# SPSS CODING WITH PROBLEM

## SPSS Coding with Problems

**Problem1: Suppose we want to Compute the "Increment" and "Present Salary" of the employee from the following data,**

**Salary**

**10000 13000 15500**
15000 14000 16500
12000 17000 17500

- Increment = (10% of the salary)+1000
- Present Salary=Salary+Increment

**Command: 1.**

```
SAVE OUTFILE='C:\Users\ASUS\Documents\Mahdi Solve
1.sav'
  /COMPRESSED.

COMPUTE Increnebt=Salary * 0.10 + 1000.
EXECUTE.
```

**Command: 2.**

```
COMPUTE Present_Salary=Salary + Increnebt.
EXECUTE.
```

**Problem 2:** Suppose we want to find the deduction from the salary for the transport facility for the employees of a firm from the following data of salary. It is given that 5% of salary is deducted if salary is Gretter then 12000 taka.

**Command :**
**IF  (Salary > 12000) Deduct=Salary * 0.05.**
**EXECUTE.**

**Problem : 3. Suppose we want to Complete the Increment of the employee who at satisfies the following condition from the following data.**
**Condition: Increment= 15% of the salary if job category =3 and Experience is Gretter then or equal to 5 years.**

| Salary | Job category | Experience (Year) |
|---|---|---|
| 10000 | 1 | 5 |
| 15000 | 2 | 6 |
| 17000 | 3 | 7 |
| 21000 | 3 | 8 |
| 18000 | 3 | 5 |

**Command:**

```
IF(Job_Cat = 3 & Experience >= 5) Increment=Salary *
0.15.
EXECUTE.
```

**Problem 4: <u>Recode Into Different Variable</u>**

**Suppose we want to create "Social-Status" on the basis of Income Variable.**

| **<u>Income</u>** | | |
|---|---|---|
| **10000** | **13000** | **15500** |
| 15000 | 14000 | 16500 |
| 12000 | 17000 | 17500 |

| Income | New Value |
|---|---|
| Less or equal 10000 | 1 |
| 10001-12000 | 2 |
| 12001-15000 | 3 |
| 15001-17000 | 4 |
| 17001-17500 | 5 |

**Command:**

```
RECODE Salary (Lowest thru 10000=1) (10001 thru
12000=2) (12001 thru 15000=3) (15001 thru 17000=4)
(17001 thru 17500=5) INTO Social_Status.
EXECUTE.
```

**Problem 5: Suppose we want to define "Educational Status" On the basis of "Year of Schooling" from the following data.**

| Year of Schooling | New Value | Value Label |
|---|---|---|
| **0** | **1** | **Illiterate** |
| **1-5** | **2** | **Primary** |
| **6-10** | **3** | **Secondary** |
| **11-12** | **4** | **Higher Secondary** |
| **13-16** | **5** | **Graduate** |
| **17** | **6** | **Post Graduate** |
| **18** | **7** | **Higher** |

**<u>Year of Schooling</u>**

**15, 07, 14, 08, 18, 00, 18, 06, 17, 11, 10, 05, 12, 16.**

**Construct a Frequency Distribution Table Using Educational Status Variable.**

**Command:**

```
RECODE Year_Of_SChooling (17=6) (18=7) (Lowest thru
0=1) (1 thru 5=2) (6 thru 10=3) (11 thru 12=4) (13
thru 16=5) INTO Educational_Status.
EXECUTE.

FREQUENCIES VARIABLES=Educational_Status
  /ORDER=ANALYSIS.
```

## <u>Table For Frequency</u>

**Educational_Status**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | illeterate | 1 | 11.1 | 11.1 | 11.1 |

| | | | | |
|---|---|---|---|---|
| Secondary | 3 | 33.3 | 33.3 | 44.4 |
| Graduate | 2 | 22.2 | 22.2 | 66.7 |
| Post Graduate | 1 | 11.1 | 11.1 | 77.8 |
| Higher | 2 | 22.2 | 22.2 | 100.0 |
| Total | 9 | 100.0 | 100.0 | |

