Sarfraj Alam

# Lab 1: For Ungrouped Data

Enter the following values in SPSS and calculate mean, s.d, ranges, mode, median.

Weights: 25, 35, 45, 55, 65, 75

**Basic Steps for SPSS**

1. Select Analyze ->Descriptive statistics ->Descriptives

2. Click the Descriptives-> Move Midvalue into Variable(s).

3. Click the Option. select Mean, s.d, ranges, mode, median

4. Click Ok

**Working Expression:**

1. **Mean:** Mean= Sum of all values divided by the number of values
   Mean= $\sum x/n$
2. **Median:** Arrange the data in ascending order.

   If n is odd: Median=Middle value

   If n is even: Median=(Middle two values sum)/2

3. **Standard Deviation:** Square root of the average of squared differences from the mean

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

4. **Range:** Range=Maximum value - Minimum value

**Output:**

Weights

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 25.00 | 1 | 14.3 | 16.7 | 16.7 |
| | 35.00 | 1 | 14.3 | 16.7 | 33.3 |
| | 45.00 | 1 | 14.3 | 16.7 | 50.0 |
| | 55.00 | 1 | 14.3 | 16.7 | 66.7 |
| | 65.00 | 1 | 14.3 | 16.7 | 83.3 |
| | 75.00 | 1 | 14.3 | 16.7 | 100.0 |
| | Total | 6 | 85.7 | 100.0 | |
| Missing | System | 1 | 14.3 | | |

| | | | | |
|---|---|---|---|---|
| Total | 7 | 100.0 | | |

Statistics

Weights

| | | |
|---|---|---|
| N | Valid | 6 |
| | Missing | 1 |
| Mean | | 50.0000 |
| Median | | 50.0000 |
| Mode | | 25.00[a] |
| Std. Deviation | | 18.70829 |
| Range | | 50.00 |

a. Multiple modes exist. The smallest value is shown

**Interpretation:**

1. Mean=50, mean the average value of weights is 50.
2. Median= 50, means the 50% of the wight is above 50 and 50% is below 50.
3. Mode=25, means the majority of times the weight is 25.
4. Std Deviation=18.708, means 18.708 is the average dispersion from mean.
5. Range= 50, means the difference between highest and lowest weight is 50.

# Lab 2: For Ungrouped Data

Enter the following values in SPSS and calculate mean, s.d., ranges, mode, median.

| Weight | Midvalue | Frequency |
|---|---|---|
| 20-30 | 25 | 4 |
| 30-40 | 35 | 6 |
| 40-50 | 45 | 7 |
| 50-60 | 55 | 21 |
| 60-70 | 65 | 23 |
| 70-80 | 75 | 2 |

**Basic Steps**:

1. Enter the Data Editor Window.

2. Select Data-> Weight Cases.

3. Move Frequency into Frequency Variable.

4. Click Ok. select Analyze ->Descriptive statistics-> Frequencies

5. Click the frequencies-> Move Midvalue into Variable(s)

6. Click the Statistics. select Mean.

7. Click continue. Click ok.

**Working Expression:**

1. **Mean:** Mean= Sum of all values divided by the number of values

   Mean= $\sum x/n$
2. **Median:** Arrange the data in ascending order.

   If n is odd: Median=Middle value

   If n is even: Median=(Middle two values sum)/2

3. **Standard Deviation**: Square root of the average of squared differences from the mean

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

4. **Range:** Range=Maximum value - Minimum value

Sarfraj Alam

midvalue

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 25.00 | 1 | 16.7 | 16.7 | 16.7 |
| | 35.00 | 1 | 16.7 | 16.7 | 33.3 |
| | 45.00 | 1 | 16.7 | 16.7 | 50.0 |
| | 55.00 | 1 | 16.7 | 16.7 | 66.7 |
| | 65.00 | 1 | 16.7 | 16.7 | 83.3 |
| | 75.00 | 1 | 16.7 | 16.7 | 100.0 |
| | Total | 6 | 100.0 | 100.0 | |

