|  |  |  |  |
| --- | --- | --- | --- |
| **S No.** | **Title** | **Date** | **Signature** |
| 1 | INTRODUCTION TO SPSS |  |  |
| 2 | Practical 1. DESCRIPTIVE STATISTICS |  |  |
| 3 | Practical 2. DESCRIPTIVE STATISTICS |  |  |
| 4 | Practical 3. CORRELATION AND LINEAR REGRESSION |  |  |
| 5 | Practical 4. CORRELATION AND LINEAR REGRESSION |  |  |
| 6 | Practical 5. CORRELATION AND LINEAR REGRESSION |  |  |
| 7 | Practical 6. CORRELATION AND LINEAR REGRESSION |  |  |
| 8 | Practical 7. CORRELATION AND LINEAR REGRESSION |  |  |
| 9 | Practical 8. PROBABILITY  DISTRIBUTIONS |  |  |
| 10 | Practical 9. PROBABILITY  DISTRIBUTIONS |  |  |
| 11 | Practical 10. DESCRIPTIVE STATISTICS |  |  |
| 12 | Practical 11. CORRELATION AND LINEAR REGRESSION |  |  |

**INDEX**

**INTRODUCTION TO SPSS**

SPSS (Statistical Package for the Social Sciences) is a software package that is widely used for statistical analysis in social sciences, psychology, healthcare, marketing, and other fields. It’s graphical user interface has two views which can be toggled by clicking on one of the two tabs in the bottom left of the SPSS Statistics window. The 'Data View' shows a spreadsheet view of the cases (rows) and variables (columns). The 'Variable View' displays the metadata dictionary where each row represents a variable and shows the variable name, variable label, value label(s), print width, measurement type, and a variety of other characteristics. The statistical output is to a proprietary file format i.e .spv file.

**History**

SPSS was originally developed by Norman H. Nie, C. Hadlai (Tex) Hull, and Dale H. Bent at Stanford University in the late 1960s. The first version of the software, called SPSS/PC, was released in 1984 for the IBM PC. The company that developed SPSS, SPSS Inc., was acquired by IBM in 2009.

**Features**

SPSS offers a wide range of features for statistical analysis, data management, and data visualization. Some of the key features of SPSS include:

* Data preparation and management: SPSS allows users to import, clean, and manage data from a variety of sources, including spreadsheets, databases, and other statistical software.
* Statistical analysis: SPSS offers a wide range of statistical procedures, including descriptive statistics, inferential statistics, factor analysis, regression analysis, and more.
* Data visualization: SPSS provides various tools for creating charts, graphs, and tables to help users visualize their data.
* Syntax editor: SPSS allows users to write and execute syntax commands for automation of repetitive tasks.
* Reporting: SPSS offers customizable templates and tools for generating reports and presentations based on statistical results.

**Advantages**

SPSS has several advantages that make it popular among researchers and analysts. Some of these advantages include:

* User-friendly interface: SPSS has an intuitive user interface that makes it easy to use for users of all levels of expertise.
* Comprehensive statistical analysis: SPSS offers a wide range of statistical procedures, making it a comprehensive tool for data analysis.
* Compatibility with other software: SPSS is compatible with other statistical software and programming languages such as R, Python, and SAS.
* Availability of support: SPSS has a large community of users, and IBM provides support and resources to help users with any issues they encounter.

**Current version**

The current version of SPSS is IBM SPSS Statistics 29, which was released in 2022. This version includes several new features and improvements such as new survival model procedure, open-source extension procedures, UI and workbook enhancements.

**Practical 1. DESCRIPTIVE STATISTICS**

Q1. One of the major measures of the quality of service provided by an organization is the speed with which it responds to customer complaints. An internet service provider, had undergone a major improvement by recruiting well trained installation crews, supervisors and office staffs. The business objective of the company was to reduce the time between when the complaint it received and when it is resolved. During a recent month, the company received 50 complaints concerning internet installation. The data from the 50 complaints, collected by ISP, represent the number of hours between the receipt and the complaint:

27, 4, 52, 30, 22, 36, 26, 20, 23, 33, 68, 165, 32, 29, 28, 29, 26, 25, 1, 14, 13, 13, 10, 5, 19, 126, 110, 110, 29, 61, 35, 94, 31, 26, 5, 12, 4, 54, 5, 35, 137, 31, 27, 152, 2, 123, 81, 74, 27, 11

a. Compute the mean, median, first quartile, and third quartile.

b. Compute the range, interquartile range, variance, standard deviation, and coefficient of variation.

c. Construct a boxplot. Are the data skewed? If so, how?

d. On the basis of results of (a) through (c), if you had to tell the president of the company how long a customer should expect to wait to have a complaint resolved, what would you say? Explain.

**Solution,**

**SYNTAX**

|  |  |  |
| --- | --- | --- |
| **Statistics** | | |
| Number of Hours | | |
| N | Valid | 50 |
| Missing | 0 |
| Mean | | 43.0400 |
| Median | | 28.5000 |
| Percentiles | 25 | 13.7500 |
| 50 | 28.5000 |
| 75 | 55.7500 |

FREQUENCIES VARIABLES=Hours

/FORMAT=NOTABLE

/NTILES=4

/STATISTICS=MEAN MEDIAN

/ORDER=ANALYSIS.

**OUTPUT**

**SYNTAX**

EXAMINE VARIABLES=Hours

/PLOT BOXPLOT

/COMPARE GROUPS

/STATISTICS DESCRIPTIVES

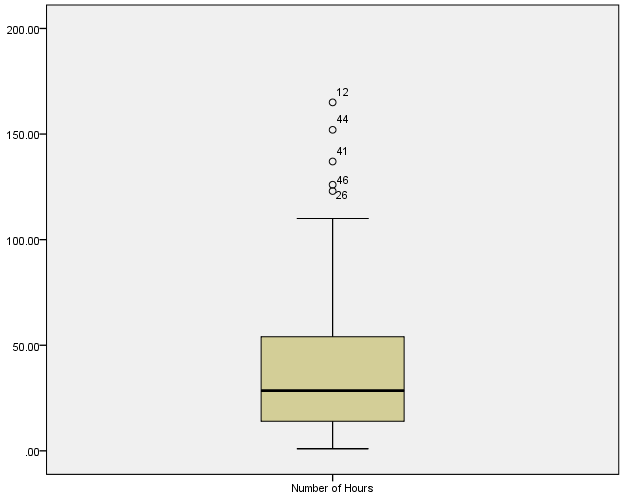
/CINTERVAL 95

/MISSING LISTWISE

/NOTOTAL.

