



# 

# STATISTICS LAB REPORT

Submitted by: Submitted to:

**NAME:** Supriya Panta **NAME**: Nanda Kumar Tharu

**SECTION**: A

# ROLL NO: 2280607 SIGNATURE

LAB REPORT INDEX-2081

Bachelor’s 2nd Semester

Name: Supriya Panta

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.N** | **Title of lab report** | **Initial date** | **Final date** | **Signature** | **Remarks** |
| 1 |  |  |  |  |  |
| 2 |  |  |  |  |  |
| 3 |  |  |  |  |  |
| 4 |  |  |  |  |  |
| 5 |  |  |  |  |  |
| 6 |  |  |  |  |  |
| 7 |  |  |  |  |  |
| 8 |  |  |  |  |  |
| 9 |  |  |  |  |  |
| 10 |  |  |  |  |  |
| 11 |  |  |  |  |  |

**Practical 1**

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 has undergone a major improvement by recruiting well-trained installation crews, supervisors and office staff. The business objective of the company was to reduce the time between when the complaint is received and when it is resolved. During the 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 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 box plot. Are the data skewed? If so, how?

d. Based on 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:**

Mean() = 

Median(Md) = \*h

First Quartile = \*h

Third Quartile = \*h

**Computation:**

|  |
| --- |
| 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 |

|  |
| --- |
| 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 |

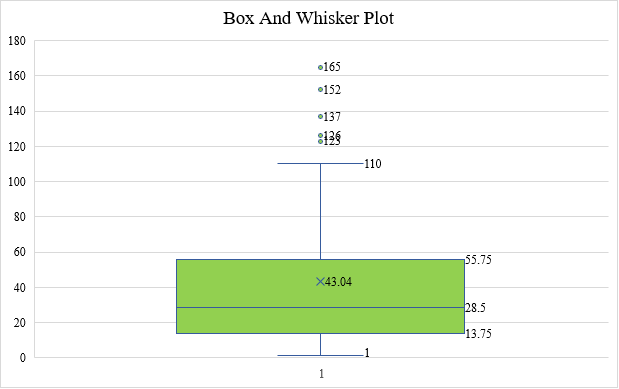
**Q.no: a**

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Symbol | Value | Formula |
| Mean |  | 43.04 | AVERAGE(B2:B51) |
| Median | Md | 28.5 | MEDIAN(B2:B51) |
| Mode | Mo | 27 | MODE(B2:B51) |
| 1st Quartile | Q1 | 15.25 | QUARTILE(B2:B51,1) |
| 3rd Quartile | Q3 | 53.5 | QUARTILE(B2:B51,3) |

**Q.no: b**

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Symbol | Value | Formula |
| Range | R | 164 | MAX(B2:B51)-MIN(B2:B51) |
| Inter Quartile Range | IQR | 38.25 | QUARTILE(B2:B51,3)-QUARTILE(B2:B51,1) |
| Variance |  | 1757.794286 | VAR.S(B2:B51) |
| Standard Deviation |  | 41.92605736 | STDEV.S(B2:B51) |
| Coefficient of Variance | CV | 97.4118433074495 | (STDEV.S(B2:B51)/AVERAGE(B2:B51))\*100 |

**Q no: c**



The whisker of the upper side is larger than that of lower side. Hence the data is positively skewed.

**Q.no: d**

Based on the result of (a) through (c), the median of 28.5 hours and the positively skewed nature of the data, I would tell the president of the company that a customer should expect to have their complaint resolved within approximately 28.5 hours. However due to the high variability and presence of outliers, some cases may take longer. It's safer to say that most complaints (about 75%) are resolved within 53.5 hours.

**Conclusion:**

Therefore, the given problem is solved.

**Practical : 2**

1**.**Construct sub-divided bar diagram to represent the data given below:

|  |  |  |
| --- | --- | --- |
| Particulars | Cost per Sector(in Rs) |  |
|  | 2005 | 2008 |
| Raw Materials | 4000 | 8000 |
| Labors | 6000 | 12000 |
| Indirect Expenses | 3000 | 7000 |
| Other Expenses | 2000 | 3000 |
| Total | 15000 | 30000 |

Sub-divided diagram is given as below:

2. 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 | 45.5 |
| Bhutan | 20.7 |
| Maldives | 40.5 |
| Pakistan | 55.2 |
| Sri Lanka | 35.8 |
| Bangladesh | 60.3 |
| Afghanistan | 48.5 |

i.Pie chart

ii. Simple bar diagram

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

|  |  |  |  |
| --- | --- | --- | --- |
| Year | No of Passed Students in | | |
|  | 1 st Division | 2nd Division | 3 rd Division |
| 2002 | 20 | 60 | 80 |
| 2003 | 30 | 75 | 110 |
| 2004 | 40 | 100 | 30 |

**Practical: 3**

The table shows the number of absences in a statistics class and the final exam 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 the number of absences is 20.
3. Find the standard error of the estimate.
4. Find the regression equation by using Data Analysis Toll Pack and predict the final exam marks if the number of absences is 20.
5. Find the coefficient of determination.

**Working Expression**:

Let X be No. of absences.