Statistics

| | | midvalue | frequency |
|---|---|---|---|
| N | Valid | 6 | 6 |
| | Missing | 0 | 0 |
| Mean | | 50.0000 | 10.5000 |
| Median | | 50.0000 | 6.5000 |
| Mode | | 25.00[a] | 2.00[a] |
| Std. Deviation | | 18.70829 | 9.09395 |
| Range | | 50.00 | 21.00 |

a. Multiple modes exist. The smallest value is shown

frequency

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 2.00 | 1 | 16.7 | 16.7 | 16.7 |
| | 4.00 | 1 | 16.7 | 16.7 | 33.3 |

| | | | | | |
|---|---|---|---|---|---|
| | 6.00 | 1 | 16.7 | 16.7 | 50.0 |
| | 7.00 | 1 | 16.7 | 16.7 | 66.7 |
| | 21.00 | 1 | 16.7 | 16.7 | 83.3 |
| | 23.00 | 1 | 16.7 | 16.7 | 100.0 |
| | Total | 6 | 100.0 | 100.0 | |

**Interpretation.**

For Mid-value.

1. Mean=50, mean the average value of weights is 50.
2. Median= 50, means the 50% of the wight is above 50 and 50% is below 50.
3. Mode=25, means the majority of times the weight is 25.
4. Std Deviation=18.708, means 18.708 is the average dispersion from mean.
5. Range= 50, means the difference between highest and lowest weight is 50.

For Frequency.

1. Mean=10.5, mean the average value of weights is 10.5.
2. Median= 6.50, means the 50% of the wight is above 6.50 and 50% is below 6.50.
3. Mode=2, means the majority of times the weight is 2.
4. Std Deviation=9.093, means 9.093 is the average dispersion from mean.
5. Range= 21, means the difference between highest and lowest weight is 21.

Sarfraj Alam

# Lab 3: Confidence Interval for Population Mean μ,(σ$^2$ Unknown and large n)

Enter the following values in SPSS and create a confidence interval

assuming normal distribution:

Length: 125, 120, 121, 123, 122, 130, 124, 122, 120, 122, 118, 119, 123, 124, 122, 124, 121, 122, 138, 149, 123, 128, 122, 130, 120, 122, 124, 134, 137, 128, 122, 121, 125, 120, 132, 130, 128, 130, 122, 124.

**Basic Steps for SPSS**

1. Enter the data.
2. Select Analyze ->Compare Means-> One sample T test.
3. Click Options-> Type % (90, 95, 99) confidence interval
4. Click on Continue and then Click OK.

**Working Expression**

95%

One-Sample Statistics

| | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|
| Length | 40 | 125.2750 | 6.14770 | .97204 |

One-Sample Test

| | Test Value = 0 | | | | | |
|---|---|---|---|---|---|---|
| | | | | | 95% Confidence Interval of the Difference | |
| | t | df | Sig. (2-tailed) | Mean Difference | Lower | Upper |
| Length | 128.879 | 39 | .000 | 125.27500 | 123.3089 | 127.2411 |

90%

Sarfraj Alam

One-Sample Statistics

| | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|
| Length | 40 | 125.2750 | 6.14770 | .97204 |

One-Sample Test

| | Test Value = 0 | | | | | |
|---|---|---|---|---|---|---|
| | | | | | 90% Confidence Interval of the Difference | |
| | t | df | Sig. (2-tailed) | Mean Difference | Lower | Upper |
| Length | 128.879 | 39 | .000 | 125.27500 | 123.6372 | 126.9128 |

99%

One-Sample Statistics

| | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|
| Length | 40 | 125.2750 | 6.14770 | .97204 |

One-Sample Test

| | Test Value = 0 | | | | | |
|---|---|---|---|---|---|---|
| | | | | | 99% Confidence Interval of the Difference | |
| | t | df | Sig. (2-tailed) | Mean Difference | Lower | Upper |
| Length | 128.879 | 39 | .000 | 125.27500 | 122.6428 | 127.9072 |

**Interpretation:**

1. For 95% confidence level confidence interval is (123.3089, 127.2411).
2. For 90% confidence level confidence interval is (123.6372, 126.9128).
3. For 99% confidence level confidence interval is (122.6428, 127.9072).

# LAB 4: Testing of Hypothesis

The following values are the lengths of 40 steel rods selected for lab test from a factory.

