximum***
No. of infected plants	10	10.00	3.018	5	15

One-Sample Kolmogorov-Smirnov Test		
		No. of infected plants
N		10
Normal Parameters ^{a,b}	***Mean***	10.00
	Std. Deviation	3.018
Most	***Absolute***	.146
	Positive	.130

Extreme Differences	***Negative***	-0.146
Test Statistic		0.146
Asymp. Sig. (2-tailed)		0.200[c,d]

a. ***Test distribution is Normal.***
b. ***Calculated from data.***
c. ***Lilliefors Significance Correction.***
d. ***This is a lower bound of the true significance.***

So, P Value = 0.200

STEP IV : Critical Value

$$\alpha = 5\% = 0.05$$

STEP V : Decision

Since P > α , accept $\boldsymbol{H_0}$ and reject $\boldsymbol{H_1}$

CONCLUSION:

Hence, the disease infected plants are uniformly distributed over the entire area.

PROJECT 3.4 (Mann Whitney U Test):

Test the hypothesis of no difference between the ages of male and female employees of a certain company, using the Mann- Whitney U test for the samples data below.

Male	35	43	26	44	40	42	33	38	25	26
Female	30	41	34	31	36	32	25	47	28	24

PROCEDURE:

1. Enter the data
2. Select Analyze => Nonparametric Test => Legacy Dialogs => 2 independent sample.
3. Click Options => Select Descriptive.
4. Click Continue => Ok.

SOLUTION:

STEP I: Null Hypothesis (H_0)

$Md_1 = Md_2$

i.e. There is no significant difference between age of male and female employees of the company.

STEP II: Alternative Hypothesis (H_1)

$Md_1 \neq Md_2$

i.e. There is significant difference between age of male and female employees of the company.

STEP III: Test Statistics

$$U_0 = \min\{U_1, U_2\}$$

where,

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

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

Calculation (From SPSS):

Descriptive Statistics					
	N	***Mean***	***Std. Deviation***	***Minimum***	***Maximum***
Age	20	34.00	7.167	24	47
Gender	20	1.50	0.513	1	2

Test Statistics[a]	
	Age of Employee
Mann-Whitney U	38.500
Wilcoxon W	93.500
Z	-0.870
Asymp. Sig. (2-tailed)	0.384
Exact Sig. [2*(1-tailed Sig.)]	0.393[b]

a. Grouping Variable : Gender

b. Not corrected for ties

So, P Value = 0.384

STEP IV : Critical Value

$$\alpha = 5\% = 0.05$$

STEP V : Decision

Since 2P > α , accept $\boldsymbol{H_0}$ and reject $\boldsymbol{H_1}$

<u>CONCLUSION:</u>

Hence, there is no significant difference between age of male and female employees of the company.

PROJECT 3.5 (Median Test):

An IQ test was given to a randomly selected 15 male and 20 female students of a university. Their scores were recorded as follows:

Male: 56, 66, 62, 81, 75, 73, 83, 68, 48, 70, 60, 77, 86, 44, 72

Female: 63, 77, 65, 71, 74, 60, 76, 61, 67, 72, 64, 65, 55, 89, 45, 53, 68, 73, 50, 81

Use median test to determine whether IQ of male and female students is same in the university.

PROCEDURE:

1. Enter the data
2. Select Analyze => Nonparametric Test => Legacy Dialogs => K independent sample.
3. Click Options => Select Descriptive.
4. Define Groups {Gender(1,2)}.
5. Click Continue => Ok.

SOLUTION:

STEP I: Null Hypothesis (H_0)

$Md_1 = Md_2$

i.e. The IQ of male and female students is same in the university.

STEP II: Alternative Hypothesis (H_1)

$Md_1 \neq Md_2$

i.e. The IQ of male and female students is not same in the university.

STEP III: Test Statistics

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

Calculation (From SPSS):

Descriptive Statistics					
	N	***Mean***	***Std. Deviation***	***Minimum***	***Maximum***
IQ	35	67.1429	11.314	44	89
Gender	35	1.57	0.502	1	2

Test Statistics[a]		
		IQ
N		35
Median		68.00
Chi-Square		0.614
df		1
Asymp. Sig.		0.433
Yates' Continuity Correction	***Chi-Square***	0.194
	df	1
	Asymp. Sig.	0.659

a. Grouping Variable : Gender

So, P Value = 0.659

STEP IV : Critical Value

$$\alpha = 5\% = 0.05$$

STEP V : Decision

Since P > α , accept H_0 and reject H_1

CONCLUSION:

Hence, The IQ of male and female students is same in the university.

PROJECT 3.6 (Wilcoxon Matched Pair Signed Rank Test):

Use Wilcoxon matched pair signed rank test to determine the equality of effectiveness of two types of drugs in suppressing pain from following data.

