# ACKNOWLEDGEMENT

I am deeply grateful to Mr. Rajan Poudel for his invaluable guidance and support, which were essential in the completion of this lab report. His expertise in probability and statistics significantly improved my understanding of the subject.

I also extend my thanks to Itahari Namuna College for providing the necessary resources and facilities for conducting the experiments and analyses presented in this report.

Special thanks go to my classmates and colleagues for their invaluable feedback and support throughout this project. Their insights and suggestions greatly enhanced the analyses and interpretations.

Lastly, I sincerely appreciate my family and friends for their unwavering support and encouragement during this study.

Thank you.

# TABLE OF CONTENT

# EXPERIMENT NO-1

The following table gives the yearly income (000Rs) of the 10 computer programmers working in a software company.

Income (000Rs.) 1780, 1760, 1680, 1780, 1830, 1940, 1100, 1800, 1060, 1950.

Find the mean, median and mode of the given individual data.

## 1.1. Objective: To find mean, median and mode of individual series using SPSS.

## 1.2. Working Expression:

**Mean**: The mean, or arithmetic mean, is a widely utilized measure of central tendency. It is often regarded as the most effective measure due to its ability to meet most criteria of a good measure. Mathematically, the mean is calculated by summing all observations and then dividing by the total number of observations.

Mean ($\bar{X}$) = $\frac{\sum X}{N}$

**Median**: Median is a widely used as measure of center tendency. This measure is appropriate to calculate the average value of the given distribution having open ended classes, unequal class size, and for the qualitative data. Mathematically, median is the value which divided the given frequency distribution into two equal parts.

Median ($M_d$) = value of $(\frac{n+1}{2})^{th}$ item.

**Mode**: Mode is a commonly used measure of central tendency. This measure is suitable when we have to find the most common value, most usual value, most frequent value, most repeated value, ideal size, favorite number, etc. of given data. Mathematically, mode is a value which repeats maximum number of time in a frequency distribution.

Mode ($M_o$) = most repeated value

## 1.3. Calculation:

**Statistics**

X

| | | |
|---|---|---|
| N | Valid | 10 |
| | Missing | 0 |
| Mean | | 1668.00 |
| Median | | 1780.00 |
| Mode | | 1780 |

**X**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 1060 | 1 | 10.0 | 10.0 | 10.0 |
| | 1100 | 1 | 10.0 | 10.0 | 20.0 |
| | 1680 | 1 | 10.0 | 10.0 | 30.0 |
| | 1760 | 1 | 10.0 | 10.0 | 40.0 |
| | 1780 | 2 | 20.0 | 20.0 | 60.0 |
| | 1800 | 1 | 10.0 | 10.0 | 70.0 |
| | 1830 | 1 | 10.0 | 10.0 | 80.0 |
| | 1940 | 1 | 10.0 | 10.0 | 90.0 |
| | 1950 | 1 | 10.0 | 10.0 | 100.0 |
| | Total | 10 | 100.0 | 100.0 | |

## 1.4. Conclusion:

The mean of the given distribution is 1668.

The median of the given distribution is 1780.

The mode of the given distribution is 1780.

# EXPERIMENT NO-2

From the data given below calculate the value of $Q_1$, $Q_3$, $D_2$, $D_9$, $P_{45}$ and $P_{57}$.

| Quantity | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 |
|---|---|---|---|---|---|---|---|---|---|
| No. of Person | 8 | 12 | 36 | 25 | 28 | 18 | 9 | 12 | 6 |

## 2.1. Objective: To find quartile, decile and percentile of discrete series using SPSS.

## 2.2. Working Expression:

**Quartiles**: Quartiles are values that divide a given frequency distribution into four equal parts. There are three quartiles in a frequency distribution: the first quartile or lower quartile (Q1), the second quartile or median (Md), and the third quartile or upper quartile (Q3).

Quartile($Q_i$) = size of $[\frac{i(n+1)}{4}]^{th}$ item.

**Deciles**: Deciles are values that divide a given frequency distribution into ten equal parts. There are nine deciles, denoted as D1, D2, D3, ..., D9, where D1 < D2 < D3 < ... < D9.

Deciles($D_i$) = size of $[\frac{i(n+1)}{10}]^{th}$ item.