**Command For Graph**

**For Pie Chart:**

```
GRAPH
  /PIE=PCT BY Educational_Status.
```

**For Bar Chart**

```
GRAPH
  /BAR(GROUPED)=PCT BY Age BY Educational_Status.
```

**Problem 6: Consider the following Table**

| ID | Gender | Age | Diabeties | Smoking Status |
|---|---|---|---|---|
| 1 | Male | 65 | Yes | Yes |
| 2 | Female | 18 | No | No |
| 3 | Female | 46 | Yes | Yes |
| 4 | Male | 57 | Yes | yes |
| 5 | male | 26 | no | no |
| 6 | male | 45 | yes | yes |
| 7 | female | 29 | no | yes |
| 8 | male | 68 | yes | no |
| 9 | male | 39 | no | yes |
| 10 | female | 41 | No | no |

- **Construct a Simple Bar Diagram for “Diabetes”**
- **Construct a Clustered Bar Diagram for “Diabetes” and “Smoking Status”**
- **Draw a Stacked bar Diagram for “Diabetes” and “gender”**
- **Perform a bivariate Analysis for “Gender” and “Diabetes”**

**Command1:**

```
GRAPH
  /BAR(SIMPLE)=PCT BY Diabeties.
```

**Command2:**

```
GRAPH
  /BAR(GROUPED)=PCT BY Diabeties BY Smoking_Status.
```

**Command3:**

```
GRAPH
  /BAR(STACK)=PCT BY Diabeties BY Gender.
```

**Command4:**

```
CROSSTABS
  /TABLES=Gender BY Diabeties
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ
  /CELLS=COUNT TOTAL
  /COUNT ROUND CELL.
```

**Table For Crosstabs & Chi Square Test**

**Gender of The Respondent * Diabeties Status Of The Respondent Crosstabulation**

| | | | Diabeties Status Of The Respondent | | Total |
|---|---|---|---|---|---|
| | | | Yes | No | |
| Gender of The Respondent | Male | Count | 4 | 2 | 6 |
| | | % of Total | 40.0% | 20.0% | 60.0% |
| | Female | Count | 1 | 3 | 4 |
| | | % of Total | 10.0% | 30.0% | 40.0% |
| Total | | Count | 5 | 5 | 10 |
| | | % of Total | 50.0% | 50.0% | 100.0% |

**Chi-Square Tests**

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | 1.667[a] | 1 | .197 | | |
| Continuity Correction[b] | .417 | 1 | .519 | | |
| Likelihood Ratio | 1.726 | 1 | .189 | | |
| Fisher's Exact Test | | | | .524 | .262 |
| Linear-by-Linear Association | 1.500 | 1 | .221 | | |
| N of Valid Cases | 10 | | | | |

a. 4 cells (100.0%) have expected count less than 5. The minimum expected count is 2.00.

b. Computed only for a 2x2 table

**Problem 7: Suppose we want to define the "Social Status" on the basis of "Income" Variable using the following data.**

| Income | New value | Meaning |
|---|---|---|
| 0-3000 | 1 | Lower Class |
| 3001-10000 | 2 | Lower Middle Class |
| 10001-25000 | 3 | Middle Class |
| 25001-100000 | 4 | Higher Middle Class |
| 100000 above | 5 | Higher Class |

**Income**

**20000, 3200, 900, 45000, 1800, 11000, 7000, 245000, 35000, 48000, 32000, 56000, 22000, 125000**

- **Construct a frequency Distribution Table for "Social Status"**
- **Draw a pie Chart and Bar Diagram for the Appropriate Variable from the above data.**

**Command1:**

```
FREQUENCIES VARIABLES=Income
  /FORMAT=NOTABLE
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM
SEMEAN MEAN MEDIAN MODE SUM SKEWNESS SESKEW KURTOSIS
SEKURT
  /ORDER=ANALYSIS.
```

**Command2:**

**For Bar Chart:**

```
GRAPH
  /BAR(SIMPLE)=PCT BY Social_Status.
```

**For Pie Chart:**

```
GRAPH
  /PIE=PCT BY Social_Status.
```

**Problem8: Let us consider the following data.**

| Sales Experience (In Year) *X* | Annual Sales *Y* |
|---|---|
| 1 | 80 |
| 3 | 97 |
| 4 | 92 |
| 4 | 102 |
| 6 | 103 |
| 7 | 98 |
| 8 | 419 |
| 10 | 123 |
| 11 | 110 |
| 13 | 125 |

