



Faculty of Information Technology

Department: All

FINAL EXAM: **July**, 2022-2023 (Summer)

Course Code and Name: STAT 6122 Descriptive Statistics

Instructor: Dr. Hategekimana Fidele

Time: [Time of the exam]

Group: [Day and Evening]

Exam Duration: 3 hours

Date: [Date of the Exam]

Total max: /40

Instructions:

1. Read carefully information of the first page of the examination booklet: Don't forget to fill the reversed places with: ***your name, your student's identification number, and your group.***
2. Every student should have his/her calculator and the geometrical instruments;
3. The student is not allowed to lend neither the calculator, nor the geometric instrument from his neighbor students.

QUESTION I. (26 Marks)

The Given below are the marks obtained by 50 students in a class test. Prepare a frequency distribution with a class interval of 10. Take the first class interval as 40 – 50. Find the mean, the first quartile, the mode.

72	74	40	60	82	115	41	61	75	53	110	76	84	50	67	78	79
56	68	69	104	80	79	79	54	73	59	66	49	77	90	84	76	42
64	69	72	50	79	52	103	90	51	86	78	83	65	81	70	94	

Answer to the following questions:

1. To which type of variable (precise the subdivision) do these data belong? **(1 mark)**

Answer: Type: quantitative, Subdivision: discret

2. Explain why these data must be presented by a grouped frequency distribution **(1 mark)**

Answer: the set of data contains more than 12 different distinct values

3. What is the minimum number of class intervals needed to present these data by a grouped frequency distribution? **(1 mark)**

Answer: The set of data has the size N=50. According to Sturggs formula,

The minimum number k of class interval is given by

$$k = 1 + 3.322 \log_{10}(N) = 1 + 3.322 \log_{10}(50) = 6.64 \approx 7$$

50 values can be grouped into just 7, 8, 9, 10, 11, or 12 class intervals

4. Present the 50 marks by a stem-and-leaf chart.

(2 marks)

Answer:

The decimal point is 1 digit(s) to the right of the |

```

4 | 0129
5 | 00123469
6 | 014567899
7 | 022345667889999
8 | 0123446
9 | 004
10 | 34
11 | 05

```

5. Construct an extended grouped frequency distribution table that has 40 as the lower limit of the first class interval with class width equal to 10

(5 marks)

Answer: Extended grouped frequency distribution table

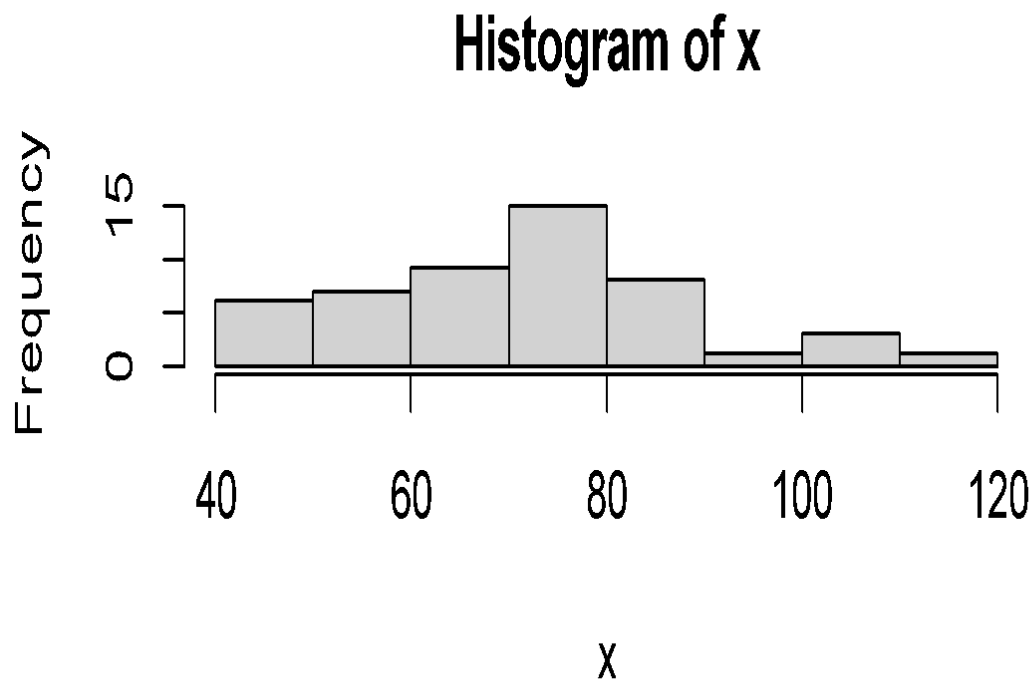
i	Classes	m_i	f_i	cf_i	$D_i = \frac{m_i - 75}{10}$	$D_i f_i$
1	40-50	45	4	4	-3	-12
2	50-60	55	8	12	-2	-16
3	60-70	65	9	21	-1	-9
4	70-80	75	15	36	0	0
5	80-90	85	7	43	1	7
6	90-100	95	3	46	2	6
7	100-110	105	2	48	3	6
8	110-120	115	2	50	4	8
$\sum_{i=1}^8$			50			-10