**Percentiles**: Percentiles are values that divide a given frequency distribution into one hundred equal parts. There are ninety-nine percentiles, denoted as P1, P2, P3, ..., P99, where P1 < P2 < P3 < ... < P99.

Percentiles($P_i$) = size of $[\frac{i(n+1)}{100}]^{th}$ item.

## 2.3. Calculation:

**Statistics**

quantity

| | | |
|---|---|---|
| N | Valid | 154 |
| | Missing | 0 |
| Percentiles | 20 | 20.00 |
| | 25 | 20.00 |
| | 45 | 25.00 |
| | 50 | 25.00 |
| | 57 | 30.00 |
| | 75 | 35.00 |
| | 90 | 45.00 |

**quantity**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 10 | 8 | 5.2 | 5.2 | 5.2 |
| | 15 | 12 | 7.8 | 7.8 | 13.0 |
| | 20 | 36 | 23.4 | 23.4 | 36.4 |
| | 25 | 25 | 16.2 | 16.2 | 52.6 |
| | 30 | 28 | 18.2 | 18.2 | 70.8 |
| | 35 | 18 | 11.7 | 11.7 | 82.5 |
| | 40 | 9 | 5.8 | 5.8 | 88.3 |
| | 45 | 12 | 7.8 | 7.8 | 96.1 |
| | 50 | 6 | 3.9 | 3.9 | 100.0 |
| | Total | 154 | 100.0 | 100.0 | |

## 2.4. Conclusion:

The value of $Q_1$ of given distribution is 20.

The value of $Q_3$ of given distribution is 35.

The value of $D_2$ of given distribution is 20.

The value of $D_9$ of given distribution is 45.

The value of $P_{45}$ of given distribution is 25.

The value of $P_{57}$ of given distribution is 30.

# EXPERIMENT NO-3

From the data given below, calculate range, variance and standard deviation.

| Marks | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 |
|---|---|---|---|---|---|
| No. of Students | 5 | 18 | 15 | 16 | 6 |

**3.1. Objective:** To find range, variance and standard deviation of continuous data using SPSS.

**3.2. Working Expression:** Range: It is a measure of dispersion defined as the difference between largest and smallest values in a given distribution.

Range = L – S

Variance: The variance of the values of a variable X is defined as the square of the standard deviation.

Var(X) = $\sigma^2$

Standard Deviation: It is the square root of mean of the squared deviations from the arithmetic mean.

$$(continuous\ series)\sigma = \sqrt{\frac{1}{N}\sum f\,(X-\bar{X})^2}$$

## 3.3. Calculation:

**Statistics**

M

| | | |
|---|---|---|
| N | Valid | 60 |
| | Missing | 0 |
| Std. Deviation | | 11.49797 |
| Variance | | 132.203 |
| Range | | 40.00 |

**M**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 5.00 | 5 | 8.3 | 8.3 | 8.3 |
| | 15.00 | 18 | 30.0 | 30.0 | 38.3 |
| | 25.00 | 15 | 25.0 | 25.0 | 63.3 |
| | 35.00 | 16 | 26.7 | 26.7 | 90.0 |
| | 45.00 | 6 | 10.0 | 10.0 | 100.0 |
| | Total | 60 | 100.0 | 100.0 | |

## 3.4. Conclusion:

The range of given distribution is 40.

The variance of given distribution is 132.203.

The standard deviation of given distribution is 11.498.