Length: 125, 120, 121, 123, 122, 130, 124, 122, 120, 122, 118, 119,

123, 124, 122, 124, 121, 122, 138, 149, 123, 128, 122, 130, 120, 122,

124, 134, 137, 128, 122, 121, 125, 120, 132, 130, 128, 130, 122, 124.

Test whether this sample of size 40 has come from a population whose mean length is 125 cm.

**Basic Steps in SPSS**

1. Enter the data in the data editor.

2. Select Analyze ->Compare Means-> One sample T test. Type in Test Value box.

3. Click Options-> Type 95 in confidence interval percentages box.

4. Click on Continue and then Ok.

**Working Expression:**

We wish to test the hypothesis that the samples differ significantly from a hypothesized population mean height of 125 cm. So, we have

**Step 1: Null Hypothesis($H_0$):** $\mu = 125$

i.e. There is no difference between sample mean and population mean.

**Step 2: Alternate Hypothesis($H_1$)**: $\mu \neq 125$

i.e. There is significant difference between sample mean and population mean.

**Step3: Test statistics**

Under $H_0$ Test statistics is given by,

p- value of test statistics (sig. (2-tailed)) = 0.00, compared to $\alpha$= 0.05.

**Step4: Decision and Conclusion**

P< $\alpha$, accept $H_1$ and reject $H_0$.

**Conclusion:** There is significant difference between sample mean and population mean.

Sarfraj Alam

**Output:**

One-Sample Statistics

| | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|
| Length | 40 | 125.2750 | 6.14770 | .97204 |

One-Sample Test

| | Test Value = 0 | | | | | |
|---|---|---|---|---|---|---|
| | | | | | 95% Confidence Interval of the Difference | |
| | t | df | Sig. (2-tailed) | Mean Difference | Lower | Upper |
| Length | 128.879 | 39 | .000 | 125.27500 | 123.3089 | 127.2411 |

**Interpretation:** There is significant difference between sample mean and population mean.

## Lab 5: Hypothesis Testing between two population Means for Matched Paired Samples

The sales of a product of a company after and before advertisement are as follows: Is advertisement effective at 5 % ?

| Month | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Before X | 120 | 140 | 160 | 140 | 180 | 190 |
| After Y | 200 | 210 | 150 | 200 | 220 | 240 |

**Basic steps in SPSS**

1.Enter the data into Data Editor

2. Select Analyze-> Compare Means ->Paired- Samples T test.

3. Click Options-> Continue-> Ok.

**Working expression:**

**Step 1: Null Hypothesis($H_0$)**: $\mu x = \mu y$

I.e There is no difference between before and after.

**Step 2: Alternate Hypothesis($H_1$)**: $\mu x < \mu y$

I.e There is significant difference between before and after.

**Step3: Test statistics**

Under $H_0$ Test statistics is given by,

Tcal=0.465

p- value of test statistics (sig. (2-tailed)) = 0.00, compared to $\alpha$= 0.05.

**Step4: Decision and Conclusion**

P< $\alpha$, accept $H_1$ and reject $H_0$.

**Conclusion:** There is significant difference between sample mean and population mean.

**Output:**

Paired Samples Statistics

| | | Mean | N | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| Pair 1 | BeforeX | 155.0000 | 6 | 26.64583 | 10.87811 |
| | AfterY | 203.3333 | 6 | 30.11091 | 12.29273 |

Sarfraj Alam

Paired Samples Correlations

| | | N | Correlation | Sig. |
|---|---|---|---|---|
| Pair 1 | BeforeX & AfterY | 6 | .374 | .465 |

Paired Samples Test

| | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | 95% Confidence Interval of the Difference | | | | |
| | | Mean | Std. Deviation | Std. Error Mean | Lower | Upper | | | |
| Pair 1 | BeforeX - AfterY | -48.33333 | 31.88521 | 13.01708 | -81.79481 | -14.87186 | -3.713 | 5 | .014 |

**Interpretation:** There is significant difference between sample mean and population mean.

## Lab 6: Hypothesis Testing When raw data for Independent Samples is given

The monthly advertising cost of a company for two products X and Y were as follows during 6-month period:

Is there sufficient evidence to conclude that average cost on advertising on product Y is more than on product X.