- Draw a Scatter Diagram and Interpret Your result.
- Perform Correlation Analysis and Interpret your Result. Also test whether there is any Correlation between *X* and *Y*

**Command1:**

```
GRAPH
  /SCATTERPLOT(BIVAR)=X WITH Y
  /MISSING=LISTWISE.
```

**<u>Interpretation:</u>** From the Scatter diagram, it can be concluded that *X* and *Y* are positively related. That means, as *X* Increase, *Y* also increase.

**Command2:**

```
CORRELATIONS
```

```
/VARIABLES=X Y
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.
```

**Table for Correlation**

**Correlations**

| | | Sales Experience (in Year) | Annual Sales |
|---|---|---|---|
| Sales Experience (in Year) | Pearson Correlation | 1 | .239 |
| | Sig. (2-tailed) | | .506 |
| | N | 10 | 10 |
| Annual Sales | Pearson Correlation | .239 | 1 |
| | Sig. (2-tailed) | .506 | |
| | N | 10 | 10 |

**Interpretation:** The value of Correlation Coefficient $r = 0.239$ Which Implies that There is a weak positive Association between $X$ and $Y$

Hypothesis to be tested

$H_0: \rho = 0$ ; There is no association between $X$ and $Y$

$H_1: \rho \neq 0$; There is association between $X$ and $Y$

**Comment:** P Value $= 0.506$ , since P value is Greeter Then 0.05. So we may accept our null Hypothesis at 5% Level of Significance and concluded that the correlation coefficient is equal to zero , There is no linear association between $X$ and $Y$.

**Problem 9: Calculate the Correlation Coefficient and Comment on your Result.**

| Price (tk) | Supply(Kg) |
|---|---|
| 10 | 30 |
| 12 | 33 |
| 18 | 45 |
| 16 | 44 |
| 15 | 42 |
| 19 | 48 |
| 18 | 47 |
| 17 | 46 |
| 19 | 30 |
| 16 | 35 |

| | |
|---|---|
| 14 | 45 |
| 10 | 44 |
| 9 | 30 |

**Command:**

```
CORRELATIONS
  /VARIABLES=X Y
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

**Table for Correlation**

**Correlations**

| | | Price | Suplly |
|---|---|---|---|
| Price | Pearson Correlation | 1 | .477 |
| | Sig. (2-tailed) | | .100 |
| | N | 13 | 13 |
| Suplly | Pearson Correlation | .477 | 1 |
| | Sig. (2-tailed) | .100 | |
| | N | 13 | 13 |

**Interpretation:** Correlation Coefficient $r = 0.477$ which implies that there is a moderate positive Correlation Between Price and supply. So the Hypothesis is to be tested

I.e. $H_0: \rho = 0;$ there is no association between Price and supply

$H_1: \rho \neq 0;$ There is association between Price and supply.

**Comment**: P value $= 0.100$ is Gretter then 0.05 then we accept our null Hypothesis with 5% level of Significance.

So that, there is no association between Pruce and Supply. There is no linear association between Price and Supply.

**Problem 10:**

| $Y$ Sales | $X_1$ (Experience) | $X_2$(Age) | $X_3$(Education) | $X_4$(Gender) | $X_5$(Region) |
|---|---|---|---|---|---|
| 15 | 2 | 30 | 10 | Male | Dhak |
| 24 | 4 | 45 | 12 | Female | Raj |
| 12 | 1 | 22 | 10 | Male | Khul |
| 16 | 3 | 25 | 13 | Male | Dhak |
| 18 | 6 | 30 | 14 | Female | Khul |
| 17 | 2 | 31 | 15 | Female | Dhak |
| 19 | 5 | 38 | 10 | Male | Raj |
| 16 | 3 | 24 | 12 | Female | Dhak |
| 13 | 2 | 19 | 13 | Male | Dhak |
| 15 | 3 | 45 | 15 | Male | Raj |

- Fit a Regression Line of $Y$ on $X_1, X_2, \dots \dots . X_5$ also Interpret your Result.
- Test Whether the covariates have significant impact on $Y$ or not Comment on your Result.

## Command:

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Y
  /METHOD=ENTER X1 X2 X3 X4 Dhak.
```