# EXPERIMENT NO-4

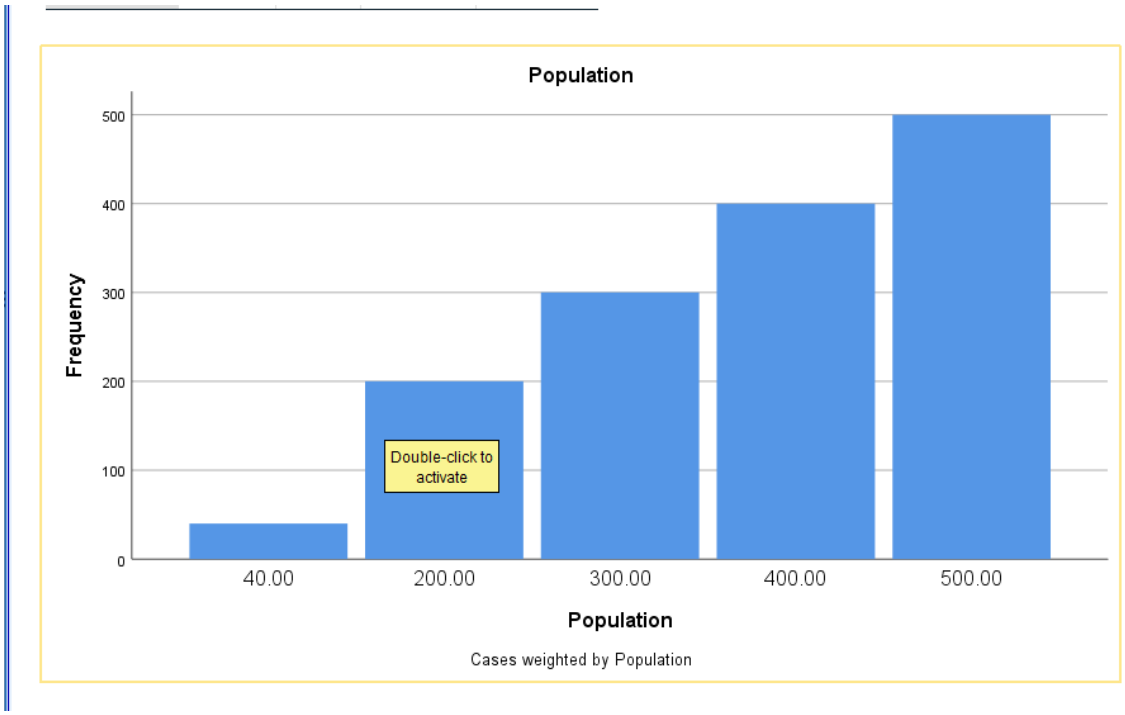
Draw a line graph using spss from the given below information.

| Village | A | | B | C | D | E |
|---|---|---|---|---|---|---|
| Population | 300 | | 400 | 500 | 40 | 200 |

**4.1. Objective:** To draw bar graph using spss.

**4.2. Working Expression:** A bar graph is a chart that uses rectangular bars to depict data, with the height of each bar corresponding to the value of the data point it represents. Bar graphs are frequently employed to compare data across various categories. They are useful for representing categorical data, such as the number of individuals in different age groups, or ordinal data, like the education levels within a group. Additionally, bar graphs can illustrate continuous data, such as the average income of people in different cities.

**4.3. Calculation:**



## Frequencies

[DataSet1]

**Statistics**

Population

| | | |
|---|---|---|
| N | Valid | 1440 |
| | Missing | 0 |

**Population**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 40.00 | 40 | 2.8 | 2.8 | 2.8 |
| | 200.00 | 200 | 13.9 | 13.9 | 16.7 |
| | 300.00 | 300 | 20.8 | 20.8 | 37.5 |
| | 400.00 | 400 | 27.8 | 27.8 | 65.3 |
| | 500.00 | 500 | 34.7 | 34.7 | 100.0 |
| | Total | 1440 | 100.0 | 100.0 | |

## 4.4. Conclusion:

We have successfully made the bar graph using spss.

# EXPERIMENT NO-5

Calculate coefficient of skewness and kurtosis from the marks of students given below.

| Marks | 59 | 61 | 63 | 65 | 67 | 69 | 71 | 73 | 75 |
|---|---|---|---|---|---|---|---|---|---|
| No. of Students | 0 | 2 | 8 | 20 | 40 | 20 | 8 | 2 | 0 |

**5.1. Objective:** To find coefficient of skewness and kurtosis of discrete data using SPSS.

**5.2. Working Expression:** Skewness: It gives us an idea about the asymmetry of the items about the measure of central tendency. If mean = median = mode, it is called symmetrical distribution while mean ≠ median ≠ mode, it is called a skewed distribution and can be positive or negative.

$S_{KB} = \frac{Q3+Q1-2Md}{Q3-Q1}$

Kurtosis: It is the measure of peaked ness or flatness of curve of given distribution.

