**OBJECTIVE :** 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 INVESTMENT BY RECRUITING WELL TRAINED INSTALLATION CREWED 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 RESOLUTION OF THE COMPLAINTS:

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

1. COMPUTE THE MEAN, MEDIAN, FIRST QUARTILE AND THIRD QUARTILE.
2. COMPUTE THE RANGE, INTERQUARTILE RANGE, VARIANCE, STANDARD DEVIATION AND COEFFICIENT OF VARIATION.
3. CONSTRUCT A BOXPLOT. ARE THE DATA SKEWED? IF SO, HOW?
4. ON THE BASIS OF THE 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.

**WORKING EXPRESSION:**

We have, Mean() =

Median(Md) = \*h

First Quartile = \*h

Third Quartile = \*h

**CALULATION:**

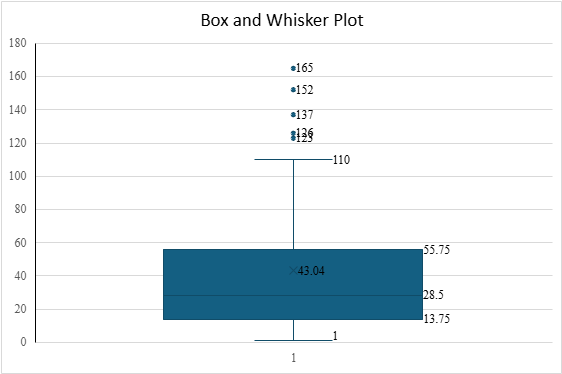
1. From Excel output:

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistics** | **Symbol** | **Value** | **Formula** |
| Mean |  | 43.04 | AVERAGE(A2:A51) |
| Median | Md | 28.5 | MEDIAN(A2:A51) |
| Mode | Mo | 27 | MODE(A2:A51) |
| First Quartile | Q1 | 15.25 | QUARTILE(A2:A51,1) |
| Third Quartile | Q3 | 53.5 | QUARTILE(A2:A51,3) |

1. From Excel output:

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistics** | **Symbol** | **Value** | **Formula** |
| Range | R | 164 | MAX(A2:A51)-MIN(A2:A51) |
| Inter-quartile Range | IQR | 38.25 | QUARTILE(A2:A51,3)-QUARTILE(A2:A51,1) |
| Variance |  | 1757.794286 | VAR.S(A2:A51) |
| Standard Deviation |  | 41.92605736 | STDEV.S(A2:A51) |
| Coefficient of Variance | cv | 97.41184331 | STDEV(A2:A51)/AVERAGE(A2:A51)\*100 |

1. Box and Whisker Plot:



Comment : The whisker of upward side is larger than that of downward side, so the data is positively skewed.

1. Given the median of 28.5 hours and the right-skewed nature of the data, I would recommend telling the president of the company that a typical customer should expect to have their complaint resolved within approximately 28.5 hours. However, it's important to acknowledge that due to the high variability and presence of outliers, some cases may take significantly longer. It's safer to say that most complaints (about 75%) are resolved within 53.5 hours.

**CONCLUSION :**

Hence, I am able to find all the required solutions of this question/topic.

**OBJECTIVE :** THE TABLE SHOWS THE NUMBER OF ABSENCES IN A STATISTICS CLASS AND THE FINAL EXAM MARKS OF CSIT STUDENTS.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NO. OF STUDENTS | 1 | 2 | 2 | 6 | 4 | 3 | 3 | 5 |
| MARKS OBTAINED | 95 | 90 | 90 | 55 | 70 | 80 | 85 | 60 |

1. FIND THE CORRELATION COEFFICIENT AND INTERPRET THE RESULTS
2. FIND THE REGRESSION EQUATION AND PREDICT THE FINAL MARKS IF NO. OF ABSENCE IS 0
3. FIND THE STANDARD ERROR OF THE ESTIMATE
4. FIND THE REGRESSION EQUATION BY DATA ANALYSIS TOOLPACK AND PREDICT THE FINAL EXAM MARKS IF THE NO. OF ABSENT IS 20
5. FIND THE COEFFICIENT OF DETERMINATION

**WORKING EXPRESSION :**

Let x = No. of absence

Y = Marks obtained

Let the regression line Y on x is ………..(i)

Where,

a = intercept

b = regression coefficient (slope)

