**Lab Report of Probability and Statistics on SPSS**

Submitted to

**Senior Lecturer Mr. Arjun Kumar Gaire and Senior Lecturer Mr. Keshav Paudel**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Submitted by | Roll No. | Group | Lab No. | Lab Date | Submission Date |
| Shronal Duwal | 790343 | B | 6 | 2081/08/04 | 2081/08/18 |
| Lab Title: Correlation and Regression Analysis | | | | | |
| Signature of Instructor: | | | | | |

**Objectives:**

1. To present the scatter plot of data to observe the relationships.
2. To compute the Karl Pearson’s correlation coefficient and interpret the result.
3. To compute the value of Regression coefficient.

**Procedure for objective 1**:

|  |  |
| --- | --- |
| **Step 1:** For the construction of the scatter plot, from the option of “Graphs”, “Legacy Dialogs” is selected where among the various graph plots, “Scatter/Dot” is selected. |  |
| **Step 2:** After following the 1st step, the window as shown in the attached image gets opened. Here, the type of scatter plot to be constructed is selected where we choose “Simple Scatter”. |  |
| **Step 3:** After the selection of the type of scatter graph, the values/variables for the construction are chosen. Here, among the various options, “Data Structure & Algorithm” and “Mathematics-III” are chosen for “Y-Axis” and “X-Axis” respectively for the scatter plot. |  |

**Output of objective 1:**

|  |  |
| --- | --- |
| After completing above steps, the Output screen of the software produces a scatter plot where the values are of the given variables “Data Structure & Algorithm” and “Mathematics–III” for Y-axis and X-axis respectively. The so-constructed scatter plot shows the relation between the values of the respective variables. |  |

**Conclusion for objective 1:**

Following the above steps as mentioned, the scatter plot for various values can be constructed using the SPSS software. The scatter plot allows the user to compare the relation between the variables through the graph created using their respective values. The software allows the variation in the axis of the values for efficient understanding of the relation between the variables through the scatter plot.

**Procedure for objective 2:**

The formula required for the objective:

1. Karl Pearson’s Correlation Coefficient, ; (Direct Method)

|  |  |
| --- | --- |
| **Step 1:** From the options in the row consisting file, edit, view, data, etc. choose the option of “Analyze” where “Correlate” is selected. Then “Bivariate” option is to be selected for the computation of Karl Pearson’s correlation coefficient. |  |
| **Step 2:** After the 1st step, “Bivariate Correlation” window gets opened, where the variables with measure-scale can be chosen for the computation. Similarly, here all the values that can be used for calculations are selected and sent as “Variables”. And from the “Options”, the table to be formed can be modified. Here, means and standard deviation is added to the table as well. |  |

**Output of objective 2:**

|  |  |
| --- | --- |
| The output screen of the SPSS software then prompts the page that shows the correlation table that shows the calculated values of all the chosen variables. Here, the values of all the subjects were given, so the correlation coefficient shows the relation between all the included subjects. |  |

**Interpretation of the result:**

The correlation table highlights the relationships between various academic variables. Mathematics-III exhibits the strongest correlation with EDC at r=0.747r = 0.747, indicating a strong positive linear relationship, and the weakest correlation with Computer Graphics at r=0.305r = 0.305. Similarly, OOAD shows a high correlation with Computer Graphics at r=0.739r = 0.739, reflecting a strong association. EDC is also notably correlated with Data Structure & Algorithm at r=0.616r = 0.616, suggesting moderate dependence.

In contrast, Applied Sociology has moderate correlations with several variables, such as Data Structure & Algorithm at r=0.624r = 0.624, and Mathematics-III at r=0.523r = 0.523. Most of the correlations are statistically significant at the 0.01 level, emphasizing the linear relationships between variables. This indicates that certain subjects, particularly Mathematics-III, EDC, and OOAD, are more closely tied to other variables in this academic dataset.

**Conclusion for objective 2:**

Computing Karl Pearson's correlation coefficient provides a quantitative measure of the strength and direction of the linear relationship between two variables. A strong positive coefficient indicates that as one variable increases, the other also increases, while a strong negative coefficient suggests an inverse relationship. A value near zero signifies little to no linear correlation. The above step allows the user to create correlation coefficients table through which the correlation between the subjects can be analyzed and interpreted as required.

**Procedure for objective 3**:

The formula required for the objective:

1. Regression equation of x on y:
2. Regression equation of y on x:
3. Regression Coefficient:

|  |  |
| --- | --- |
| **Step 1:** For the calculation of the regression coefficient, the option can be obtained by first clicking on “Analyze”, then “Regression” and finally “Linear” in the sequential order. |  |
| **Step 2:** After the 1st step, “Linear Regression” window gets opened where “Dependent” variable and “Independent(s)” variable are selected for the computation and tabulation. After the respective selections of the variables, clicking on the Ok leads the software to create table as shown in the output table. |  |

**Output of objective 3:**

|  |  |
| --- | --- |
| The output screen of the SPSS software then prompts the page that shows the regression coefficient table that shows the calculated values of all the chosen variables.  Supposing that Mathematics (y) is the dependent variable and Data Structure & Algorithm (x) is the independent variable, the regression equation from the table is:  Y = 15.127 + 0.515X |  |

**Conclusion for objective 3:**

The regression coefficient quantifies the relationship between variables, indicating the average change in the dependent variable for a one-unit change in the independent variable. This measure is essential for understanding variable influence and building predictive models. The SPSS output shows that the regression analysis resulted in an adjusted R² of 0.280, indicating the model explains 28% of the variance in the dependent variable. The F-value is 18.899 with a significance level of 0.000, confirming the model's validity. The predictor variable has a significant positive effect on the dependent variable.