**OUTPUT**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Descriptives** | | | | |
|  | | | Statistic | Std. Error |
| Number of Hours | Mean | | 43.0400 | 5.92924 |
| 95% Confidence Interval for Mean | Lower Bound | 31.1247 |  |
| Upper Bound | 54.9553 |  |
| 5% Trimmed Mean | | 39.1444 |  |
| Median | | 28.5000 |  |
| Variance | | 1757.794 |  |
| Std. Deviation | | 41.92606 |  |
| Minimum | | 1.00 |  |
| Maximum | | 165.00 |  |
| Range | | 164.00 |  |
| Interquartile Range | | 42.00 |  |
| Skewness | | 1.488 | .337 |
| Kurtosis | | 1.309 | .662 |

**Boxplot**

**CONCLUSION**

Mean = 43.04, Median = 28.5, First Quartile = 13.75, Third Quartile = 55.75

Range = 164, Interquartile Range = 42, Variance = 1757.794, Standard Deviation = 41.92606, Coefficient of Variation = 97.4

Data is Positively Skewed.

Minimum Hour is 1, Maximum hour is 165 & Mean Hour is 43.04.

**Practical 2. DESCRIPTIVE STATISTICS**

Q2. Use values below

61 55 18 90 32 27 85 26 75 61 51 88 34 65 48 55 67 50 62 3 7 81 20 42 68 36 10 50 68 33 23 64 22 51 82 64

a. Calculate values of descriptive statistics.

b. Construct the frequency distribution of size 15.

c. Present the grouped data in pie chart & histogram.

**Solution,**

**SYNTAX**

FREQUENCIES VARIABLES=values

/NTILES=4

/NTILES=10

/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM SEMEAN MEAN MEDIAN MODE SUM SKEWNESS SESKEW KURTOSIS SEKURT

/ORDER=ANALYSIS.

**OUTPUT**

|  |  |  |
| --- | --- | --- |
| **Statistics** | | |
| values | | |
| N | Valid | 36 |
| Missing | 0 |
| Mean | | 49.28 |
| Std. Error of Mean | | 4.003 |
| Median | | 51.00 |
| Mode | | 50a |
| Std. Deviation | | 24.015 |
| Variance | | 576.721 |
| Skewness | | -.173 |
| Std. Error of Skewness | | .393 |
| Kurtosis | | -.883 |
| Std. Error of Kurtosis | | .768 |
| Range | | 87 |
| Minimum | | 3 |
| Maximum | | 90 |
| Sum | | 1774 |
| Percentiles | 10 | 15.60 |
| 20 | 24.20 |
| 25 | 28.25 |
| 30 | 33.10 |
| 40 | 46.80 |
| 50 | 51.00 |
| 60 | 61.00 |
| 70 | 64.00 |
| 75 | 66.50 |
| 80 | 68.00 |
| 90 | 82.90 |

|  |
| --- |
| a. Multiple modes exist. The smallest value is shown |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **values** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 3 | 1 | 2.8 | 2.8 | 2.8 |
| 7 | 1 | 2.8 | 2.8 | 5.6 |
| 10 | 1 | 2.8 | 2.8 | 8.3 |
| 18 | 1 | 2.8 | 2.8 | 11.1 |
| 20 | 1 | 2.8 | 2.8 | 13.9 |
| 22 | 1 | 2.8 | 2.8 | 16.7 |
| 23 | 1 | 2.8 | 2.8 | 19.4 |
| 26 | 1 | 2.8 | 2.8 | 22.2 |
| 27 | 1 | 2.8 | 2.8 | 25.0 |
| 32 | 1 | 2.8 | 2.8 | 27.8 |
| 33 | 1 | 2.8 | 2.8 | 30.6 |
| 34 | 1 | 2.8 | 2.8 | 33.3 |
| 36 | 1 | 2.8 | 2.8 | 36.1 |
| 42 | 1 | 2.8 | 2.8 | 38.9 |
| 48 | 1 | 2.8 | 2.8 | 41.7 |
| 50 | 2 | 5.6 | 5.6 | 47.2 |
| 51 | 2 | 5.6 | 5.6 | 52.8 |
| 55 | 2 | 5.6 | 5.6 | 58.3 |
| 61 | 2 | 5.6 | 5.6 | 63.9 |
| 62 | 1 | 2.8 | 2.8 | 66.7 |
| 64 | 2 | 5.6 | 5.6 | 72.2 |
| 65 | 1 | 2.8 | 2.8 | 75.0 |
| 67 | 1 | 2.8 | 2.8 | 77.8 |
| 68 | 2 | 5.6 | 5.6 | 83.3 |
| 75 | 1 | 2.8 | 2.8 | 86.1 |
| 81 | 1 | 2.8 | 2.8 | 88.9 |
| 82 | 1 | 2.8 | 2.8 | 91.7 |
| 85 | 1 | 2.8 | 2.8 | 94.4 |
| 88 | 1 | 2.8 | 2.8 | 97.2 |
| 90 | 1 | 2.8 | 2.8 | 100.0 |
| Total | 36 | 100.0 | 100.0 |  |

**SYNTAX**

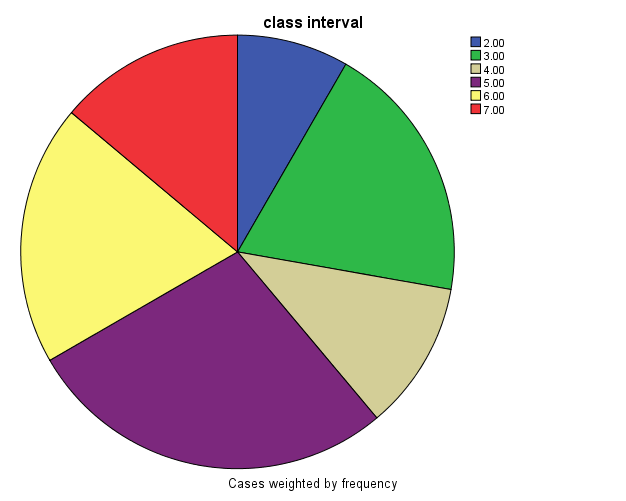
RECODE values (Lowest thru 2=1) (3 thru 17=2) (18 thru 32=3) (33 thru 47=4) (48 thru 62=5) (63 thru 77=6) (78 thru 92=7) (93 thru Highest=8) INTO class.

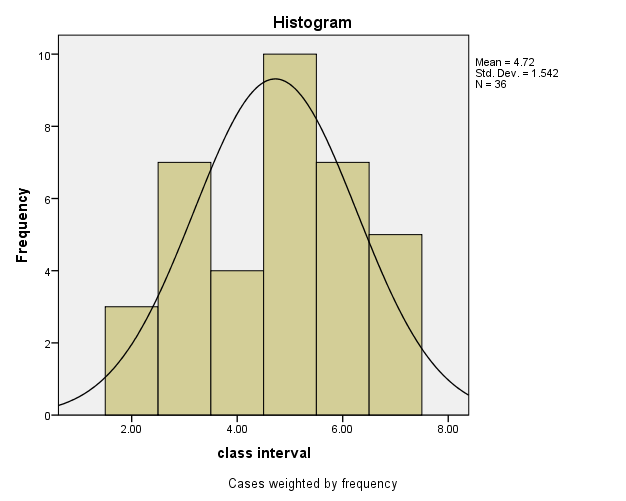
VARIABLE LABELS class 'class interval'.