The fitted regression model is

And Karl Pearson Correlation is

**CALCULATON :**

1. From Excel:

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Symbol** | **Value** | **Formula** |
| Correlation | r | -0.98602 | (C17\*D15-B15\*C15)/((sqrt(C17\*E15-B15^2)\*(sqrt(C17\*F15-C15^2)) |
| -0.98602 | CORREL(B7:B14,C7:C14) |

1. From Excel:

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Symbol** | **Value** | **Formula** |
| Intercept | a | 106.6667 | INTERCEPT(C7:C14,B7:B14) |
| Regression coefficient | b | -8.78205 | SLOPE(C7:C14,B7:B14) |

So, the fitted regression model is

1. From Excel,

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Symbol** | **Value** | **Formula** |
| Standard error of estimate |  | 39.99286 | STDEV.S(B7:B14,C7:C14) |

1. When x = 20,

(v) Coefficient of Determination:

From Excel,

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Symbol** | **Value** | **Formula** |
| Coefficient of determination |  | 0.972235 | D19^2 |

**CONCLUSION :**

1. Correlation coefficient: -0.9860

The results show a strong negative correlation between absences and final exam marks. As absences increase, marks tend to decrease significantly.

1. Regression equation :

Predicted marks for 0 absences: 106.6667

1. Standard error of the estimate: 2.6755
2. Predicted marks for 20 absences: -68.9744
3. Coefficient of determination (R-squared): 0.9722

**OBJECTIVE :** YOU HAVE COLLECTED DATA ON THE BATTERY LIFE (IN HOURS) OF A PARTICULAR MODEL OF LAPTOPS. THE DATA SET CONSISTS OF 20 INDIVIDUAL BATTERY LIFE MEASUREMENTS AS FOLLOWS:

7.8, 6.9, 8.2, 7.1, 6.4, 7.9, 7.3, 8.5, 7.6, 7.2, 6.5, 8.3, 7.7, 7.0, 6.8, 7.5, 8.0, 6.6, 7.4, 8.1

1. CALCULATE THE MEASURES OF CENTRAL TENDENCY (MEAN, MEDIAN, AND MODE) TO DETERMINE THE AVERAGE BATTERY LIFE OF THE LAPTOPS.
2. DETERMINE THE MEASURES OF DISPERSION (RANGE, VARIANCE, AND STANDARD DEVIATION) TO ASSESS THE VARIABILITY IN BATTERY LIFE AND UNDERSTAND HOW MUCH THEY DEVIATE FROM THE AVERAGE.
3. IDENTIFY ANY OUTLIERS OR UNUSUAL DATA POINTS IN THE BATTERY LIFE DATA THAT SIGNIFICANTLY DEVIATE FROM THE REST OF THE DATA.

**WORKING EXPRESSION :**

We have, Mean() =

Mode = Highest repeated frequency

**CALCULATION :**

1. From Excel,

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Symbol** | **Formula** | **Value** |
| Mean |  | AVERAGE(A2:A21) | 7.44 |
| Median | Md | MEDIAN(A2:A21) | 7.45 |
| Mode | Mo | None | |

1. From Excel,

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Symbol** | **Value** | **Formula** |
| Range | R | 2.1 | MAX(A2:A21)-MIN(A2:A21) |
| Variance | σ | 0.388842105 | VAR.S(A2:A21) |
| Standard Deviation |  | 0.623572053 | STDEV.S(A2:A21) |

1. From Excel,

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Value** | **Formula** |
| Upper-bound | 9.2 | QUARTILE(A2:A21,3)+1.5\*(QUARTILE(A2:A21,3)-QUARTILE(A2:A21,1)) |
| Lower-bound | 5.6 | QUARTILE(A2:A21,1)-1.5\*(QUARTILE(A2:A21,3)-QUARTILE(A2:A21,1)) |
| Outliers | None | |

**CONCLUSION :**

Hence, I am able to find all the required solutions of this question/topic.

**OBSERVATION :** WEBSITE TRAFFIC AND PAGE LOAD TIMES TEAM COLLECTED DATA ON THE WEBSITE TRAFFIC (IN NUMBER OF VISITORS) AND THE CORRESPONDING PAGE LOAD TIMES (IN SECONDS) FOR A WEBSITE OVER A PERIOD OF TIME. THE DATA SET CONSISTS OF 18 OBSERVATIONS AS FOLLOWS:

WEBSITE TRAFFIC: 1200, 1500, 1100, 1350, 1250, 1300, 1400, 1150, 1550, 1250, 1450, 1300, 1500, 1200, 1350, 1250, 1400, 1300