## Table for Regression

### Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .922[a] | .850 | .663 | 1.959 |

a. Predictors: (Constant), Dhak, Gender, Education, Experience, Age

### ANOVA[a]

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 87.152 | 5 | 17.430 | 4.543 | .084[b] |
| | Residual | 15.348 | 4 | 3.837 | | |
| | Total | 102.500 | 9 | | | |

a. Dependent Variable: Sales

b. Predictors: (Constant), Dhak, Gender, Education, Experience, Age

### Coefficients[a]

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 8.213 | 4.713 | | 1.743 | .156 |
| | Experience | .796 | .548 | .359 | 1.453 | .220 |
| | Age | .243 | .094 | .659 | 2.587 | .061 |
| | Education | -.541 | .375 | -.313 | -1.443 | .223 |
| | Gender | 3.006 | 1.472 | .460 | 2.042 | .111 |
| | Dhak | 1.610 | 1.705 | .251 | .944 | .399 |

a. Dependent Variable: Sales

## Interpret & Regression Coefficient

Here the Regression equation is $Y = 8.213 + 0.796X_1 + 0.243X_2 - 0.541X_3 + 3.006X_4 + 1.610X_5$

$\alpha = 8.213$ When all the independent variables are Equal to Zero Then the average value of $Y$ is 8.213.
$\hat{\beta}_{Exp} = 0.796$ For 1 Unit increase in $X_1$, then the average value of $Y$ increase 0.796
$\hat{\beta}_{Age} = 0.243$ For 1 Unit Increase in $X_1$ it increase the average value of $Y$ is 0.243
$\hat{\beta}_{Edu} = -0.541$ For 1 Unit Increase in $X_1$ it Decrease the average value of $Y$ is 0.541
$\hat{\beta}_{Gender} = 3.006$ On the average male Individual Sales 3.006 Higher Compared to Female Individuals.
$\hat{\beta}_{Dhak} = 1.610$ On the average region Individual sales 1.610 Higher compared to Khulna.

**Hypothesis:**

$$H_{01}: \beta_{Exp} = 0$$
$$H_{11}: \beta_{Exp} \neq 0$$

**Comment:** Hence The P Value is $0.220$ is Gretter Then 0.05 then we accept $H_0$ at 5% level of Significance. So that $X_1$ has no significant Impact on $Y$

**Problem 12: Consider the following data:**

| Gender | No of House Worked |
|---|---|
| Male | 40 |
| Female | 38 |
| Female | 60 |
| Male | 44 |
| Male | 40 |
| Female | 55 |
| Female | 42 |
| Male | 50 |
| Female | 60 |
| Male | 50 |
| Female | 40 |
| Male | 45 |
| Female | 50 |
| Female | 48 |
| Male | 40 |
| Female | 50 |
| Female | 50 |
| Male | 40 |
| Male | 40 |
| Male | 60 |

- Test Whether the Student work 40 Hours work in average or not.
- Weather the male and Female work the same average number of hours or not.

**Command 1:**

```
T-TEST
  /TESTVAL=40
  /MISSING=ANALYSIS
  /VARIABLES=House_Worked
  /CRITERIA=CI(.95).
```

**Interpret:** Hence the Hypohesis is

$$H_{01}: \mu = 40$$
$$H_{11}: \mu \neq 40$$

There the P Value Is 0.000 is less then 0.05 so the null Hypothesis is rejected at 5% level of significant. So the Students worked not 40 Average.

**Table for T Test**

**One-Sample Test**

| | Test Value = 40 | | | | | |
|---|---|---|---|---|---|---|
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
| | | | | | Lower | Upper |
| No of House Worked | 4.313 | 19 | .000 | 7.100 | 3.65 | 10.55 |

**Command 2:**

```
T-TEST GROUPS=Gender(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=House_Worked
  /CRITERIA=CI(.95).
```

**Table for Groups T Test**

**Interpret: Hypothesis:**

$$\mathbf{H_{01}: \mu_1 = \mu_2}$$
$$\mathbf{H_{01}: \mu_1 \neq \mu_2}$$

Here $\mu_1$ = Average Hour Worked Student by the male Students
And $\mu_2$ = Average Hour Worked Student by the female Students

**<u>Comment:</u>** Here the P Value = 0.189 is Gretter Then 0.05 so that we accept our Null Hypothesis at 5% level of significance. So that males and females work the same average number of hours.