EXECUTE.

**OUTPUT**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **class interval** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 2.00 | 3 | 8.3 | 8.3 | 8.3 |
| 3.00 | 7 | 19.4 | 19.4 | 27.8 |
| 4.00 | 4 | 11.1 | 11.1 | 38.9 |
| 5.00 | 10 | 27.8 | 27.8 | 66.7 |
| 6.00 | 7 | 19.4 | 19.4 | 86.1 |
| 7.00 | 5 | 13.9 | 13.9 | 100.0 |
| Total | 36 | 100.0 | 100.0 |  |

**Piechart**



**CONCLUSION**

Mean of above values is 49.29.

Median of above values is 51.

Mode of above values is 50.

Standard Deviation of above values is 24.015.

Range is 87.

Minimum Value is 3 & maximum value is 90.

Mean of frequencies is 4.72.

Standard Deviation of frequencies is 1.542.

**Practical 3. CORRELATION AND LINEAR REGRESSION**

Q3. Omprakash Sharma, owner of the Kathmandu Precast Company, has hired you as a part-time analyst. He was extremely pleased when you uncovered a positive relationship between the number of building permits issued and the amount of work available to his company. Now he wonders if it’s possible to use knowledge of interest rates on first mortgages to predict the number of building permits that will be issued each month. You collect a sample of data covering nine months.

|  |  |  |
| --- | --- | --- |
| Month | Building Permits (Y) | Interest rate (X) |
| 1 | 786 | 10.2 |
| 2 | 494 | 12.6 |
| 3 | 289 | 13.5 |
| 4 | 892 | 9.7 |
| 5 | 343 | 10.8 |
| 6 | 888 | 9.5 |
| 7 | 509 | 10.9 |
| 8 | 987 | 9.2 |
| 9 | 187 | 14.2 |

i. Calculate the correlation coefficient between building permits and interest rates and test its significance at 1%.

ii. Estimate the best fitting regression line and compute residual for month 9.

iii. Compute the coefficient of determination and interpret its meaning.

iv. Predict building permits when the interest rate increases by 9.7%.

**Solution,**

**SYNTAX**

CORRELATIONS

/VARIABLES=X Y

/PRINT=TWOTAIL NOSIG

/MISSING=PAIRWISE

|  |  |  |  |
| --- | --- | --- | --- |
| **Correlations** | | | |
|  | | Interest Rate | Building Permits |
| Interest Rate | Pearson Correlation | 1 | -.891\*\* |
| Sig. (2-tailed) |  | .001 |
| N | 9 | 9 |
| Building Permits | Pearson Correlation | -.891\*\* | 1 |
| Sig. (2-tailed) | .001 |  |
| N | 9 | 9 |
| \*\*. Correlation is significant at the 0.01 level (2-tailed). | | | |
| The correlation coefficient between building permits and interest rate is -0.891. | | | |

**SYNTAX**

REGRESSION

/DESCRIPTIVES MEAN STDDEV CORR SIG N

/MISSING LISTWISE

/STATISTICS COEFF OUTS R ANOVA CHANGE ZPP

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT Y

/METHOD=ENTER X

/SCATTERPLOT=(\*ZPRED ,\*ZRESID)

/SAVE PRED RESID.

**Regression**

|  |  |  |  |
| --- | --- | --- | --- |
| **Descriptive Statistics** | | | |
|  | Mean | Std. Deviation | N |
| Building Permits | 597.22 | 296.934 | 9 |
| Interest Rate | 11.178 | 1.8247 | 9 |

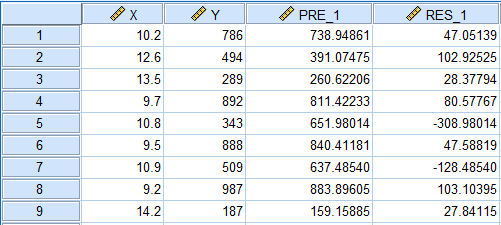
|  |  |  |  |
| --- | --- | --- | --- |
| **Variables Entered/Removeda** | | | |
| Model | Variables Entered | Variables Removed | Method |
| 1 | Interest Rateb | . | Enter |
| a. Dependent Variable: Building Permits | | | |
| b. All requested variables entered. | | | |

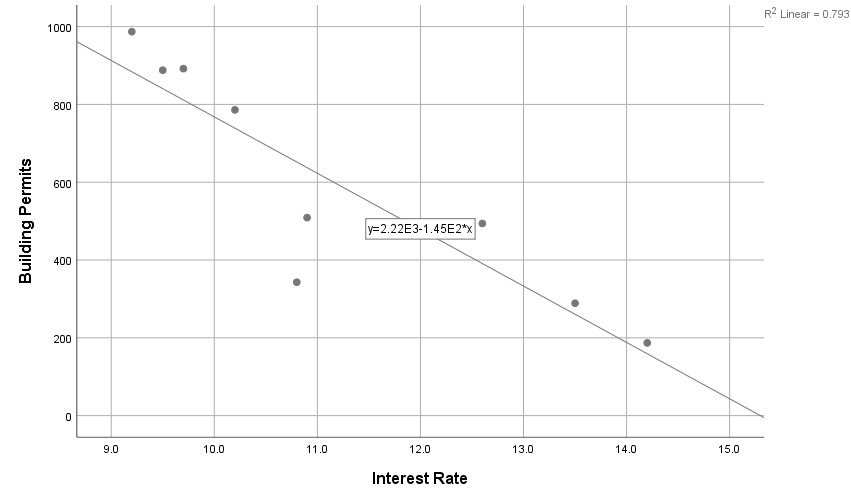
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summaryb** | | | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
| R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .891a | .793 | .764 | 144.298 | .793 | 26.876 | 1 | 7 | .001 |
| a. Predictors: (Constant), Interest Rate | | | | | | | | | |
| b. Dependent Variable: Building Permits | | | | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 559606.629 | 1 | 559606.629 | 26.876 | .001b |
| Residual | 145752.926 | 7 | 20821.847 |  |  |
| Total | 705359.556 | 8 |  |  |  |
| a. Dependent Variable: Building Permits | | | | | | |
| b. Predictors: (Constant), Interest Rate | | | | | | |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations | | | |
| B | Std. Error | Beta | Zero-order | Partial | Part |
| 1 | (Constant) | 2217.412 | 316.204 |  | 7.013 | .000 |  |  |  |
| Interest Rate | -144.947 | 27.959 | -.891 | -5.184 | .001 | -.891 | -.891 | -.891 |
| a. Dependent Variable: Building Permits | | | | | | | | | | |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations | | |
| B | Std. Error | Beta | Zero-order | Partial | Part |
| 1 | (Constant) | 2217.412 | 316.204 |  | 7.013 | .000 |  |  |  |
| Interest Rate | -144.947 | 27.959 | -.891 | -5.184 | .001 | -.891 | -.891 | -.891 |
| a. Dependent Variable: Building Permits | | | | | | | | | | |





**CONCLUSION**

The best fitting regression line is

Y= -144.947 + 2217.412 X

The residual for month 9 is 27.84115.