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Cost I (X) | 220 | 240 | 160 | 240 | 280 | 290 | - |
| Cost II (Y) | 100 | 110 | 150 | 100 | 120 | 140 | 145 |

**Basic Steps for SPSS**

1. Enter the data into Data editor.

2. Select Analyze ->Compare means ->Independent samples T Test

3. Move value into Test variable(s) and type into grouping variable

4. Click Define groups and type 1 and 2 into group 2

5. Click Options-> Continue ->Ok.

**Working Expression:**

**Step 1: Null Hypothesis($H_0$):** $\mu 1 = \mu 2$

i.e. There is no significant difference between population mean of two group.

**Step 2: Alternate Hypothesis($H_1$):** $\mu 1 < \mu 2$

i.e. There is significant difference between population mean of two group.

**Step3: Test statistics**

Under $H_0$ Test statistics is given by,

Tcal=5.862

**Step4: Critical Value**

For $\alpha$=5% level of significance and (n1+n2-2) degree of freedom

Ttab= 1.796

**Step4: Decision and Conclusion**

Ttab<Tcal, accept $H_1$ and reject $H_0$.

**Conclusion:** There is significant difference between population mean of two groups.

Sarfraj Alam

**Output:**

Group Statistics

| | group | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| cost | 1.00 | 6 | 238.3333 | 46.65476 | 19.04673 |
| | 2.00 | 7 | 123.5714 | 21.35304 | 8.07069 |

Independent Samples Test

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 95% Confidence Interval of the Difference | |
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | Lower | Upper |
| cost | Equal variances assumed | 1.357 | .269 | 5.862 | 11 | .000 | 114.76190 | 19.57600 | 71.67541 | 157.84840 |
| | Equal variances not assumed | | | 5.548 | 6.775 | .001 | 114.76190 | 20.68608 | 65.51535 | 164.00846 |

Independent Samples Effect Sizes

| | | Standardizer[a] | Point Estimate | 95% Confidence Interval | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| cost | Cohen's d | 35.18658 | 3.262 | 1.502 | 4.964 |
| | Hedges' correction | 37.83666 | 3.033 | 1.396 | 4.617 |
| | Glass's delta | 21.35304 | 5.374 | 2.213 | 8.512 |

a. The denominator used in estimating the effect sizes.

Cohen's d uses the pooled standard deviation.

Hedges' correction uses the pooled standard deviation, plus a correction factor.

Glass's delta uses the sample standard deviation of the control group.

**Interpretation:** There is significant difference between population mean of two groups.

## Lab 7: 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.

**Basic Steps for SPSS:**

1.Enter the data in data editor window.

2. Analyze/ Nonparametric test / Legacy Dialogs / Runs

3. Click Options Select Descriptive and continue

4. Click OK

**Working Expression:**

**Step 1: Null Hypothesis($H_0$):**

I.e. Sample observation is random.

**Step 2: Alternate Hypothesis($H_1$):** $\mu 1 < \mu 2$

I.e. Sample observation is not random.

**Step3: Test statistics:**

Under $H_0$ Test statistics is given by,

Number of runs(R)=22

Number of head(n1)=14

Number of tails(n2)=16

**Step 4: Critical value**

For $\alpha$= 0.05 level of significance, n1=16 and n2=14 tabulated value is

Rtab=(10,22)

**Step4: Decision and Conclusion**

R=22$\in$ (10,22), accept H0 and reject H1.

**Conclusion:** Sample observation is random.

**Output:**

Descriptive Statistics

| | N | Mean | Std. Deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| Runs | 30 | 10.9333 | 6.32964 | 1.00 | 22.00 |

Runs Test

| | Runs |
|---|---|
| Test Value[a] | 10.50 |
| Cases < Test Value | 15 |
| Cases >= Test Value | 15 |
| Total Cases | 30 |
| Number of Runs | 2 |
| Z | -5.017 |
| Asymp. Sig. (2-tailed) | .000 |

**Interpretation:** Sample observation is random.

Sarfraj Alam

# Lab 8: Binomial Test

Test whether the coin is unbiased from following observations.