## <u>Design of Experiments</u>

Problem 13: Consider the following data;

| Fertilizer 1 | Fertilizer 2 | Fertilizer 3 |
|---|---|---|
| 77 | 72 | 76 |
| 81 | 58 | 85 |
| 71 | 74 | 82 |
| 76 | 60 | 80 |
| 80 | 70 | 77 |

Perform ANNOVA and test null hypothesis concerning Fertilizers at 5% level of Significant Interpret your Result.

**Command:**

```
ONEWAY Production BY Fertilizers
  /MISSING ANALYSIS.
```

**<u>Interpret:</u>** Hypothesis: $H_0: \mu_1 = \mu_2 = \mu_3$

$H_1: \mu_1 \neq \mu_2 \neq \mu_3$/ at least two are not equal.

Here $\mu_1$is average Production Under Fertilizer 1

$\mu_2$ Average Production Under Fertilizer 2

$\mu_3$ Average Production under Fertilizer 3

**<u>Table for One way Classification</u>**

### ANOVA

Production of the land

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 478.800 | 2 | 239.400 | 8.737 | .005 |
| Within Groups | 328.800 | 12 | 27.400 | | |
| Total | 807.600 | 14 | | | |

**<u>Comment:</u>** Hence The P Value Is 0.005 is Less then the 0.05 so we reject our null hypothesis at 5% level of significance that means average production significantly differs under Fertilizers.

Since $H_0$ is Rejected that's why we have to perform LSD for pairwise comparison.

Here $H_{01}: \mu_1 = \mu_2$

$H_{11}: \mu_1 \neq \mu_2$

**<u>Command For LSD:</u>**

```
ONEWAY Production BY Fertilizers
  /PLOT MEANS
  /MISSING ANALYSIS
  /POSTHOC=LSD ALPHA(0.05).
```

**Comment:** Here From the table we find that the P Value is 0.010 is less then 0.05 so we reject our null hypothesis. That means Average yield under Fertilizers 1 & 2 Significantly Differs.

**Hypothesis:** $H_{02}: \mu_2 = \mu_3$

$H_{12}: \mu_2 \neq \mu_3$

**Table for LSD**

**Multiple Comparisons**

Dependent Variable: Production of the land

LSD

| (I) Fertilizer Of the land | (J) Fertilizer Of the land | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 1 | 2 | 10.200* | 3.311 | .010 | 2.99 | 17.41 |
| | 3 | -3.000 | 3.311 | .383 | -10.21 | 4.21 |
| 2 | 1 | -10.200* | 3.311 | .010 | -17.41 | -2.99 |
| | 3 | -13.200* | 3.311 | .002 | -20.41 | -5.99 |
| 3 | 1 | 3.000 | 3.311 | .383 | -4.21 | 10.21 |
| | 2 | 13.200* | 3.311 | .002 | 5.99 | 20.41 |

*. The mean difference is significant at the 0.05 level.

**Comment:** Here From the table we find the P Value is 0.002 is less then 0.05 So we accept our Null Hypothesis at 5% level of significance. That means average yield of fertilizer 2 & 3 significantly differs.

**Hypothesis:** $H_{03}: \mu_1 = \mu_3$

$H_{13}: \mu_1 \neq \mu_3$

**Comment:** Here we find that P Value is 0.383 is Gretter then 0.05 Then we Accept our null Hypothesis at 5% level of significance. That means average yield of fertilizer 1 & 3 not significantly Differs.

**Problem 14: Consider the following data**.

| Treatment | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| | 39 | 26 | 35 | 26 | 24 | 28 |
| | 45 | 43 | 28 | 27 | 28 | 36 |
| | 28 | 32 | 36 | 30 | 31 | 45 |
| Yield | 32 | 40 | 24 | 25 | 29 | 41 |
| | 38 | 29 | 31 | 42 | 44 | 25 |

Perform ANNOVA and test the significance of difference between treatment means at 5% Level of significance. Also interpret your result.