1. Coefficient of determination = R^2 = 0.793 = 79.3%

It means that 79.3% of the variation is explained by Interest rate and the rest 20.7% is due to unknown factors.

1. Here,

Predicted building permits is 811.42233 when the increase rate increases by 9.7%.

**Practical 4. CORRELATION AND REGRESSION**

Q4. Management of a soft drink bottling company wants to develop a method for allocating delivery cost to customers. Aldo one cost clearly relates good travel time within a particular route, another variable, another variable cost reflects the time required to unload the cases of soft drink at the delivery point. A sample of 10 deliveries within a territory was selected. The delivery times and the number of cases delivered were recorded as follows:

|  |  |  |
| --- | --- | --- |
| Customer | Number of cases | Delivery times (minutes) |
| 1 | 52 | 32.1 |
| 2 | 64 | 34.8 |
| 3 | 95 | 37.8 |
| 4 | 116 | 38.5 |
| 5 | 143 | 44.2 |
| 6 | 161 | 43.0 |
| 7 | 184 | 49.4 |
| 8 | 218 | 56.8 |
| 9 | 254 | 61.2 |
| 10 | 267 | 58.2 |

1. Find the correlation coefficient between delivery times and the number of cases delivered.
2. Develop a regression model to predict delivery time, based on the number of cases delivered.
3. Interpret the meaning of slope in this problem.
4. Predict the delivery time for 150 cases of soft drink.
5. Compute the standard error of the estimate and in interpret its meaning.
6. Determine the coefficient of determination and explain its meaning in this problem.
7. Compute residual for customer 7.

**Solution,**

**SYNTAX**

CORRELATIONS

/VARIABLES=X Y

/PRINT=TWOTAIL NOSIG

/MISSING=PAIRWISE.

|  |  |  |  |
| --- | --- | --- | --- |
| **Correlations** | | | |
|  | | Number of cases | Delivery time (minutes) |
| Number of cases | Pearson Correlation | 1 | .981\*\* |
| Sig. (2-tailed) |  | .000 |
| N | 10 | 10 |
| Delivery time (minutes) | Pearson Correlation | .981\*\* | 1 |
| Sig. (2-tailed) | .000 |  |
| N | 10 | 10 |
| \*\*. Correlation is significant at the 0.01 level (2-tailed). | | | |

**SYNTAX**

\*Curve Estimation. TSET

NEWVAR=NONE.CURVEFIT

/VARIALBLES=Y WITH X

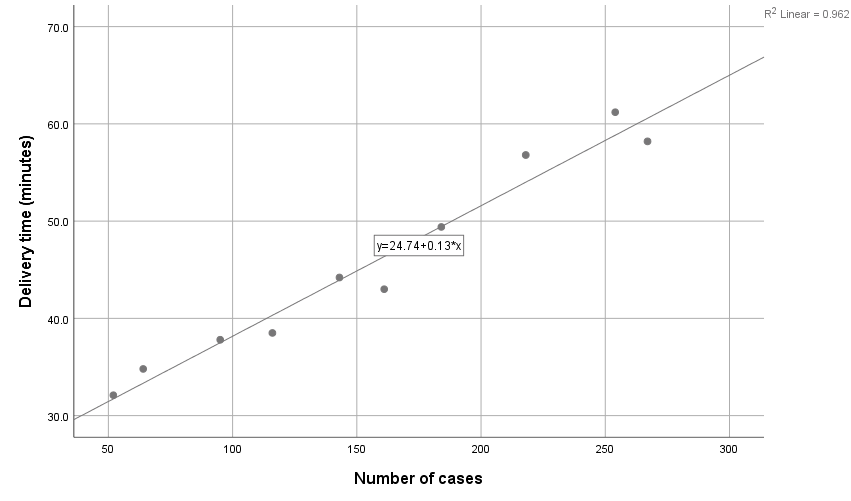
/CONSTANT

/MODEL=LINEAR

/PRINT ANOVA

/PLOT FIT.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations | | |
| B | Std. Error | Beta | Zero-order | Partial | Part |
| 1 | (Constant) | 24.744 | 1.603 |  | 15.432 | .000 |  |  |  |
| Number of cases | .134 | .009 | .981 | 14.323 | .000 | .981 | .981 | .981 |
| a. Dependent Variable: Delivery time (minutes) | | | | | | | | | |



**SYNTAX**

REGRESSION

/MISSING LISTWISE

/STATISTICS COEFF OUTS R ANOVA CHANGE ZPP

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT Y

/METHOD=ENTER X

/SCATTERPLOT=(\*ZPRED ,\*ZRESID)

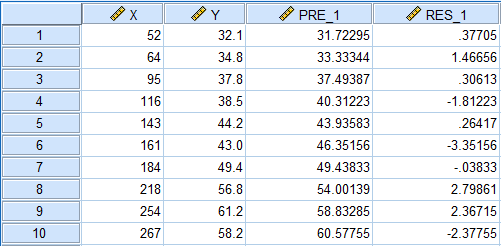
/SAVE PRED RESID.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summaryb** | | | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
| R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .981a | .962 | .958 | 2.1232 | .962 | 205.149 | 1 | 8 | .000 |
| a. Predictors: (Constant), Number of cases | | | | | | | | | |
| b. Dependent Variable: Delivery time (minutes) | | | | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 924.797 | 1 | 924.797 | 205.149 | .000b |
| Residual | 36.063 | 8 | 4.508 |  |  |
| Total | 960.860 | 9 |  |  |  |
| a. Dependent Variable: Delivery time (minutes) | | | | | | |
| b. Predictors: (Constant), Number of cases | | | | | | |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations | | |
| B | Std. Error | Beta | Zero-order | Partial | Part |
| 1 | (Constant) | 24.744 | 1.603 |  | 15.432 | .000 |  |  |  |
| Number of cases | .134 | .009 | .981 | 14.323 | .000 | .981 | .981 | .981 |
| a. Dependent Variable: Delivery time (minutes) | | | | | | | | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Residuals Statisticsa** | | | | | |
|  | Minimum | Maximum | Mean | Std. Deviation | N |
| Predicted Value | 31.723 | 60.578 | 45.600 | 10.1368 | 10 |
| Residual | -3.3516 | 2.7986 | .0000 | 2.0018 | 10 |
| Std. Predicted Value | -1.369 | 1.478 | .000 | 1.000 | 10 |
| Std. Residual | -1.579 | 1.318 | .000 | .943 | 10 |
| a. Dependent Variable: Delivery time (minutes) | | | | | |



**CONCLUSION**

The correlation coefficient between delivery time and the number of cases delivered is 0.981.