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

**Basic Steps in SPSS**

1.Enter the data in data editor window.

2. Analyze/ Nonparametric test / Legacy Dialogs / Binomial test

3. Click Options Select Descriptive and continue

4. Click OK

**Working Expression:**

**Step 1: Null Hypothesis($H_0$):** P=1/2

I.e. Head and Tails are equally likely.

**Step 2: Alternate Hypothesis($H_1$):** $P \neq 1/2$

I.e. Head and Tails are not equally likely.

**Step3: Test statistics**

Under $H_0$ Test statistics is given by,

Number of Toss(n)=50

Number of head(n1)=20

Number of tails(n2)=30

**Step 4: Critical value**

Pvalue=0.203

Step4: Decision and Conclusion

Since Pvalue = 0.203>0.05, accept H0 and reject H1.

**Conclusion:** Heads and tails are equally likely.

**Output:**

Binomial Test

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

**Interpretation:** Heads and tails are equally likely.

## Lab 9: One Sample K-S 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 Kolmogorov 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 |

**Basic Steps in SPSS**

1. Start the SPSS program. In the Data editor window, type in the data.

2. Select Analyze-> Nonparametric tests-> Legacy Dialogs-> 1-Samples K-S.

3. Move X into Test Variable List. Then, Click OK.

**Working Expression**:

**Step 1: Null Hypothesis($H_0$)**: $F_e(x)=F_0(x)$

I.e. disease infected plants are uniformly distributed over the entire area.

**Step 2: Alternate Hypothesis($H_1$):** $F_e(x) \neq F_0(x)$

I.e. disease infected plants are not uniformly distributed over the entire area.

**Step3: Test statistics:**

Under $H_0$ Test statistics is given by,

Number of Toss(n)=50

Number of head(n1)=20

Number of tails(n2)=30

**Step 4: Critical value**

Pvalue=0.200

Step4: Decision and Conclusion

Since Pvalue = 0.200>0.05, accept H0 and reject H1.

**Conclusion:** Disease infected plants are uniformly distributed over the entire area.

Sarfraj Alam

**Output:**

One-Sample Kolmogorov-Smirnov Test

| | | | Noofinfectedplants |
|---|---|---|---|
| N | | | 10 |
| Normal Parameters[a,b] | Mean | | 10.0000 |
| | Std. Deviation | | 3.01846 |
| Most Extreme Differences | Absolute | | .146 |
| | Positive | | .130 |
| | Negative | | -.146 |
| Test Statistic | | | .146 |
| Asymp. Sig. (2-tailed)[c] | | | .200[d] |
| Monte Carlo Sig. (2-tailed)[e] | Sig. | | .786 |
| | 99% Confidence Interval | Lower Bound | .775 |
| | | Upper Bound | .796 |

a. Test distribution is Normal.

**Interpretation:** Disease infected plants are uniformly distributed over the entire area.

Sarfraj Alam

# Lab 10: 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. Use α = 0.01

Male 35 43 26 44 40 42 33 38 25 26

Female: 30 41 34 31 36 32 25 47 28 24

**Basic Steps in SPSS:**

1. Go to **Analyze → Nonparametric Tests → Legacy Dialogs → 2 Independent Samples**.
2. In the **Two-Independent-Samples Tests** dialog box:
   - Move **Age** to the **Test Variable List**.
   - Move **Gender** to the **Grouping Variable** box.
   - Click **Define Groups** and enter:
     - Group 1: 1 (Male)
     - Group 2: 2 (Female)
   - Ensure **Mann-Whitney U** is checked.
3. Click **OK**.

**Working Expression**:

**Step 1: Null Hypothesis($H_0$)**:

I.e. There is no difference in age distributions between male and female employees.

**Step 2: Alternate Hypothesis($H_1$):**

I.e. There is a difference in age distributions between male and female employees.

**Step3: Test statistics:**

Under $H_0$ Test statistics is given by,

Mann-Whitney U = 42

Wilcoxon W = 97

Z = -0.605

Asymp. Sig. (2-tailed) = 0.545

**Step 4: Critical value**

For α = 0.01 and n1=10 and n2=10

Pvalue=0.545

**Step 5: Decision and Conclusion**

Since Pvalue = 0.545>0.01, accept H0 and reject H1.