Leptokurtic curve is more peaked than mesokurtic curve.(>0.263)

Mesokurtic curve is neither peaked nor flat.(0.263)

Platykurtic curve is more flat compared to mesokurtic curve.(<0.263)

$K = \frac{p75-p25}{2(P90-p10)}$

## 5.3. Calculation:

**Statistics**

marks

| | | |
|---|---|---|
| N | Valid | 100 |
| | Missing | 0 |
| Skewness | | .000 |
| Std. Error of Skewness | | .241 |
| Kurtosis | | .234 |
| Std. Error of Kurtosis | | .478 |

**marks**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 61 | 2 | 2.0 | 2.0 | 2.0 |
| | 63 | 8 | 8.0 | 8.0 | 10.0 |
| | 65 | 20 | 20.0 | 20.0 | 30.0 |
| | 67 | 40 | 40.0 | 40.0 | 70.0 |
| | 69 | 20 | 20.0 | 20.0 | 90.0 |
| | 71 | 8 | 8.0 | 8.0 | 98.0 |
| | 73 | 2 | 2.0 | 2.0 | 100.0 |
| | Total | 100 | 100.0 | 100.0 | |

## 5.4. Conclusion:

The skewness of given distribution is 0(symmetrical).

The kurtosis of given distribution is 0.234(platykurtic).

# EXPERIMENT NO-6

Calculate Karl Pearson's correlation coefficient and Spearman's rank correlation between sales and repairs of computer as given below.

| Sales | 50 | 55 | 55 | 60 | 65 | 70 | 65 | 60 |
|---|---|---|---|---|---|---|---|---|
| Repair | 11 | 13 | 14 | 16 | 16 | 15 | 15 | 20 |

**6.1. Objective:** To find Karl Pearson's correlation coefficient and Spearman's rank correlation using SPSS.

**6.2. Working Expression:** **Correlation**: Two variables are said to be in correlation, if change in one variable is accompanied by change in another variable.

Karl Pearson's correlation coefficient: Let X and Y be two variables then, it is denoted by r which is defined as ratio of Covariance between X and Y to multiple of SD of X and SD of y.

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{n\sum X^2 - (\sum X)^2}\sqrt{n\sum Y^2 - (\sum Y)^2}}$$

**Spearman's Rank Correlation**: Spearman's rank correlation coefficient, often denoted by the symbol $\rho$ (rho), is a non-parametric measure of the strength and direction of a monotonic relationship between two variables. Unlike Pearson's correlation, Spearman's rank correlation is based on the ranks of the data rather than the actual data values.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