Hence, the required regression model is

Y = 24.74 + 0.13 X

Here, we have a regression equation,

Y = 24.74 + 0.13 X

Hence, The value changes by 0.13 units for every one-unit increase in X.

Here, we have

Y = 24.74 + 0.13 X = 24.74 + 0.13 \* 150 = 44.24

Therefore, the delivery time for 150 cases of soft drinks is 44.24 min.

The standard error of the estimate is 2.1232.

Hence, the average variation or scatteredness of the observed data point around the regression line is 2.1232.

Coefficient of determination = R^2 = 0.962 = 96.2%

It means that 96.2% of the variation is explained by the number of cases and the rest 3.8% is due to unknown factors.

Hence, the residual for customer 7 is -0.03833.

**Practical 5. CORRELATION AND LINEAR REGRESSION**

Q5. Calculate Karl Pearson’s correlation coefficient, test its significance and find the limits of population correlation coefficient. Find coefficient of determination.

|  |  |
| --- | --- |
| Nutrition | Child Mortality |
| 12.1 | 9.5 |
| 9.1 | 9.2 |
| 26 | 11.8 |
| 6.4 | 6.4 |
| 9.5 | 7.3 |
| 18.5 | 20.3 |
| 22.8 | 24.4 |
| 17.4 | 21.1 |
| 13.9 | 10.7 |
| 3.2 | 3.5 |
| 30.2 | 11.8 |
| 15.7 | 12.3 |
| 8.7 | 11.8 |
| 5.6 | 9.4 |
| 11.2 | 8.3 |
| 9.8 | 9 |
| 8.4 | 4.7 |

**Solution,**

**SYNTAX**

CORRELATIONS

/VARIABLES=X Y

/PRINT=TWOTAIL NOSIG

/MISSING=PAIRWISE.

|  |  |  |  |
| --- | --- | --- | --- |
| **Correlations** | | | |
|  | | Nutrition | Child Mortality |
| Nutrition | Pearson Correlation | 1 | .626\*\* |
| Sig. (2-tailed) |  | .007 |
| N | 17 | 17 |
| Child Mortality | Pearson Correlation | .626\*\* | 1 |
| Sig. (2-tailed) | .007 |  |
| N | 17 | 17 |
| \*\*. Correlation is significant at the 0.01 level (2-tailed). | | | |

Therefore, The Karl Pearson’s correlation coefficient is 0.626.

The significance is 0.007<0.01, so it is significant.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summaryb** | | | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
| R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .626a | .391 | .351 | 4.5963 | .391 | 9.647 | 1 | 15 | .007 |
| a. Predictors: (Constant), Nutrition | | | | | | | | | |
| b. Dependent Variable: Child Mortality | | | | | | | | | |

Here, The coefficient of determination = R^2 = 0.391 = 39.1%

It means that 39.1% of the variation is explained by the nutrition and the rest 60.9% is due to unknown factors.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 203.806 | 1 | 203.806 | 9.647 | .007b |
| Residual | 316.893 | 15 | 21.126 |  |  |
| Total | 520.699 | 16 |  |  |  |
| a. Dependent Variable: Child Mortality | | | | | | |
| b. Predictors: (Constant), Nutrition | | | | | | |

**CONCLUSION**

The Karl Pearson’s correlation coefficient is 0.626.

The significance is 0.007<0.01, so it is significant.

Here, The coefficient of determination = R^2 = 0.391 = 39.1%

It means that 39.1% of the variation is explained by the nutrition and the rest 60.9% is due to unknown factors.

|  |
| --- |
| a. Dependent Variable: Child Mortality |
| b. Predictors: (Constant), Nutrition |

**Practical 6. CORRELATION AND LINEAR REGRESSION**

Q6. Calculate Spearman’s rank correlation coefficient test its significance and find the limits of population correlation coefficient.

|  |  |
| --- | --- |
| Nutrition | Child Mortality |
| 12.1 | 9.5 |
| 9.1 | 9.2 |
| 26 | 11.8 |
| 6.4 | 6.4 |
| 9.5 | 7.3 |
| 18.5 | 20.3 |
| 22.8 | 24.4 |
| 17.4 | 21.1 |
| 13.9 | 10.7 |
| 3.2 | 3.5 |
| 30.2 | 11.8 |
| 15.7 | 12.3 |
| 8.7 | 11.8 |
| 5.6 | 9.4 |
| 11.2 | 8.3 |
| 9.8 | 9 |
| 8.4 | 4.7 |

**Solution,**

**SYNTAX**

NONPAR CORR

/VARIABLES=X Y

/PRINT=SPEARMAN TWOTAIL NOSIG

/MISSING=PAIRWISE.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **OUTPUT Correlations** | | | | |
|  | | | Nutrition | Child Mortality |
| Spearman's rho | Nutrition | Correlation Coefficient | 1.000 | .779\*\* |
| Sig. (2-tailed) | . | .000 |
| N | 17 | 17 |
| Child Mortality | Correlation Coefficient | .779\*\* | 1.000 |
| Sig. (2-tailed) | .000 | . |
| N | 17 | 17 |
| \*\*. Correlation is significant at the 0.01 level (2-tailed). | | | | |

**CONCLUSION**

Here, The Spearman’s Rank correlation coefficient is 0.779.

Since The significance level is 0 < 0.01, so it is significant.

**Practical 7. CORRELATION AND LINEAR REGRESSION**

Q7. Find the regression equation of percentage change in production on production and present the data using scattered plot. Find the estimated value of dependent variable for all observed value of independent variable. Test the significance of regression coefficients and overall regression coefficient. Find the coefficient of determination.

|  |  |
| --- | --- |
| Nutrition | Child Mortality |
| 12.1 | 9.5 |
| 9.1 | 9.2 |
| 26 | 11.8 |
| 6.4 | 6.4 |
| 9.5 | 7.3 |
| 18.5 | 20.3 |
| 22.8 | 24.4 |
| 17.4 | 21.1 |
| 13.9 | 10.7 |
| 3.2 | 3.5 |
| 30.2 | 11.8 |
| 15.7 | 12.3 |
| 8.7 | 11.8 |
| 5.6 | 9.4 |
| 11.2 | 8.3 |
| 9.8 | 9 |
| 8.4 | 4.7 |

**Solution,**

**SYNTAX**

REGRESSION

/DESCRIPTIVES MEAN STDDEV CORR SIG N

/MISSING LISTWISE

/STATISTICS COEFF OUTS R ANOVA COLLIN TOL CHANGE ZPP

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT Child\_Mortality

/METHOD=ENTER Nutrition

/SCATTERPLOT=(\*ZRESID ,\*ZPRED)

/RESIDUALS HISTOGRAM(ZRESID).