**Command:**

```
ONEWAY Producton BY Fertilizers
  /PLOT MEANS
```

```
  /MISSING ANALYSIS
  /POSTHOC=LSD ALPHA(0.05).
```

**Hypothesis:** $H_0: \mu_1 = \mu_2 = \mu_3$
$H_1: \mu_1 \neq \mu_2 \neq \mu_3$/at least two are not equal

**ANNOVA Table For One Way Classification**

**ANOVA**

Production

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 167.900 | 5 | 33.580 | .676 | .646 |
| Within Groups | 1192.800 | 24 | 49.700 | | |
| Total | 1360.700 | 29 | | | |

**Comment:** Hence the P value is 0.646 is Gretter Then 0.05 so the null hypothesis is accepted at 5% level of significance. That means average yield of Fertilizer A B C D E and F are not significantly Differs.

Problem 15: The Following data relate a RBD with 4 Treatments in 3 Blocks.

| Treatment | | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Blocks | 1 | 22 | 26 | 25 | 25 |
| | 2 | 25 | 27 | 28 | 26 |
| | 3 | 26 | 29 | 38 | 20 |

Perform ANNOVA & test the null hypothesis concerning Block & treatment at 5% level of significance Interpret your result.

**Command (for Block):**

```
UNIANOVA Yield BY Block
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /POSTHOC=Block(LSD)
  /CRITERIA=ALPHA(0.05)
  /DESIGN=Block.
```

**Hypothesis:** Hypothesis to be tested For Block $H_{01}: \mu_{B_1} = \mu_{B_2} = \mu_{B_3}$
$H_{11}$: at least two are not equal

Here $\mu_{\beta_1}$ = Mean Production Under Block 1
$\mu_{\beta_2}$ = Mean Production Under Block 2
$\mu_{\beta_3}$ = Mean Production Under Block 3

**Table (For Block)**

**Tests of Between-Subjects Effects**

Dependent Variable: Yield

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | 106.167[a] | 2 | 53.083 | 4.324 | .048 |
| Intercept | 9185.333 | 1 | 9185.333 | 748.127 | .000 |
| Block | 106.167 | 2 | 53.083 | 4.324 | .048 |
| Error | 110.500 | 9 | 12.278 | | |
| Total | 9402.000 | 12 | | | |
| Corrected Total | 216.667 | 11 | | | |

a. R Squared = .490 (Adjusted R Squared = .377)

**Comment:** From the table we have the P value is 0.048 is Less then 0.05 so the null hypothesis is rejected at 5% level of significant. That means there is significant difference among the Blocks.

**Command for Treatments:**

```
UNIANOVA Yield BY Treatment
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /POSTHOC=Treatment(LSD)
  /CRITERIA=ALPHA(0.05)
  /DESIGN=Treatment.
```

**Table (For Treatment)**

## Tests of Between-Subjects Effects

Dependent Variable: Yield

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | 58.000[a] | 3 | 19.333 | .975 | .451 |
| Intercept | 9185.333 | 1 | 9185.333 | 463.126 | .000 |
| Treatment | 58.000 | 3 | 19.333 | .975 | .451 |
| Error | 158.667 | 8 | 19.833 | | |
| Total | 9402.000 | 12 | | | |
| Corrected Total | 216.667 | 11 | | | |

a. R Squared = .268 (Adjusted R Squared = -.007)

**Comment:** Here P Value is 0.451 is Gretter then 0.05 so we accept our null Hypothesis at 5% Level of Significance. So we write that there is no significant difference among the treatment.