PAGE LOAD TIMES (IN SECONDS): 2.5, 3.2, 2.1, 2.8, 2.6, 2.7, 2.9, 2.3, 3.5, 2.6, 3.0, 2.7, 3.2, 2.4, 2.8, 2.6, 2.9, 2.7

1. CALCULATE THE CORRELATION COEFFICIENT BETWEEN THE WEBSITE TRAFFIC AND PAGE LOAD TIMES TO DETERMINE THE STRENGTH AND DIRECTION OF THE RELATIONSHIP BETWEEN THE TWO VARIABLES.
2. PERFORM A SIMPLE LINEAR REGRESSION ANALYSIS TO PREDICT PAGE LOAD TIMES BASED ON THE WEBSITE TRAFFIC. CALCULATE THE REGRESSION EQUATION, INTERPRET THE COEFFICIENTS, AND ASSESS THE GOODNESS OF FIT OF THE REGRESSION MODEL.
3. INTERPRET THE RESULTS OF THE CORRELATION AND REGRESSION ANALYSIS IN THE CONTEXT OF THE WEBSITE'S PERFORMANCE. DISCUSS ANY INSIGHTS OR OBSERVATIONS YOU CAN DRAW FROM THE ANALYSIS.

**WORKING EXPRESSION :**

Let x = Website Traffic

Y = Page load times

Let the regression line Y on x is ………..(i)

Where,

a = intercept

b = regression coefficient (slope)

The fitted regression model is

And Karl Pearson Correlation is

**CALCULATON :**

1. From Excel,

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Symbol** | **Value** | **Formula** |
| Correlation | r | 0.978600548 | (C22\*D20-B20\*C20)/(SQRT(C22\*E20-B20^2)\*SQRT(C22\*F20-C20^2)) |
| 0.978600548 | CORREL(B2:B19,C2:C19) |

1. From Excel,

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Symbol** | **Value** | **Formula** |
| Intercept | a | -0.677333333 | INTERCEPT(C2:C19,B2:B19) |
| Regression Coefficient | b | 0.002586667 | SLOPE(C2:C19,B2:B19) |

So, the fitted regression model is

iii) The predicted page load times for the minimum and maximum traffic are:

Predicted load time for minimum traffic (1100 visitors): 2.16 seconds

Predicted load time for maximum traffic (1550 visitors): 3.35 seconds

**CONCLUSION :**

1. Correlation coefficient: 0.9865

This indicates a very strong positive correlation between website traffic and page load times.

1. Simple linear regression:

Regression equation:

The regression equation shows that for every additional visitor, the page load time increases by 0.0026 seconds.

1. Interpretation:

This analysis suggests that as website traffic increases, page load times tend to increase significantly. This could indicate potential scalability issues with the website's infrastructure.

1. YOU ARE GIVEN THE INFORMATION REGARDING HUMAN POVERTY INDEX (HPI) OF SAARC COUNTRIES. SHOW THE DATA USING SIMPLE BAR DIAGRAM AND PIE CHART.

|  |  |
| --- | --- |
| **COUNTRY** | **HPI** |
| NEPAL | 50.5 |
| INDIA | 45.8 |
| BHUTAN | 20.7 |
| MALDIVES | 40.3 |
| PAKISTAN | 55.2 |
| SRI LANKA | 35.8 |
| BANGLADESH | 60.3 |
| AFGHANISTAN | 48.5 |

Solution:

Construction of simple Bar diagram and Pie chart

1. CONSTRUCT MULTIPLE BAR DIAGRAM TO REPRESENT THE DATA GIVEN BELOW:

|  |  |  |  |
| --- | --- | --- | --- |
| **YEAR** | **NO. OF PASSED STUDENTS IN** | | |
| **1 DIVISION** | **2 DIVISION** | **3 DIVISION** |
| 2002 | 20 | 60 | 80 |
| 2003 | 30 | 75 | 110 |
| 2004 | 40 | 100 | 90 |

Solution:

Construction of multiple bar diagram

1. CONSTRUCT SUB DIVIDED BAR DIAGRAM TO REPRESENT THE DATA GIVEN BELOW:

|  |  |  |
| --- | --- | --- |
| **PARTICULAR** | **COST PER SCOOTER(IN NPR)** | |
| **2005** | **2008** |
| RAW MATERIALS | 4000 | 8000 |
| LABORS | 6000 | 12000 |
| INDIRECT EXPENSES | 3000 | 7000 |
| OTHER EXPENSES | 2000 | 3000 |
| **TOTAL** | **15000** | **30000** |

Solution:

Construction of sub divided bar diagram