**OUTPUT**

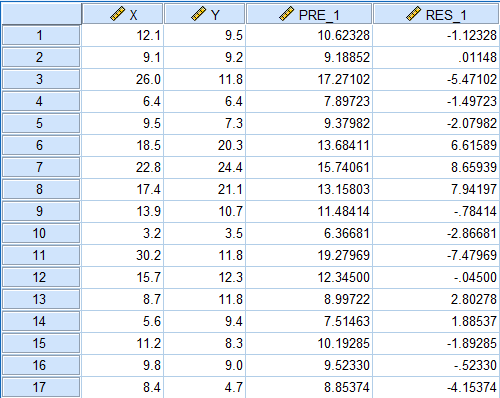
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summaryb** | | | | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
| R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .626a | .391 | .351 | 4.5963 | .391 | 9.647 | 1 | 15 | .007 |
| a. Predictors: (Constant), Nutrition | | | | | | | | | | |
| b. Dependent Variable: Child Mortality | | | | | | | | | | |

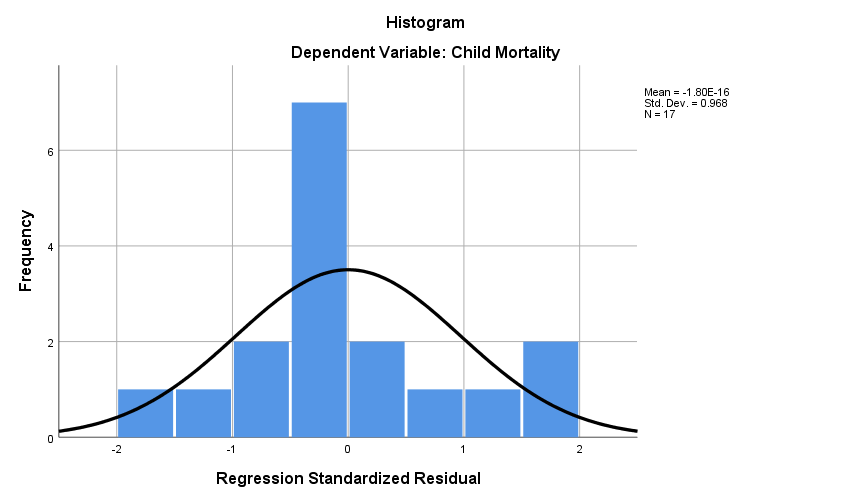
Here, The coefficient of determination = R^2 = 0.391

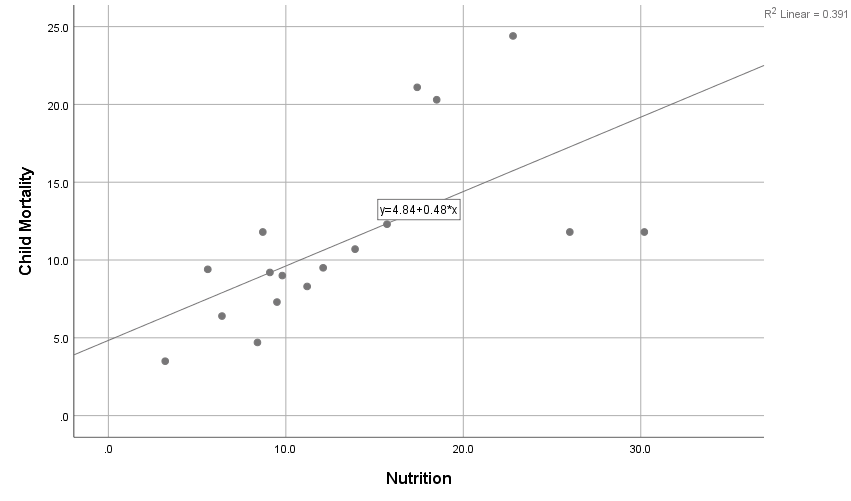
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 203.806 | 1 | 203.806 | 9.647 | .007b |
| Residual | 316.893 | 15 | 21.126 |  |  |
| Total | 520.699 | 16 |  |  |  |
| a. Dependent Variable: Child Mortality | | | | | | |
| b. Predictors: (Constant), Nutrition | | | | | | |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations | | |
| B | Std. Error | Beta | Zero-order | Partial | Part |
| 1 | (Constant) | 4.836 | 2.351 |  | 2.057 | .057 |  |  |  |
| Nutrition | .478 | .154 | .626 | 3.106 | .007 | .626 | .626 | .626 |
| a. Dependent Variable: Child Mortality | | | | | | | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Residuals Statisticsa** | | | | | |
|  | Minimum | Maximum | Mean | Std. Deviation | N |
| Predicted Value | 6.367 | 19.280 | 11.265 | 3.5690 | 17 |
| Residual | -7.4797 | 8.6594 | .0000 | 4.4504 | 17 |
| Std. Predicted Value | -1.372 | 2.246 | .000 | 1.000 | 17 |
| Std. Residual | -1.627 | 1.884 | .000 | .968 | 17 |
| a. Dependent Variable: Child Mortality | | | | | |







Here, The regression equation of % change in production is

Y = 4.84 + 0.84\*X

|  |  |  |  |
| --- | --- | --- | --- |
| **Correlations** | | | |
|  | | Child Mortality | Nutrition |
| Pearson Correlation | Child Mortality | 1.000 | .626 |
| Nutrition | .626 | 1.000 |
| Sig. (1-tailed) | Child Mortality | . | .004 |
| Nutrition | .004 | . |
| N | Child Mortality | 17 | 17 |
| Nutrition | 17 | 17 |

Since 0.004<level of significance (i.e. 0.05) so it is significant.

**CONCLUSION**

The regression equation of % change in production is Y = 4.84 + 0.84X.

Regression Coefficients are significant since 0.004<level of significance (i.e. 0.05).

The coefficient of determination = R^2 = 0.391.

**Practical 8. PROBABILITY DISTRIBUTIONS**

Q8.Fit binomial distribution and find the probabilities of less than 4, less than equal to 4, 4, not equals to 4 and greater than 4 using data x and f.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| f | 5 | 25 | 35 | 48 | 65 | 41 | 28 | 9 | 4 |

**Solution,**

Here, Mean = 3.72

**SYNTAX**

COMPUTE ex=PDF.BINOM(x,8,3.72/8).

EXECUTE.

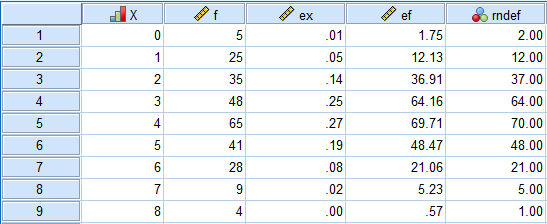
COMPUTE ef=260\*ex.

EXECUTE.

COMPUTE rndef=rnd(ef).

EXECUTE.