# Advanced Statistical Analysis In SPSS For Windows

**Problem 1:** From SPSS Data "**Employee data.sav**" Analyze Correlation Analysis between Salary and Salary begin also interpret your Result

## Coding:

```
CORRELATIONS
  /VARIABLES=salary salbegin
  /PRINT=TWOTAIL NOSIG
  /STATISTICS DESCRIPTIVES
  /MISSING=PAIRWISE.
```

## Table:

**Descriptive Statistics**

| | Mean | Std. Deviation | N |
|---|---|---|---|
| Current Salary | $34,419.57 | $17,075.661 | 474 |
| Beginning Salary | $17,016.09 | $7,870.638 | 474 |

**Correlations**

| | | Current Salary | Beginning Salary |
|---|---|---|---|
| Current Salary | Pearson Correlation | 1 | .880** |
| | Sig. (2-tailed) | | .000 |
| | N | 474 | 474 |
| Beginning Salary | Pearson Correlation | .880** | 1 |
| | Sig. (2-tailed) | .000 | |
| | N | 474 | 474 |

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

**Interpretation:** According to this result we can mention that the **Coefficient of correlation $r = 0.880$** which is **highly positively correlated which is 1% level of significance.**

**Problem 2:** Fit a multiple regression to the data and test the significance of the regression equation.

| Yield (t/ha): y | Tillers (no/hill): $x_1$ | Plant Height (Cm): $x_2$ |
|---|---|---|
| 5.5 | 13.6 | 110.7 |
| 5.7 | 13.9 | 109.8 |
| 5.9 | 14.6 | 106.5 |
| 6.3 | 15.2 | 105.9 |
| 6.5 | 15.7 | 104.8 |
| 6.5 | 16.2 | 106.9 |
| 6.7 | 17.3 | 95.7 |
| 7.2 | 17.8 | 93.4 |
| 7.3 | 18.2 | 84.6 |
| 7.5 | 18.5 | 80.4 |
| 7.8 | 19 | 78.6 |
| 7.9 | 19.2 | 76.6 |

## Coding:

```
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA CHANGE ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT y
  /METHOD=ENTER x1 x2
  /SCATTERPLOT=(y ,*ZPRED)
  /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).
```

## Table:

**Correlations**

| | | y | x1 | x2 |
|---|---|---|---|---|
| Pearson Correlation | y | 1.000 | .988 | -.955 |
| | x1 | .988 | 1.000 | -.950 |
| | x2 | -.955 | -.950 | 1.000 |
| Sig. (1-tailed) | y | . | .000 | .000 |
| | x1 | .000 | . | .000 |
| | x2 | .000 | .000 | . |
| N | y | 12 | 12 | 12 |
| | x1 | 12 | 12 | 12 |
| | x2 | 12 | 12 | 12 |

## Coefficients[a]

| Model | | Unstandardized Coefficients B | Unstandardized Coefficients Std. Error | Standardized Coefficients Beta | t | Sig. | 95.0% Confidence Interval for B Lower Bound | 95.0% Confidence Interval for B Upper Bound | Correlations Zero-order | Correlations Partial | Correlations Part |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | (Constant) | 2.178 | 1.943 | | 1.121 | .291 | -2.216 | 6.573 | | | |
| | x1 | .335 | .063 | .827 | 5.339 | .000 | .193 | .477 | .988 | .872 | .259 |
| | x2 | -.011 | .010 | -.169 | -1.094 | .302 | -.032 | .011 | -.955 | -.343 | -.053 |

a. Dependent Variable: y

## ANOVA[a]

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 7.054 | 2 | 3.527 | 207.807 | .000[b] |
| | Residual | .153 | 9 | .017 | | |
| | Total | 7.207 | 11 | | | |

a. Dependent Variable: y

b. Predictors: (Constant), x2, x1

**Interpretation:** From the above outputs, we have

$$a = Constant = 2.178$$

$$b_1 = Partial\ Regression\ Coefficient\ of\ x_1 = 0.335$$

$$b_2 = Partial\ Regression\ coefficient\ of\ x_2 = -0011$$

Hence estimated regression equation is $y = 2.178 + 0.335x_1 - 0.011x_2$ and the effect of $x_1$ on $y$ is highly significant at 1% level of significant.

Again, The calculated value of $R^2 = 0.979$ and calculated value of F= $207.80$ so, F Ratio is highly significant.

Therefore, the combined effect of tiller number and plant height on yield is highly significant.

Since $R^2 = 0.979$ i.e. 97.9% variation of the yield is explained by tiller number and plant height.