Patient No.	***Drug A***	***Drug B***	***Patient No.***	***Drug A***	***Drug B***
1	6.5	3.5	***11***	5.4	5.5
2	3.7	3.7	***12***	4.0	4.1
3	3.9	4.7	***13***	5.7	4.1
4	6.7	5.0	***14***	3.9	4.2
5	6.2	5.6	***15***	3.6	3.7
6	6.7	4.3	***16***	4.9	4.1
7	6.1	5.4	***17***	3.9	5.4
8	4.3	5.8	***18***	5.8	3.7
9	5.5	4.3	***19***	4.9	4.1
10	6.8	4.3	***20***	3.9	4.1

PROCEDURE:

1. Enter the data

2. Select Analyze => Nonparametric Test => Legacy Dialogs =>2 Related sample.

3. Click Options => Select Drug A and Drug B (in test pair variable 1 - drug A and variable 2 - drug B)

4. Click Continue => Ok.

SOLUTION:

STEP I: Null Hypothesis (H_0)

$Md_1 = Md_2$

i.e. Both Drug A and Drug B are equally effective in suppressing the pain.

STEP II: Alternative Hypothesis (H_1)

$Md_1 \neq Md_2$

i.e. Both Drug A and Drug B are not equally effective in suppressing the pain.

STEP III: Test Statistics

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

where,

S (+) = Sum of ranks of + sign

S (-) = Sum of ranks of - sign

<u>Calculation (From SPSS):</u>

Descriptive Statistics					
	N	***Mean***	***Std. Deviation***	***Minimum***	***Maximum***
Drug A	20	5.170	1.1164	3.6	6.8
Drug B	20	4.480	0.7157	3.5	5.8

Test Statistics[a]	
	Drug B – Drug A
Z	-2.076[b]
Asymp. Sig. (2-tailed)	.038

a. Grouping Variable : Gender

b. Based on positive ranks

So, P Value = 0.038

STEP IV : Critical Value

$$\alpha = 5\% = 0.05$$

STEP V : Decision

Since P > α , accept H_0 and reject H_1

<u>CONCLUSION:</u>

Hence, Both Drug A and Drug B are equally effective in suppressing the pain.

PROJECT 3.7 (Friedman F Test):

A survey was conducted in four hospitals in a particular city to obtain the number of babies born over a 12 months' period. This time period was divided into four seasons to test the hypothesis that the birth rate is constant over all the four seasons. The results of the survey were as follows:

	No. of Births			
Hospital	***Winter***	***Spring***	***Summer***	***Fall***
A	92	72	94	77
B	15	16	10	17
C	58	71	51	62
D	19	26	20	18

Analyze the data using Friedman two ANOVA test.

PROCEDURE:

1. Enter the data
2. Select Analyze => Nonparametric Test => Legacy Dialogs =>K Related sample.
3. Click Test Variable => Select Friedman.
4. Click Continue => Ok.

SOLUTION:

STEP I: Null Hypothesis (H_0)

$Md_1 = Md_2 = Md_3 = Md_4$

i.e. The birth rate is constant over all four seasons.

STEP II: Alternative Hypothesis (H_1)

$Md_1 \neq Md_2 = Md_3 = Md_4$

i.e. The birth rate isn't constant over all four seasons.

STEP III: Test Statistics

$$F_r = \frac{12}{nk(k+1)} \sum_{i=1}^{k} R_i^2 - 3n(k+1)$$

If tie occurs, then test statistics is

$$F_r = \frac{\frac{12}{nk(k+1)} \sum_{i=1}^{k} R_i^2 - 3n(k+1)}{1 - \sum \frac{(t_i{}^3 - t_i)}{n(k^3 - k)}}$$

Calculation (From SPSS):

<table>
<tr><th colspan="4">Test Statistics[a]</th></tr>
<tr><td colspan="3">N</td><td>4</td></tr>
<tr><td colspan="3">Chi-Square</td><td>.900</td></tr>
<tr><td colspan="3">df</td><td>3</td></tr>
<tr><td colspan="3">Asymp. Sig.</td><td>.825</td></tr>
<tr><td rowspan="3">Monte Carlo Sig.</td><td colspan="2">Sig.</td><td>.750</td></tr>
<tr><td rowspan="2">95% Confidence Interval</td><td>Lower Bound</td><td>.326</td></tr>
<tr><td>Upper Bound</td><td>1.000</td></tr>
</table>

a. Friedman Test

So, P Value = 0.825

STEP IV : Critical Value

$$\alpha = 5\% = 0.05$$

STEP V : Decision

Since P > α , accept $\boldsymbol{H_0}$ and reject $\boldsymbol{H_1}$

CONCLUSION:

Hence, the birth rate is constant over all four seasons.