Adventist University of Central Africa
P.O. Box 2461 Kigali, Rwanda | www.auca.ac.rw | info@auca.ac.rw

**Faculty of Information Techonolgy (Undergraduate)**
**Final Exam, 2022/2023**
**Course Name: STAT 8125 Descriptive Discrete Statistics, Group: E and J (Evening)**
**Instructor: Dr. HATEGEKIMANA Fidele** **Exam Duration: 2 :30 Hours**
**Date: (15/05/2023)**

---

**Question 1.**

In a city, 30 houses were surveyed as to know how many domestic appliances they have and the following were the result of the survey:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 5 | 1 | 5 | 2 | 1 | 4 | 2 | 3 |
| 4 | 2 | 4 | 3 | 2 | 6 | 3 | 2 | 4 | 3 |
| 6 | 2 | 2 | 3 | 3 | 7 | 2 | 3 | 0 | 2 |

1. What is the type of variable do the data define? In which scale of measurement is the variable measured? **(2 marks)**

   **Answer*:***
   ***Expression <u>how many</u> conveys the quantity obtained by counting and zero domestic appliance indicate the absolute zero.***
   - ***Variable: Quantitative, Discrete***
   - ***Scale of measurement: Ratio scale***

2. Construct extended frequency distribution table to represent these data and add the columns $x_i f_i$ and $x_i^2 f_i$ **(6 marks)**

   **Answer:**

| $i$ | $x_i$ | $f_i$ | $cf_i$ | $x_i f_i$ | $x_i^2 f_i$ |
|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 0 | 0 |
| 2 | 1 | 3 | 4 | 3 | 3 |
| 3 | 2 | 10 | 14 | 20 | 40 |
| 4 | 3 | 7 | 21 | 21 | 63 |
| 5 | 4 | 4 | 25 | 16 | 64 |
| 6 | 5 | 2 | 27 | 10 | 50 |
| 7 | 6 | 2 | 29 | 12 | 72 |

| | | | | | |
|---|---|---|---|---|---|
| 8 | 7 | 1 | 30 | 7 | 49 |
| $\sum_{i=1}^{8}$ | | **30** | | **89** | **341** |

3. Compute the mean $\bar{x}$, the median $M_e$ and the mode $M_0$ **(4.5 marks)**

**Answer:**

**The mean**:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{8} x_i f_i = \frac{89}{30} = 2.97$$

**The median**:

*The median, $M_e$, is the positional value which is the $(N + 1)\times 0.5 = 31\times 0.5 = 15.5$th value of the array data sorted in ascending order. The median is the medpoint of the two values in respective positions 15 and 16. By reading the positions on the column of cumulative frequency distribution, $cf_i$, the 15th and 16th values are 4 i.e. $M_e = \frac{3+3}{2} = 3$,*

**The mode:**

*The mode, $M_0$, is the most repeated value. i.e. value with highest frequency distribution. For that, $M_0 = 2$*

4. Which form of the distribution the the mode, median and mean predict? **(1 mark)**

**Answer:**

*The value of the mean, the mode and median are in the following compairison: $M_0 < \bar{x} \approx M_e$. They are predicting a **left skewed distribution**.*

5. Compute the variance, standard deviation and the coefficient of variation. **(2.5 marks)**

**Answer:**

**The variance:**

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{8} x_i^2 f_i = \frac{341}{30} - 2.97^2 = 2.55$$

**The standard deviation:**

$$\sigma = \sqrt{\sigma^2} = \sqrt{2.55} = 1.6$$

**The coefficient of variation**:

$$CV = \frac{\sigma}{\bar{x}}\% = \frac{1.6}{2.97} = 53.87\%$$

6. What does predict the value of the coefficient of variation? **(1 mark)**

**Answer:**

*The coefficient of variantion **CV** is greater than **15%** and then predicts the non-homogeneity of the distribution.*

7. Use the information of the extended frequency distribution table to find out the values of $Q_1$, $Q_2$ and $Q_3$ **(4.5 marks)**

**Answer:**

$Q_1$ is the value in position $(N+1)\times 25\% = 31\times 0.25 = 7.75$. $Q_1$ is the medpont of the values in respective positions 7 and 8; which are both equal to **2** ,

$$Q_1 = \frac{2+2}{2} = 2$$

$Q_2$ is the same as the median. i.e.