6. Present these marks by an histogram

(2 marks)

Answer:

Histogram



7. Find the arithmetic mean using step-deviation formula

(1.5 marks)

Answer:

$$\bar{x} = 75 + \frac{10}{50} \sum_{i=1}^8 D_i f_i = 75 + \frac{-10 \times 10}{50} = 73$$

8. Find the Mode (1.5 marks)

Answer:

Modal class interval is 70-80

$$M_0 = L + \frac{(f_1 - f_0) \times C}{(f_1 - f_0) + (f_1 - f_2)} = 70 + \frac{(15-9) \times 10}{(15-9) + (15-7)} = 74.28571$$

9. Find the quartiles (4.5 marks)

Answer:

First quartile

Class interval that contains the Q_1 is that in which fall $(N+1) \times 25\% = 51 \times 0.25 = 12.75$

Cumulative frequency; which is the class interval: 60-70

$$Q_1 = L + \left(\frac{N \times 25\% - cf}{f} \right) \times C = 60 + \left(\frac{50 \times 0.25 - 12}{9} \right) \times 10 = 60.556$$

Second quartile

Class interval that contains the Q_2 is that in which fall $(N+1) \times 50\% = 51 \times 0.5 = 25.5$

Cumulative frequency; which is the class interval: 70-80

$$Q_2 = L + \left(\frac{N \times 50\% - cf}{f} \right) \times C = 70 + \left(\frac{50 \times 0.5 - 21}{15} \right) \times 10 = 72.666$$

Second quartile

Class interval that contains the Q_3 is that in which fall $(N+1) \times 75\% = 51 \times 0.75 = 38.25$

Cumulative frequency; which is the class interval: 80-90

$$Q_3 = L + \left(\frac{N \times 75\% - cf}{f} \right) \times C = 80 + \left(\frac{50 \times 0.75 - 36}{7} \right) \times 10 = 82.14$$

10. Find the skewness and kurtosis coefficient (1 mark)

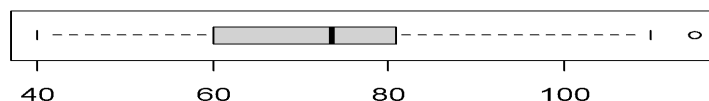
Answer:

11. Present these marks by a box-whisker plot. Do the marks have outliers?

What are they?

(4 marks)

Answer:



There is only one outlier which is 115

12. Which form does the distribution of the marks have? (Give 5 evidences supporting your answer)

(1 mark)

Answer:

The distribution has a left skewed and the following are the reasons that justify the answer;

- The stem-and-leaf diagram has more rows of large length at up the row of the modal stem (7)

- The histogram has more rectangles of higher heights at the left of the modal class interval (70-80) than those who are at its right.
- The Mode is greater than the mean and the median
- The skewness coefficient is negative
- The central vertical line inside the boxplot seems to move from the central part of the box toward the right hand side; leaving more values at its left.