## 6.3. Calculation:

**Correlations**

| | | Sales | Repair |
|---|---|---|---|
| Sales | Pearson Correlation | 1 | .500 |
| | Sig. (2-tailed) | | .207 |
| | N | 8 | 8 |
| Repair | Pearson Correlation | .500 | 1 |
| | Sig. (2-tailed) | .207 | |
| | N | 8 | 8 |

```
NONPAR CORR
  /VARIABLES=Sales Repair
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

**Nonparametric Correlations**

**Correlations**

| | | | Sales | Repair |
|---|---|---|---|---|
| Spearman's rho | Sales | Correlation Coefficient | 1.000 | .595 |
| | | Sig. (2-tailed) | . | .120 |
| | | N | 8 | 8 |
| | Repair | Correlation Coefficient | .595 | 1.000 |
| | | Sig. (2-tailed) | .120 | . |
| | | N | 8 | 8 |

## 6.4. Conclusion:

The Karl Pearson's Correlation Coefficient is 0.500(moderate positive).

The Spearman's Rank Correlation is 0.595(moderate positive).

## EXPERIMENT NO-7

Find the regression equation of X on Y from following data.

| | | | | | |
|---|---|---|---|---|---|
| X | 5 | 15 | 20 | 25 | 30 |
| Y | 50 | 60 | 80 | 110 | 130 |

**7.1. Objective:** To find regression equation of X on Y using SPSS.

**7.2. Working Expression:** Regression: It is a statistical tool used to determine how the variables are related and making estimate or prediction from that relationship. The unknown variable we are going to estimate is called dependent variable and the known variable is known as independent variable.

For regression equation of X on Y:

Equation be x = a + by

To estimate a and b using principle of least square by minimizing error sum of square:

$$\sum X = na + b\sum Y$$

$$\sum XY = a\sum Y + b\sum Y^2$$

Solving these two equations, get a and b and substitute in x = a + by

## 7.3. Calculation:

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .951[a] | .905 | .873 | 3.430 |

a. Predictors: (Constant), Y

**ANOVA[a]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 334.712 | 1 | 334.712 | 28.456 | .013[b] |
| | Residual | 35.288 | 3 | 11.763 | | |
| | Total | 370.000 | 4 | | | |

a. Dependent Variable: X

b. Predictors: (Constant), Y

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | | |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | t | Sig. |
| 1 | (Constant) | -4.403 | 4.648 | | -.947 | .413 |
| | Y | .272 | .051 | .951 | 5.334 | .013 |

a. Dependent Variable: X

## 7.4. Conclusion:

The regression equation of X on Y is x = -4.403 + 0.272y.

# EXPERIMENT NO-8

From the following data of marks obtained by 60 students in an exam of C programming, Calculate mean, median and mode.

| Marks | 20 | 30 | 40 | 50 | 60 | 70 |
|---|---|---|---|---|---|---|
| No. of Students | 8 | 12 | 20 | 10 | 6 | 4 |

**8.1. Objective:** To find mean, median and mode of discrete data using SPSS.

**8.2. Working Expression:** Mean: Mean is a commonly used measure of central tendency. The arithmetic mean is considered as the best measure of central tendency because it satisfied the most of all requisites of good measure. Mathematically, mean is defined as the sum of all the observation divided by total number of observation.

Mean $(\bar{X}) = \frac{\sum fX}{N}$

Median: Median is a widely used as measure of center tendency. This measure is appropriate to calculate the average value of the given distribution having open ended classes, unequal class size, and for the qualitative data. Mathematically, median is the value which divided the given frequency distribution into two equal parts.

Median ($M_d$) = Value corresponding to the cf greater than or equal to the size of md,

Size of md = $\frac{N+1}{2}$ th term

Mode: Mode is a commonly used measure of central tendency. This measure is suitable when we have to find the most common value, most usual value, most frequent value, most repeated value, ideal size, favorite number, etc. of given data. Mathematically, mode is a value which repeats maximum number of time in a frequency distribution.

Mode ($M_o$) = most repeated value

## 8.3. Calculation:

**Statistics**

marks

| | | |
|---|---|---|
| N | Valid | 60 |
| | Missing | 0 |
| Mean | | 41.00 |
| Median | | 40.00 |
| Mode | | 40 |

**marks**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 20 | 8 | 13.3 | 13.3 | 13.3 |
| | 30 | 12 | 20.0 | 20.0 | 33.3 |
| | 40 | 20 | 33.3 | 33.3 | 66.7 |
| | 50 | 10 | 16.7 | 16.7 | 83.3 |
| | 60 | 6 | 10.0 | 10.0 | 93.3 |
| | 70 | 4 | 6.7 | 6.7 | 100.0 |
| | Total | 60 | 100.0 | 100.0 | |

## 8.4. Conclusion:

The mean of given distribution is 41.

The median of given distribution is 40.

The mode of given distribution is 40.