**OUTPUT**



**CONCLUSION**

Hence, the fitted binomial distribution is:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| P | 2 | 12 | 37 | 64 | 70 | 48 | 21 | 5 | 1 |

**Practical 9. PROBABILITY DISTRIBUTIONS**

Q9. Fit Poisson distribution and find the probabilities of less than 2, less than equal to 3, 5, not equals to 3 and greater than 1 using data X and f.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| f | 211 | 250 | 154 | 68 | 20 | 12 | 7 | 3 | 1 |

**Solution,**

Here, Mean = 1.34

**SYNTAX**

COMPUTE px=PDF.POISSON(x,1.34).

EXECUTE.

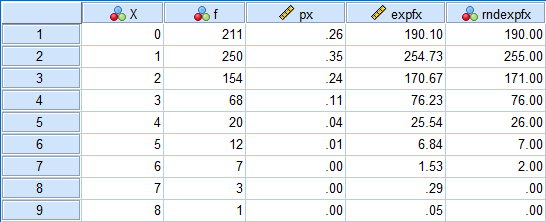
COMPUTE expfx=726\*px.

EXECUTE.

COMPUTE rndexpfx=RND(expfx).

EXECUTE.

**OUTPUT**



**CONCLUSION**

Hence, the fitted Poisson distribution is:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| f | 190 | 255 | 171 | 76 | 26 | 7 | 2 | 0 | 0 |

**Practical 10. DESCRIPTIVE STATISTICS**

Q10. Use the data given below to construct the frequency distribution.

| Gender | Weight (kg) | Height (cm) | Day of the week |

|--------|-------------|-------------|----------------|

| Others | 64.00 | 153.00 | Wednesday |

| Female | 78.00 | 163.00 | Wednesday |

| Male | 57.00 | 179.00 | Monday |

| Female | 60.00 | 150.00 | Saturday |

| Male | 75.00 | 60.00 | Tuesday |

| Female | 51.00 | 153.00 | Sunday |

| Male | 69.00 | 167.00 | Saturday |

| Others | 60.00 | 153.00 | Wednesday |

| Female | 77.00 | 166.00 | Friday |

| Female | 74.00 | 160.00 | Saturday |

| Others | 68.00 | 190.00 | Monday |

| Male | 81.00 | 187.00 | Saturday |

| Female | 56.00 | 191.00 | Saturday |

| Female | 72.00 | 151.00 | Monday |

| Male | 50.00 | 157.00 | Monday |

| Male | 57.00 | 151.00 | Sunday |

| Female | 48.00 | 161.00 | Friday |

| Female | 58.00 | 189.00 | Monday |

| Female | 47.00 | 170.00 | Saturday |

| Female | 52.00 | 185.00 | Saturday |

| Male | 72.00 | 165.00 | Thursday |

| Male | 53.00 | 168.00 | Wednesday |

| Female | 57.00 | 173.00 | Wednesday |

| Male | 59.00 | 180.00 | Saturday |

**Solution,**

**SYNTAX**

FREQUENCIES VARIABLES=Weight Height

/NTILES=4

/NTILES=10

/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM SEMEAN MEAN MEDIAN MODE SUM SKEWNESS SESKEW

KURTOSIS SEKURT

/HISTOGRAM NORMAL

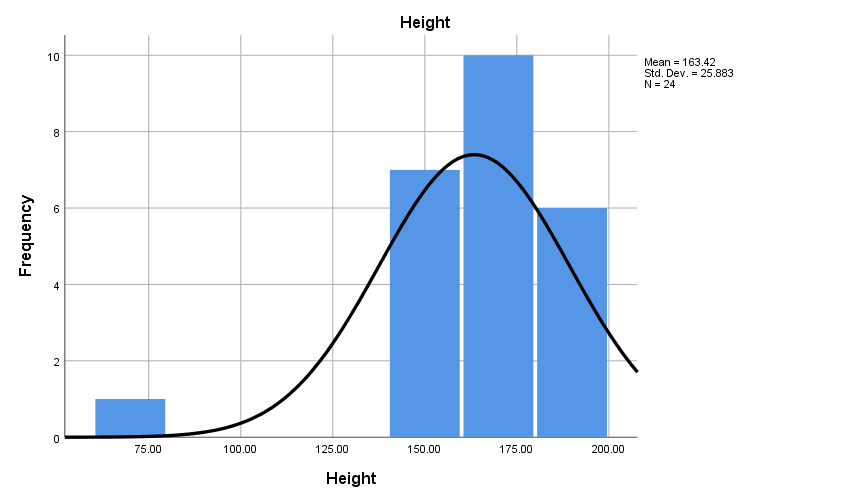
/ORDER=ANALYSIS.