**Conclusion:** There is no difference in age distributions between male and female employees.

**Output:**

**Ranks**

| | gender | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| age | male | 10 | 11.30 | 113.00 |
| | female | 10 | 9.70 | 97.00 |
| | Total | 20 | | |

**Test Statistics[a]**

| | age |
|---|---|
| Mann-Whitney U | 42.000 |
| Wilcoxon W | 97.000 |
| Z | -.605 |
| Asymp. Sig. (2-tailed) | .545 |
| Exact Sig. [2*(1-tailed Sig.)] | .579[b] |

a. Grouping Variable: gender

b. Not corrected for ties.

**Interpretation:** There is no difference in age distributions between male and female employees.

Sarfraj Alam

# Lab 11: Mann - Whitney U Test

The following are the scores which random samples of students from 2 minority groups obtained on a current event test:

Groups I :73 82 39 68 91 75 89 67 50 86 57 65

Groups II: 51 42 36 53 88 59 49 66 25 64 18 76

Use Mann - Whitney U test at the 0.05 level of sig. to test whether or not students from the two minority groups can be expected to score equally well on the test.

**Basic Steps in SPSS:**

1. Go to **Analyze → Nonparametric Tests → Legacy Dialogs → 2 Independent Samples**.
2. In the **Two-Independent-Samples Tests** dialog box:
   - Move **Score** to the **Test Variable List**.
   - Move **Group** to the **Grouping Variable** box.
   - Click **Define Groups** and enter:
     - Group 1: 1 (group1)
     - Group 2: 2 (group2)
   - Ensure **Mann-Whitney U** is checked.
3. Click **OK**.

**Working Expression**:

**Step 1: Null Hypothesis($H_0$)**:

I.e. There is no difference in the distribution of test scores between Group I and Group II.

**Step 2: Alternate Hypothesis($H_1$):**

I.e. There is difference in the distribution of test scores between Group I and Group II.

**Step3: Test statistics:**

Under $H_0$ Test statistics is given by,

Mann-Whitney U = 34

Wilcoxon W = 112

Z = -2.194

Asymp. Sig. (2-tailed) = 0.28

**Step 4: Critical value**

For α = 0.01 and n1=12 and n2=12

Pvalue=0.28

**Step 5: Decision and Conclusion**

Since Pvalue = 0.28>0.01, accept H0 and reject H1.

Sarfraj Alam

**Conclusion:** There is no difference in the distribution of test scores between Group I and Group II.

**Output:**

**Descriptive Statistics**

| | N | Mean | Std. Deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| group | 24 | 61.2083 | 20.04556 | 18.00 | 91.00 |
| score | 24 | 1.5000 | .51075 | 1.00 | 2.00 |

**Mann-Whitney Test**

**Ranks**

| | score | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| group | group1 | 12 | 15.67 | 188.00 |
| | group2 | 12 | 9.33 | 112.00 |
| | Total | 24 | | |

**Test Statistics**[a]

| | group |
|---|---|
| Mann-Whitney U | 34.000 |
| Wilcoxon W | 112.000 |
| Z | -2.194 |
| Asymp. Sig. (2-tailed) | .028 |
| Exact Sig. [2*(1-tailed Sig.)] | .028[b] |

a. Grouping Variable: score

b. Not corrected for ties.

**Interpretation:** There is no difference in the distribution of test scores between Group I and Group II.

Sarfraj Alam

## Lab 12: Friedman 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:

| Hospital | No Of Births | | | |
|---|---|---|---|---|
| | 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.

**Basic Steps in SPSS:**

1. Start the SPSS program. In the Data Editor window, type in the data.(Hospital, ...... and values {1, A......} in hospital)
2. Select Analyze-> Nonparametric tests ->Legacy Dialogs ->K Related sample.
3. Click test variable and select test type.
4. Test type Friedman and Click OK.

**Working Expression**:

**Step 1: Null Hypothesis($H_0$)**:

I.e. There is no difference in the number of births across seasons (distributions are the same).

**Step 2: Alternate Hypothesis($H_1$):**

I.e. There is a difference in the number of births across at least two seasons.