## EXPERIMENT NO-9

From the following data represents waiting time of students in bus stand to be seated. Calculate mean, median and mode.

| Waiting time | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 |
|---|---|---|---|---|---|---|
| No. of Students | 5 | 10 | 25 | 30 | 20 | 10 |

**9.1. Objective:** To find mean, median and mode of continuous data using SPSS.

**9.2. Working Expression:** Mean: Mean is a commonly used measure of central tendency. The arithmetic mean is considered as the best measure of central tendency because it satisfied the most of all requisites of good measure. Mathematically, mean is defined as the sum of all the observation divided by total number of observation.

Mean $(\bar{X}) = \frac{\Sigma fm}{N}$

Median: Median is a widely used as measure of center tendency. This measure is appropriate to calculate the average value of the given distribution having open ended classes, unequal class size, and for the qualitative data. Mathematically, median is the value which divided the given frequency distribution into two equal parts.

Median $(M_d) = L + \frac{\frac{N}{2} - cf}{f} * h$

Mode: Mode is a commonly used measure of central tendency. This measure is suitable when we have to find the most common value, most usual value, most frequent value, most repeated value, ideal size, favorite number, etc. of given data. Mathematically, mode is a value which repeats maximum number of time in a frequency distribution.

Mode $(M_o) = L + \frac{f1-f0}{2f1-f0-f2} * h$

If mode is ill defined, mode = 3 Median – 2 Mean

## 9.3. Calculation:

**Statistics**

TimeM

| | | |
|---|---|---|
| N | Valid | 100 |
| | Missing | 0 |
| Mean | | 33.0000 |
| Median | | 35.0000 |
| Mode | | 35.00 |

**TimeM**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 5.00 | 5 | 5.0 | 5.0 | 5.0 |
| | 15.00 | 10 | 10.0 | 10.0 | 15.0 |
| | 25.00 | 25 | 25.0 | 25.0 | 40.0 |
| | 35.00 | 30 | 30.0 | 30.0 | 70.0 |
| | 45.00 | 20 | 20.0 | 20.0 | 90.0 |
| | 55.00 | 10 | 10.0 | 10.0 | 100.0 |
| | Total | 100 | 100.0 | 100.0 | |

## 9.4. Conclusion:

The mean of given distribution is 33.

The median of given distribution is 35.

The mode of given distribution is 35.

# EXPERIMENT NO. 10:

An experiment was conducted to determine the effects of different dates of planting and different methods of planting on the yield of sugarcane. Does the method of planting affect mean yield and date of planting affect mean yield?

| Method of Planting | Date of Planting | | | | |
|---|---|---|---|---|---|
| | | Oct | Nov | Feb | Mar |
| | I | 7 | 4 | 5 | 2 |
| | II | 10 | 5 | 5 | 3 |
| | III | 8 | 4 | 5 | 2 |

**10.1. Objective:** To carry out two way analysis of variance (ANOVA) using SPSS.

**10.2. Working Expression:** ANOVA is the systematic process for achieving the variation. It is used for test of significance.

**Types of ANOVA:**

- One-way ANOVA
- Two-way ANOVA
- Multi-way ANOVA

The total variation present in any set of numerical data is classified according to two factors is called two-way classification.

**Setting Of Hypothesis:**

$H_{0R}$: $\mu_{I.} = \mu_{II.} = \mu_{III.}$

$H_{1R}$: At least one $\mu_i$ is different, I = I, II, III

$H_{0C}$: $\mu_{.1} = \mu_{.2} = \mu_{.3}$

$H_{1C}$: At least one $\mu_{.j}$ is different, j = 1(Oct), 2(Nov), 3(Feb), 4(Mar)

**LEVEL OF SIGNIFICANCE:**

$\alpha = 5\%$

**TEST STATISTICS:**

$F_R = \frac{MSR}{MSE}$ $F_C = \frac{MSC}{MSE}$

Where,

MSR = SSR/m-1

MSC = SSC/n-1

MSE = SSE/(m-1)(n-1)

## 10.3. Calculation:

**Tests of Between-Subjects Effects**

Dependent Variable: Value

| Source | Type II Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared |
|---|---|---|---|---|---|---|
| Corrected Model | 59.500[a] | 5 | 11.900 | 28.560 | .000 | .960 |
| Intercept | 300.000 | 1 | 300.000 | 720.000 | .000 | .992 |
| DateC | 56.000 | 3 | 18.667 | 44.800 | .000 | .957 |
| MethodR | 3.500 | 2 | 1.750 | 4.200 | .072 | .583 |
| Error | 2.500 | 6 | .417 | | | |
| Total | 362.000 | 12 | | | | |
| Corrected Total | 62.000 | 11 | | | | |

a. R Squared = .960 (Adjusted R Squared = .926)

**Critical Value:**

Degree of freedom (Row) = {m-1,(m-1)(n-1} = (2,6)

Degree of freedom (Column) = {n-1,(m-1)(n-1} = (3,6)

$F_{0.05\,(2,6)} = 5.14$

$F_{0.05\,(3,6)} = 4.76$

**Decision:**

$F_R = 4.200 < F_{0.05\,(2,6)} = 5.14$, accept $H_{0R}$ at 5% level of significance.

$F_C = 44.800 > F_{0.05\,(3,6)} = 4.76$, reject $H_{0C}$ at 5% level of significance.

## 10.4. Conclusion:

Methods of planting do not affect the mean yield of sugarcane but dates of planting affect the mean yields of sugarcane.