



Faculty of Information Technology (Undergraduate)

Final Exam, 2022/2023

Course Name: STAT 6125 Descriptive Statistics, Group: K, L, I, J, G (Day)

Instructor: Dr. HATEGEKIMANA Fidele

Exam Duration: 2:30 Hours

Date: (13/12/2022)

Question 1.

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.

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

- a. Find values of central tendency (mean, median, mode and quartiles) **(6 marks)**

Answer:

i	$a - b$	m_i	f_i	cf_i	$\frac{D_i - d_i}{10}$	$D_i f_i$	$D_i^2 f_i$
1	40 - 50	45	4	4	-3	-12	36
2	50 - 60	55	8	12	-2	-16	32
3	60 - 70	65	9	21	-1	-9	9
4	70 - 80	75	15	36	0	0	0
5	80 - 90	85	7	43	1	7	7
6	90 - 100	95	3	46	2	6	12
7	100 - 110	105	2	48	3	6	18
8	110 - 120	115	2	50	4	8	32
$\sum_{i=1}^8$			50			-10	146

The mean:
$$\bar{x} = A + \frac{C}{N} \sum_{i=1}^8 D_i f_i = 75 + \frac{10}{50}(-10) = 73$$

The median:
$$M_e = L + \frac{N * 0.5 - cf}{f} * c = 70 + \frac{25 - 21}{15} \times 10 = 72.66667$$

$$M_0 = L + \frac{|f_1 - f_0|}{|f_1 - f_0| + |f_1 - f_2|} = 70 + \frac{|15 - 9|}{|15 - 9| + |15 - 7|} = 70 + \frac{6}{6 + 8} = 74.28571$$

The mode:

$$Q_1 = L + \frac{N \times 25\% - cf}{f} \times c = 60 + \frac{12.5 - 12}{9} \times 10 = 60.5556$$

The first quantile:

$$Q_2 = M_e$$

The second quantile:

$$Q_3 = L + \frac{N \times 75\% - cf}{f} \times c = 80 + \frac{37.5 - 36}{7} \times 10 = 80 + \frac{15}{7} = 82.14286$$

The third quantile:

- b. Find the variance, standard deviation and coefficient of variation. **(3 marks)**

Answer:

$$\sigma^2 = \left(\frac{\sum_{i=1}^8 D_i^2 f_i}{N} - \left(\frac{\sum_{i=1}^8 D_i f_i}{N} \right)^2 \right) = \left(\frac{146}{50} - \left(\frac{-10}{50} \right)^2 \right) \times 10^2 = 288$$

The variance:

$$\text{The standard deviation: } \sigma = \sqrt{288} = 16.97056$$

$$\text{The coefficient of variation } CV = \frac{\sigma}{x} \times 100\% = \frac{16.97056}{73} \times 100 = 23.24735\%$$

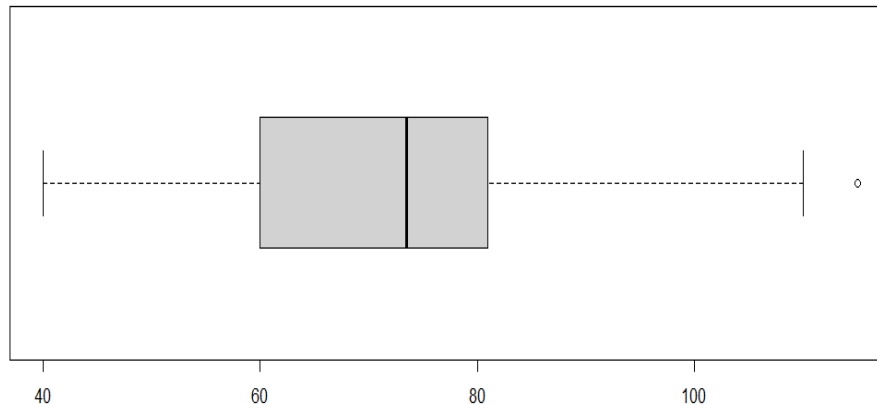
- c. Is this distribution homogeneous? **(1 mark)**

Answer:

No because, $CV < 15\%$

- d. Draw the boxplot and make a short comments of it. **(5 marks)**

Answer:



Comment: The boxplot shows the depart from the symmetric form toward the left skewed distribution form. Further the data presents one value considered as the outlier; this value is 115.

Question 2.

From the following data, obtain the correlation coefficient take assumed mean $A=5$ and $B=4$ for X and Y respectively. What is the approximate value of X when $Y=3$

X	10	6	10	6	8
Y	6	2	10	4	8

- Represent X and Y by a scatter diagram; **(5 marks)**
- Compute the linear correlation coefficient between age and blood pressure and characterise its rank; **(5 marks)**
- What is the approximate value for the blood pressure at 52 years old? **(5 marks)**

Question 3.

Let 80% of freshmen at a college take statistics, 50% take physics, and 40% take both statistics and physics.

- What is the probability that a freshman at this college takes at least one of these courses? **(2.5 marks)**

Answer:

Consider the following events:

Event A : "the student takes statistics", $p(A) = 80\% = 0.8$

Event B : "the student takes physics", $p(B) = 50\% = 0.5$

Event : "both statistics and physics", $p(A \cap B) = 40\% = 0.4$

Probability for a student to take at least one course is

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) = 0.8 + 0.5 - 0.4 = 0.9$$

- What is the probability that a freshman at this college takes only statistics? **(2.5 marks)**

Answer:

$$p(A - B) = p(A \cap \bar{B}) = p(A) - p(A \cap B) = 0.8 - 0.4 = 0.4$$

- If a freshman is known to take statistics what is the probability that he (she) takes the physics? **(2.5 marks)**

Answer:

$$p(B|A) = \frac{p(A \cap B)}{p(A)} = \frac{0.4}{0.8} = 0.5$$

- Are events taking statistics and physics independent? (Justify your answer) **(2.5 marks)**

Answer:

Yes, because $p(B|A) = 0.5 = p(B)$