**Step3: Test statistics:**

Under $H_0$ Test statistics is given by,

| | |
|---|---|
| N | 4 |
| Chi-Square | .900 |
| df | 3 |
| Asymp. Sig. | .825 |

**Step 4: Critical value**

For α = 0.05 ,N = number of blocks (hospitals) = 4, k = number of seasons = 4 p value is given by

Pvalue=0.825

**Step 5: Decision and Conclusion**

Since Pvalue = 0.825>0.05, accept H0 and reject H1.

**Conclusion:** There is no difference in the number of births across seasons.

# Output:

**Ranks**

| | Mean Rank |
|---|---|
| Winter | 2.25 |
| Spring | 3.00 |
| Summer | 2.25 |
| Fall | 2.50 |

**Test Statistics[a]**

| | |
|---|---|
| N | 4 |
| Chi-Square | .900 |
| df | 3 |
| Asymp. Sig. | .825 |

a. Friedman Test

Sarfraj Alam

**Nonparametric Correlations**

## Correlations

| | | | x | y |
|---|---|---|---|---|
| Kendall's tau_b | x | Correlation Coefficient | 1.000 | -.810* |
| | | Sig. (2-tailed) | . | .011 |
| | | N | 7 | 7 |
| | y | Correlation Coefficient | -.810* | 1.000 |
| | | Sig. (2-tailed) | .011 | . |
| | | N | 7 | 7 |
| Spearman's rho | x | Correlation Coefficient | 1.000 | -.929** |
| | | Sig. (2-tailed) | . | .003 |
| | | N | 7 | 7 |
| | y | Correlation Coefficient | -.929** | 1.000 |
| | | Sig. (2-tailed) | .003 | . |
| | | N | 7 | 7 |

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

**Interpretation:** There is no difference in the number of births across seasons

## Lab 13: Median Test

**a.** Large Samples size

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

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

**Basic Steps in SPSS:**

1. Start the SPSS program. In the Data Editor window, type in the data.(in scores and Gender)
2. Select Analyze-> Nonparametric tests ->Legacy Dialogs ->K -independent sample.
3. Defined Groups {Gender (1,2)}.
4. Click OK.

**Working Expression**:

**Step 1: Null Hypothesis($H_0$)**:

I.e. The median IQ of male and female students is the same.

**Step 2: Alternate Hypothesis($H_1$):**

I.e. The median IQ of male and female students is not the same.

**Step3: Test statistics:**

Under $H_0$ Test statistics is given by,

**Test Statistics[a]**

| | | Iqscore |
|---|---|---|
| N | | 34 |
| Median | | 67.5000 |
| Chi-Square | | 1.074 |
| df | | 1 |
| Asymp. Sig. | | .300 |
| Yates' Continuity Correction | Chi-Square | .477 |
| | df | 1 |
| | Asymp. Sig. | .490 |

a. Grouping Variable: Gender

**Step 4: Critical value**

For α = 0.05 ,N = number of blocks (hospitals) = 4, k = number of seasons = 4 p value is given by

Pvalue=0.825

**Step 5: Decision and Conclusion**

Since Pvalue = 0.825>0.05, accept H0 and reject H1.

**Conclusion:** There is no difference in the number of births across seasons.

**Descriptive Statistics**

| | N | Mean | Std. Deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| Iqscore | 34 | 66.7353 | 11.22041 | 44.00 | 89.00 |
| Gender | 34 | 1.5588 | .50399 | 1.00 | 2.00 |

**Frequencies**

| | | Gender | |
|---|---|---|---|
| | | Male | Female |
| Iqscore | > Median | 9 | 8 |
| | <= Median | 6 | 11 |

**Test Statistics[a]**

| | | Iqscore |
|---|---|---|
| N | | 34 |
| Median | | 67.5000 |
| Chi-Square | | 1.074 |
| df | | 1 |
| Asymp. Sig. | | .300 |
| Yates' Continuity Correction | Chi-Square | .477 |
| | df | 1 |
| | Asymp. Sig. | .490 |

a. Grouping Variable: Gender

Sarfraj Alam

**Correlations**

| | | x | y |
|---|---|---|---|
| x | Pearson Correlation | 1 | -.929** |
| | Sig. (2-tailed) | | .003 |
| | N | 7 | 7 |
| y | Pearson Correlation | -.929** | 1 |
| | Sig. (2-tailed) | .003 | |
| | N | 7 | 7 |