$$Q_2 = M_e = 3$$

$Q_3$ is the value in position $(N+1)\times 75\% = 31\times 0.75 = 23.25$; $Q_3$ is the medpont of the values in respective positions 23 and 24 which are both equal to **3** ,

$$Q_3 = \frac{4+4}{2} = 4$$

8. What are the value of $IQR$ and the limits of the inner fence of the boxplot. **(3 marks)**

**Answer:**

$$IQR = Q_3 - Q_1 = 4 - 2 = 2$$

*The lower limit of the inner fence:* $Q_1 - 1.5\,IQR = 2 - 1.5(2) = -1$

*The upper limit of the inner fence:* $Q_3 + 1.5\,IQR = 4 + 1.5(2) = 7$

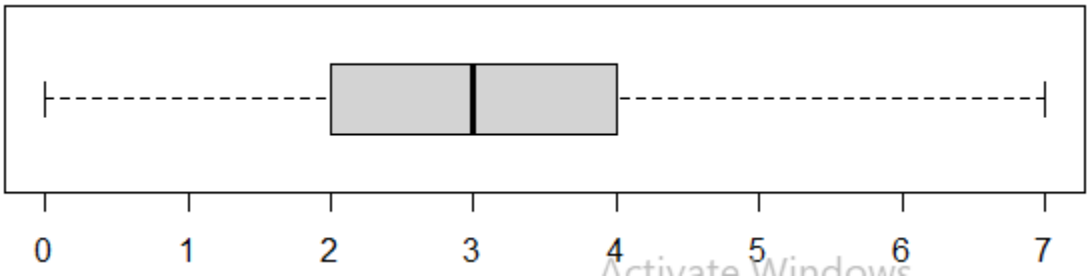
9. Does the data have outliers? If yes what are they? **(1 mark)**

**Answer:**

**No, because outliers are values outside the inner fence.**

10. Draw the boxplot and make comments about it. **(4.5 marks)**

**Answer:**




**Question 2.**

1. The following are marks of 7 students in exams of the course A and B out of 50 each:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Marks $x$ in A | 46 | 42 | 44 | 40 | 43 | 41 | 45 |
| Marks $y$ in B | 40 | 38 | 36 | 35 | 39 | 37 | 41 |

a) Calculate the covariance of the marks and conclude about the relationship between the marks of the two courses **(3 marks)**

**Answer:**

| $i$ | $x$ | $y$ | $X$ | $Y$ | $X^2$ | $Y^2$ | $XY$ |
|---|---|---|---|---|---|---|---|
| 1 | 46 | 40 | 3 | 2 | 9 | 4 | 6 |
| 2 | 42 | 38 | -1 | 0 | 1 | 0 | 0 |
| 3 | 44 | 36 | 1 | -2 | 1 | 4 | -2 |
| 4 | 40 | 35 | -3 | -3 | 9 | 9 | 9 |
| 5 | 43 | 39 | 0 | 1 | 0 | 1 | 0 |
| 6 | 41 | 37 | -2 | 1 | 4 | 1 | 2 |
| 7 | 45 | 41 | 2 | 3 | 4 | 9 | 6 |
| $\sum_{i=1}^{7}$ | **301** | **266** | | | **28** | **28** | **21** |

**<u>The Covariance</u>**:

$$Cov(x,y) = \frac{\sum_{i=1}^{7} XY}{n-1} = \frac{21}{7-1} = 3.5$$

***The two variables are not independent. The two decrease or increase in the same direction or positive coefficient of correlation.***

b) Compute the linear correlation coefficient between the two marks.
How do your rank the correlation coefficient? **(3 marks)**

**Answer:**

***The correlation coefficient:***

$$r = \frac{\sum_{i=1}^{7} XY}{\sqrt{\sum_{i=1}^{7} X^2 \times \sum_{i=1}^{7} Y^2}} = \frac{21}{\sqrt{28\times 28}} = \frac{21}{28} = 0.75$$

***Rank: High positive linear correlation coefficient.***

c) What marks should a student expect in course A when he has got 34 marks in B **( 4 marks)**

**Answer:**

*The regression equation of $x$ on $y$ is given by the formula:*

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

$$b = \frac{\Sigma XY}{\Sigma Y^2} = \frac{21}{28} = 0.75$$

*The regression equation is* $x = 0.75y + \left(\frac{301}{7} - 0.75 * \frac{266}{7}\right)$

$$x = 0.75y + 14.5$$

*For* $y = 34, \quad x = 0.75(34) + 14.5 = 40, \rightarrow\rightarrow\rightarrow x = 40$