Y = Marks obtained

Let the regression line Y on x is,

Y = a + bx -------------(i)

Where ,

a = intercept

b = regression coefficient (slope)



The fitted regression model is



And Karl Pearson Correlation is

A black text on a white background

Description automatically generated

**Computation:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | Y | XY | X^2 | Y^2 |
| 1 | 95 | 95 | 1 | 9025 |
| 2 | 90 | 180 | 4 | 8100 |
| 2 | 90 | 180 | 4 | 8100 |
| 6 | 55 | 330 | 36 | 3025 |
| 4 | 70 | 280 | 16 | 4900 |
| 3 | 80 | 240 | 9 | 6400 |
| 3 | 85 | 255 | 9 | 7225 |
| 5 | 60 | 300 | 25 | 3600 |
| 26 | 625 | 1860 | 104 | 50375 |

N=8

Q.no: i

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Symbol | Value | Formula |
| Correlation | r | -0.98602 | CORREL(B7:B14,C7:C14) |

Q.no: ii

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Symbol | Value | Formula |
| Intercept | a | 106.6667 | INTERCEPT(C7:C14,B7:B14) |
| Regression Coefficient | b | -8.78205 | SLOPE(C7:C14,B7:B14) |
| No. of absences | X | 20 | Given |
| Final Exam Marks | Y | -68.9744 | P7 + P8 \*P9 |

The fitted regression model is=



Q.no: iii

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

Q.no: iv

When x = 20,





Q. no: v

Coefficient of Determination: 





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

**Conclusion:**

1. Correlation coefficient: -0.9860

There is a high degree negative correlation between absences and final exam marks which means as absences increase, marks tend to decrease .

1. Regression equation : 

a=106.6 means when number of absences is 0, then mark obtained by student is 106.6

b= -8.78 means, when the number of absences is 1, then mark obtained decreases by -8.78.

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

**Practical:4**

You have collected data of 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

a) Calculate the measures of central tendency (mean, median, and mode) to determine the average battery life of the laptops.

b) Determine the measures of dispersion (range, variance, and standard deviation) tv assess the variability in battery life and understand how much they deviate from the average.

c) Identify any outliers unusual data prints in the battery life data that significantly deviate from the rest of the data.

**Working Expression:** Mean() = 



Mode = Highest repeated frequency

**Computation:**

|  |
| --- |
| No. of battery life in hr |
| 6.4 |
| 6.5 |
| 6.6 |
| 6.8 |
| 6.9 |
| 7 |
| 7.1 |
| 7.2 |
| 7.3 |
| 7.4 |
| 7.5 |
| 7.6 |
| 7.7 |
| 7.8 |
| 7.9 |
| 8 |
| 8.1 |
| 8.2 |
| 8.3 |
| 8.5 |

N=20

Q.no: a

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Symbol | Value | Formula |
| Mean |  | 7.45 | AVERAGE(B2:B21) |
| Median | Md | 7.44 | MEDIAN(B2:B21) |
| Mode | Mo | 0.01 | MEDIAN(B2,B21)-AVERAGE(B2:B21) |

Q.no: b

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Symbol | Value | Formula |
| Range | R | 2.1 | MAX(B2:B21)-MIN(B2:B21) |
| Inter Quartile Range | IQR | 0.95 | QUARTILE(B2:B21,3)-QUARTILE(B2:B21,1) |
| Variance |  | 0.388842 | VAR.S(B2:B21) |
| Standard Deviation |  | 0.623572 | STDEV.S(B2:B21) |
| Coefficient of Variance | CV | 8.381345 | (STDEV.S(B2:B21)/AVERAGE(B2:B21))\*100 |

Q.no: c

|  |  |  |
| --- | --- | --- |
| Parameter | Value | Formula |
| Maximum | 5.55 | QUARTILE(B2:B21,3)+1.5\*((QUARTILE(B2:B21,3)-QUARTILE(B2:B21,1)) |
| Minimum | 8.4 | QUARTILE(B2:B21,1)-1.5\*((QUARTILE(B2:B21,3)-QUARTILE(B2:B21,1)) |

There is no outlier.

**Conclusion:**

Therefore, the given problem is solved.

**Practical: 5**

Website Traffic and Page Lead Times team collected data in 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

a) 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.

b) 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.

c) 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

A black text on a white background

Description automatically generated

**Computation:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | Y | XY | X^2 | Y^2 |
| 1200 | 2.5 | 3000 | 1440000 | 6.25 |
| 1500 | 3.2 | 4800 | 2250000 | 10.24 |
| 1100 | 2.1 | 2310 | 1210000 | 4.41 |
| 1350 | 2.8 | 3780 | 1822500 | 7.84 |
| 1250 | 2.6 | 3250 | 1562500 | 6.76 |
| 1300 | 2.7 | 3510 | 1690000 | 7.29 |
| 1400 | 2.9 | 4060 | 1960000 | 8.41 |
| 1150 | 2.3 | 2645 | 1322500 | 5.29 |
| 1550 | 3.5 | 5425 | 2402500 | 12.25 |
| 1250 | 2.6 | 3250 | 1562500 | 6.76 |
| 1450 | 3 | 4350 | 2102500 | 9 |
| 1300 | 2.7 | 3510 | 1690000 | 7.29 |
| 1500 | 3.2 | 4800 | 2250000 | 10.24 |
| 1200 | 2.4 | 2880 | 1440000 | 5.76 |
| 1350 | 2.8 | 3780 | 1822500 | 7.84 |
| 1250 | 2.6 | 3250 | 1562500 | 6.76 |
| 1400 | 2.9 | 4060 | 1960000 | 8.41 |
| 1300 | 2.7 | 3510 | 1690000 | 7.29 |
| 23800 | 49.5 | 66170 | 31740000 | 138.09 |

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Symbol | Value | Formula |
| Correlation | r | 0.986456 | CORREL(B39:B56,C39:C56) |

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Symbol | Value | Formula |
| Intercept | a | -0.677333333 | INTERCEPT(C39:C56,B39:B56) |
| Regression Coefficient | b | 0.002586667 | SLOPE(C39:C56,B39:B56) |

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.