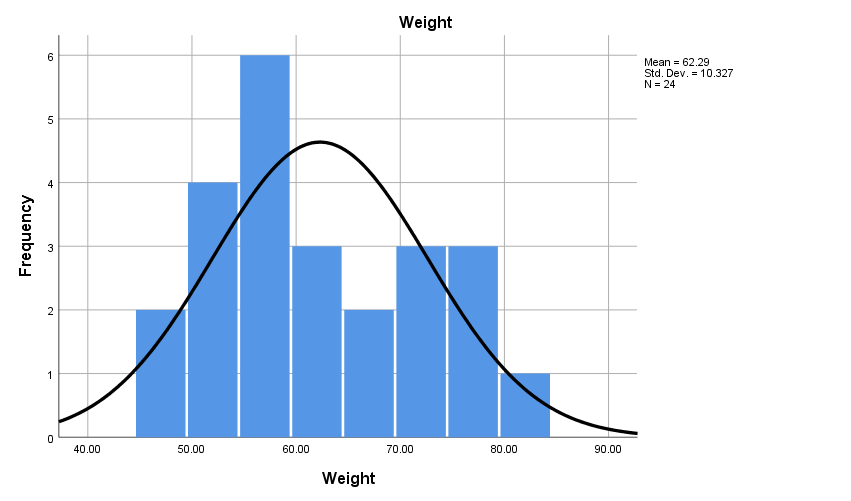
|  |  |  |  |
| --- | --- | --- | --- |
| Statistics | | | |
|  | | Weight | Height |
| N | Valid | 24 | 24 |
| Missing | 0 | 0 |
| Mean | | 62.2917 | 163.4167 |
| Std. Error of Mean | | 2.10802 | 5.28325 |
| Median | | 59.5000 | 165.5000 |
| Mode | | 57.00 | 153.00 |
| Std. Deviation | | 10.32717 | 25.88254 |
| Variance | | 106.650 | 669.906 |
| Skewness | | .304 | -2.808 |
| Std. Error of Skewness | | .472 | .472 |
| Kurtosis | | -1.179 | 11.246 |
| Std. Error of Kurtosis | | .918 | .918 |
| Range | | 34.00 | 131.00 |
| Minimum | | 47.00 | 60.00 |
| Maximum | | 81.00 | 191.00 |
| Sum | | 1495.00 | 3922.00 |
| Percentiles | 10 | 49.0000 | 150.5000 |
| 20 | 52.0000 | 153.0000 |
| 25 | 53.7500 | 153.0000 |
| 30 | 56.5000 | 155.0000 |
| 40 | 57.0000 | 161.0000 |
| 50 | 59.5000 | 165.5000 |
| 60 | 64.0000 | 168.0000 |
| 70 | 70.5000 | 176.0000 |
| 75 | 72.0000 | 179.7500 |
| 80 | 74.0000 | 185.0000 |
| 90 | 77.5000 | 189.5000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Weight | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 47.00 | 1 | 4.2 | 4.2 | 4.2 |
| 48.00 | 1 | 4.2 | 4.2 | 8.3 |
| 50.00 | 1 | 4.2 | 4.2 | 12.5 |
| 51.00 | 1 | 4.2 | 4.2 | 16.7 |
| 52.00 | 1 | 4.2 | 4.2 | 20.8 |
| 53.00 | 1 | 4.2 | 4.2 | 25.0 |
| 56.00 | 1 | 4.2 | 4.2 | 29.2 |
| 57.00 | 3 | 12.5 | 12.5 | 41.7 |
| 58.00 | 1 | 4.2 | 4.2 | 45.8 |
| 59.00 | 1 | 4.2 | 4.2 | 50.0 |
| 60.00 | 2 | 8.3 | 8.3 | 58.3 |
| 64.00 | 1 | 4.2 | 4.2 | 62.5 |
| 68.00 | 1 | 4.2 | 4.2 | 66.7 |
| 69.00 | 1 | 4.2 | 4.2 | 70.8 |
| 72.00 | 2 | 8.3 | 8.3 | 79.2 |
| 74.00 | 1 | 4.2 | 4.2 | 83.3 |
| 75.00 | 1 | 4.2 | 4.2 | 87.5 |
| 77.00 | 1 | 4.2 | 4.2 | 91.7 |
| 78.00 | 1 | 4.2 | 4.2 | 95.8 |
| 81.00 | 1 | 4.2 | 4.2 | 100.0 |
| Total | 24 | 100.0 | 100.0 |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Height | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 60.00 | 1 | 4.2 | 4.2 | 4.2 |
| 150.00 | 1 | 4.2 | 4.2 | 8.3 |
| 151.00 | 2 | 8.3 | 8.3 | 16.7 |
| 153.00 | 3 | 12.5 | 12.5 | 29.2 |
| 157.00 | 1 | 4.2 | 4.2 | 33.3 |
| 160.00 | 1 | 4.2 | 4.2 | 37.5 |
| 161.00 | 1 | 4.2 | 4.2 | 41.7 |
| 163.00 | 1 | 4.2 | 4.2 | 45.8 |
| 165.00 | 1 | 4.2 | 4.2 | 50.0 |
| 166.00 | 1 | 4.2 | 4.2 | 54.2 |
| 167.00 | 1 | 4.2 | 4.2 | 58.3 |
| 168.00 | 1 | 4.2 | 4.2 | 62.5 |
| 170.00 | 1 | 4.2 | 4.2 | 66.7 |
| 173.00 | 1 | 4.2 | 4.2 | 70.8 |
| 179.00 | 1 | 4.2 | 4.2 | 75.0 |
| 180.00 | 1 | 4.2 | 4.2 | 79.2 |
| 185.00 | 1 | 4.2 | 4.2 | 83.3 |
| 187.00 | 1 | 4.2 | 4.2 | 87.5 |
| 189.00 | 1 | 4.2 | 4.2 | 91.7 |
| 190.00 | 1 | 4.2 | 4.2 | 95.8 |
| 191.00 | 1 | 4.2 | 4.2 | 100.0 |
| Total | 24 | 100.0 | 100.0 |  |

Histogram





**CONCLUSION**

Hence the histogram above shows the frequency distribution.

**Practical 11. CORRELATION AND LINEAR REGRESSION**

Q11. Calculate the coefficient of determination and show it in a scatter diagram of the following data:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 50 | 55 | 55 | 60 | 65 | 70 | 65 | 60 |
| Y | 11 | 13 | 14 | 16 | 16 | 15 | 15 | 20 |

**SYNTAX**

REGRESSION

/MISSING LISTWISE

/STATISTICS CHANGE

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT Y

/METHOD=ENTER X.

**OUTPUT**

Regression

|  |  |  |  |
| --- | --- | --- | --- |
| Variables Entered/Removeda | | | |
| Model | Variables Entered | Variables Removed | Method |
| 1 | Xb | . | Enter |
| a. Dependent Variable: Y | | | |
| b. All requested variables entered. | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model Summary | | | | | |
| Model | Change Statistics | | | | |
| R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .250a | 2.571 | 1 | 6 | .160 |
| a. Predictors: (Constant), X | | | | | |

**SYNTAX**

\* Chart Builder.

GGRAPH

/GRAPHDATASET NAME="graphdataset" VARIABLES=X Y MISSING=LISTWISE REPORTMISSING=NO

/GRAPHSPEC SOURCE=INLINE

/FITLINE TOTAL=YES.

BEGIN GPL

SOURCE: s=userSource(id("graphdataset"))

DATA: X=col(source(s), name("X"), unit.category())

DATA: Y=col(source(s), name("Y"), unit.category())

GUIDE: axis(dim(1), label("X"))

GUIDE: axis(dim(2), label("Y"))

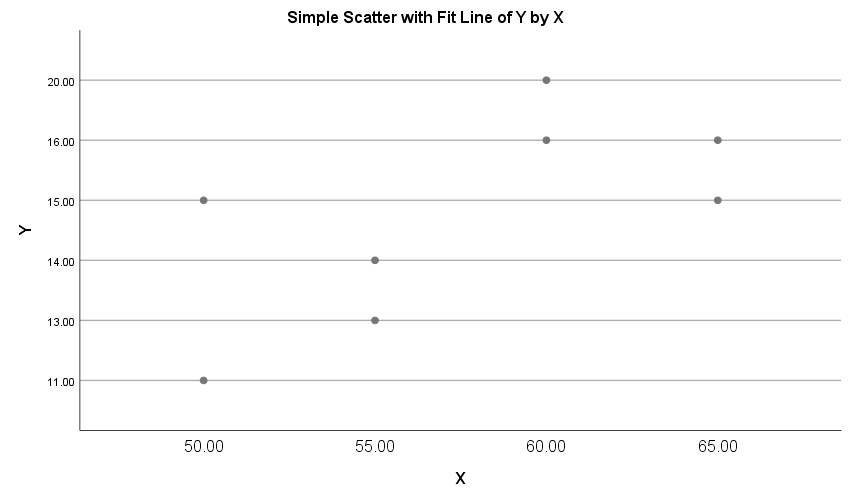
GUIDE: text.title(label("Simple Scatter with Fit Line of Y by X"))

ELEMENT: point(position(X\*Y))

END GPL.

**OUTPUT**

GGraph



**CONCLUSION**

The coefficient of determination obtained from the data above is 0.25.