QUESTION II. (14 marks)

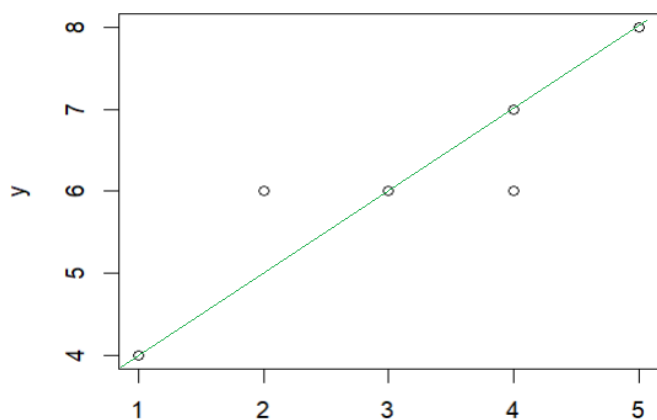
A set of bivariate data consists of these measurements on two variables, x and y expressed in the following pairs of real numbers (x, y) : (3, 6) (5, 8) (2, 6) (1, 4) (4, 7) (4, 6).

1. Draw the scatter diagram of these data. Which structure of relationship does the scatter diagram describe? (3.5 marks)

Answer:

The scatter plot predict an existence of linear relationship between x and y

Scatter Plot of y versus x



Activate Windows
Go to Settings to activate Windows.

2. Compute the covariance of the variables x and y . Give the interpretation of the calculated covariance $Cov(x, y)$ (5 marks)

Answer:

i	x	y	$X = x - \bar{x}$	$Y = y - \bar{y}$	X^2	Y^2	XY
1	3	6	-0.17	-0.17	0.03	0.03	0.03
2	5	8	1.83	1.83	3.35	3.35	3.35
3	2	6	-1.17	-1.17	1.37	0.03	0.20
4	1	4	-2.17	-2.17	4.71	4.71	4.71
5	4	7	0.83	0.83	0.69	0.69	0.69
6	4	6	0.83	-0.17	0.69	0.03	-0.14

$\sum_{i=1}^6$	19	37			10.84	8.84	8.84
----------------	----	----	--	--	-------	------	------

Mean of x: $\bar{x} = \frac{19}{6} = 3.17$

Mean of y: $\bar{y} = \frac{37}{6} = 6.17$

Covariance of x and y :

$$Cov(x, y) = \frac{1}{n-1} \sum_{i=1}^6 XY = \frac{1}{5} \times 8.84 = 1.768$$

3. Find the Karl Pearson correlation coefficient r and give its classification.

(2.5 marks)

Answer:

$$r = \frac{Cov(x,y)}{S_x S_y} = \frac{1.768}{\sqrt{\frac{10.84}{5} \times \frac{8.84}{5}}} = 0.898$$

The linear relationship is **very strong**

4. Find the regression equations and evaluate

- a. The approximate value of y when $x = 5.2$

(1.5 marks)

Answer:

$$y = bx - (\bar{y} - b\bar{x})$$

$$b = \frac{\sum XY}{\sum X^2} = \frac{8.84}{10.84} = 0.82$$

Thus,

$$y = 0.82x - (6.17 - 0.82 * 3.17)$$

$$y = 0.82x - 3.98$$

For $x = 5.2$,

$$y = 0.82(5.2) - 3.98 = 0.284$$

- b. The approximate value of x for $y = 5$

(1.5 marks)

Answer:

$$x = by - (\bar{x} - b\bar{y})$$

$$b = \frac{\Sigma XY}{\Sigma Y^2} = \frac{8.84}{8.84} = 1$$

Thus,

$$x = y - (2.67 - 1 * 6.17)$$

$$x = y + 3.5$$

For $y = 5$,

$$x = 5 + 3.5 = 8.5$$

ALL THE BEST IN JESUS CHRIST!