**. Correlation is significant at the 0.01 level (2-tailed).

**Interpretation:** There is no difference in the number of births across seasons.

## Lab 14: Simple Regression

Enter the following values in SPSS and find the regression equation of y on x:

| X | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Y | 6 | 7 | 5 | 4 | 3 | 1 | 2 |

**Basic Steps in SPSS:**

1. Enter X and Y values into SPSS.
2. Go to **Analyze** → **Regression** → **Linear**.
3. Move **X** to Independent(s) and **Y** to Dependent(s).
4. Click **OK**.

Output:

**Coefficients**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | | |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | t | Sig. |
| 1 | (Constant) | 7.714 | .742 | | 10.392 | .000 |
| | X | -.929 | .166 | -.929 | -5.594 | .003 |

a. Dependent Variable: Y

**Conclusion:**

The computed regression equation is:

$Y = 7.714 - 0.929X$

This means:

- **Intercept ($b_0$)** = 7.714
- **Slope ($b_1$)** = -0.929

This regression line represents the best linear fit for predicting Y based on X using least squares estimation.

## Lab 15: Simple Correlation

Enter the following values in SPSS and find the correlation between X and Y:

| X | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Y | 6 | 7 | 5 | 4 | 3 | 1 | 2 |

**Basic Steps in SPSS:**

1. Enter the values for **X** and **Y**.
2. Go to **Analyze → Correlate → Bivariate**.
3. Move both **X** and **Y** into the Variables box.
4. Select the desired correlation method (e.g., **Pearson**, **Kendall's tau-b**, **Spearman**).
5. Click **OK** to view the correlation result.

**Output:**

**Correlations**

| | | X | Y |
|---|---|---|---|
| X | Pearson Correlation | 1 | -.929** |
| | Sig. (2-tailed) | | .003 |
| | N | 7 | 7 |
| Y | Pearson Correlation | -.929** | 1 |
| | Sig. (2-tailed) | .003 | |
| | N | 7 | 7 |

**. Correlation is significant at the 0.01 level (2-tailed).

```
NONPAR CORR
  /VARIABLES=X Y
  /PRINT=BOTH TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.
```

Sarfraj Alam

**Nonparametric Correlations**

**Correlations**

| | | | X | Y |
|---|---|---|---|---|
| Kendall's tau_b | X | Correlation Coefficient | 1.000 | -.810* |
| | | Sig. (2-tailed) | . | .011 |
| | | N | 7 | 7 |
| | Y | Correlation Coefficient | -.810* | 1.000 |
| | | Sig. (2-tailed) | .011 | . |
| | | N | 7 | 7 |
| Spearman's rho | X | Correlation Coefficient | 1.000 | -.929** |
| | | Sig. (2-tailed) | . | .003 |
| | | N | 7 | 7 |
| | Y | Correlation Coefficient | -.929** | 1.000 |
| | | Sig. (2-tailed) | .003 | . |
| | | N | 7 | 7 |

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

### Interpretation:

The Pearson correlation measures the linear relationship between two variables.

Using the formula or software (e.g., Excel, SPSS, Python), we get:

$r \approx -0.929$

Interpretation of Pearson Result:

- Value: -0.929 (Very strong negative correlation)
- Meaning: As X increases, Y strongly decreases in a linear fashion.
- Strength: The correlation is very strong because the value is close to -1.
- Direction: It's negative, meaning there is an inverse relationship.

Spearman & Kendall (for ranked data)

If you're dealing with ordinal or non-normally distributed data, use:

- Spearman's rho: Also shows a strong negative correlation (~ -0.929).
- Kendall's tau-b: Slightly weaker in magnitude but still strongly negative (~ -0.81).

**Conclusion:**

There is a **very strong negative correlation** between variables **X** and **Y**. As one increases, the other tends to decrease in a near-perfect linear pattern. This is evident both visually and statistically.