VAV

Unit 3 Test MDM4U

(K/U)

(1)

a) The group of plants grown using water with neutral pH serves as the control group. while each ~~of the~~ other group grown using water with increasingly acidic pH levels are the experimental groups.

b) Using a group of plants for each pH level allows the botanist to observe and account for variations in individual plant response. It helps in obtaining more reliable and generalizable results by reducing the impact of outliers and providing a basis for comparison and statistical analysis.

(2)

Observation table

$N = \Sigma f = 50$

Distance (m)	frequency	Cumulative frequency
0-5	1	1
5-10	0	1+0 = 1
10-15	6	1+6 = 7
15-20	12	7+12 = 19
20-25	15	19+15 = 34
25-30	5	34+5 = 39
30-35	7	39+7 = 46
35-40	1	46+1 = 47
40-45	3	47+3 = 50

(a) The Median is given by

Median = The value of $\left(\frac{N+1}{2}\right)^{th}$ observation.

= The value of $\left(\frac{50+1}{2}\right)^{th}$ observation

= The value of $(25.5)^{th}$ observation.

The median class is the class just greater than 25.5 which means the median class is 20-25

The formula is.

$$\text{Median} = L + \frac{\left(\frac{N}{2}\right) - Cf}{f} * W$$

Where L = Lower limit of median class = 20

Cf = Cumulative frequency of previous of median class

Cf = 19

f = Frequency of median class = 15

W = Class width = 5

= Upper limit - Lower limit

$$\text{Median} = L + \frac{\left(\frac{N}{2} - Cf\right) W}{f}$$

$$= 20 + \frac{\left(\frac{50}{2} - 19\right) 5}{15}$$

$$= 20 + \frac{(25-19)\,5}{15_3}$$

$$= 20 + \frac{(6)^2}{3}$$

$$= 22$$

Median = 22

Range = Upper limit - Lower limit

= 45 - 0

Range = 45

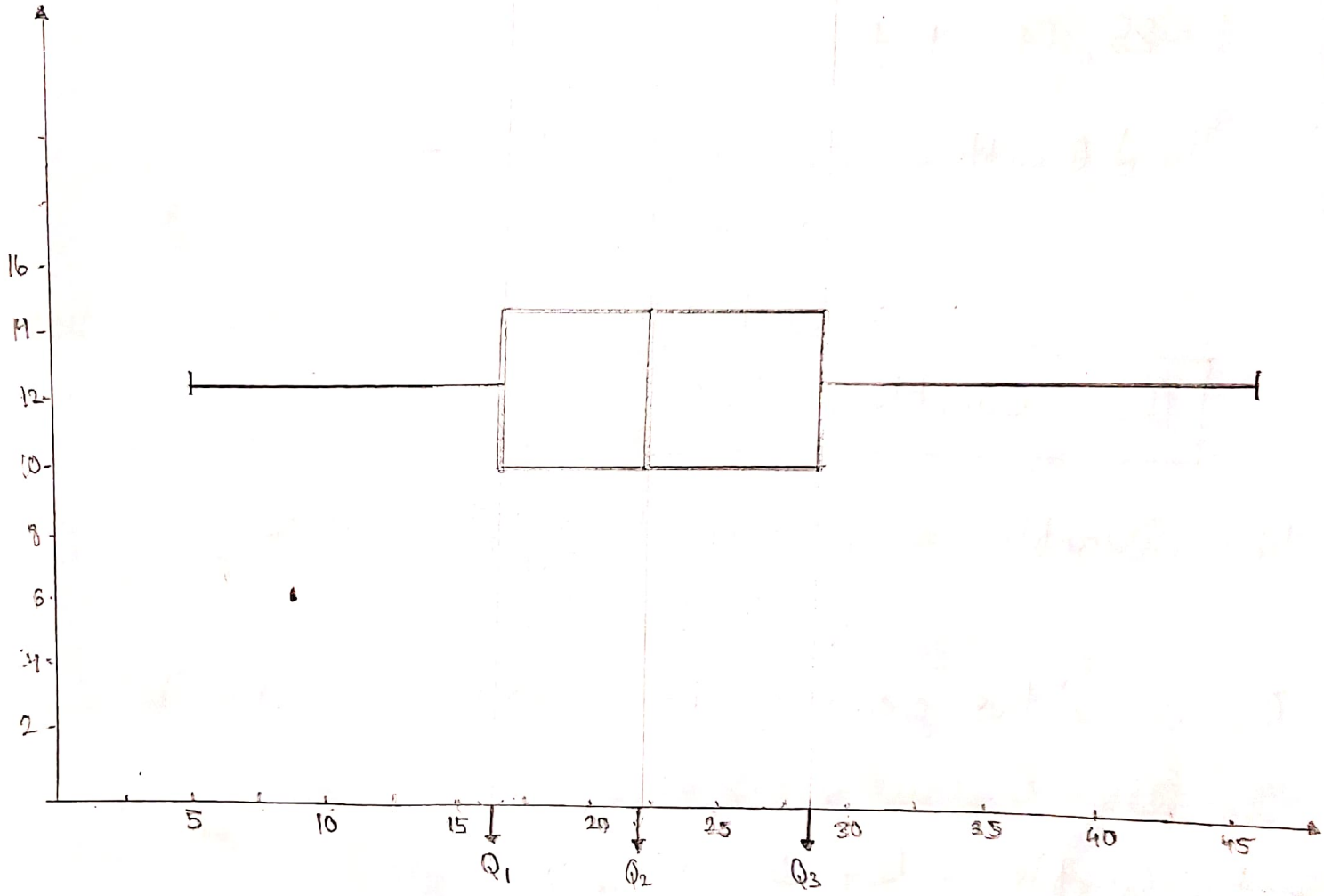
nter Quartile Range $= Q_3 - Q_1$

$= 28.5 - 17.3$

$= 11.2$

Inter Quartile Range (IQR) = 11.2

Box and Whisker Plot



(b) The whiskers would extend from 0 meters to 40 meters. This means the farthest any turtle travelled was 40m and the closest any turtle traveled was 0m.

The box cover the range of distance between 10m and 30m This suggest that most turtle traveled between 10 and 30m in 15 minutes.

(2) cont'd

Third Quartile = Q_3 = value of $3\left(\frac{N+1}{4}\right)^{th}$ Observation

$= Q_3$ = value of $(37.25)^{th}$ observation

The cumulative frequency just greater than 37.25 is 39.

The Quartile class is 25-30

$$\text{Third Quartile} = L + \frac{3\left(\frac{N}{4}\right) - Cf}{f} * W$$

$L = 25,\ Cf = 34,\ f = 5,\ W = 5$

$$\text{Third Quartile} = 25 + \frac{3\left(\frac{50}{4}\right) - 34}{5} * 5$$

$= 25 + 3.5$

Third Quartile = 28.5

First Quartile = value of $\left(\frac{N+1}{4}\right)^{th}$ observation.

$= 12.75^{th}$ observation

The cumulative frequency just greater 12.75 is 19.

The Quartile class is 15-20

$$\text{First Quartile} = L + \frac{\frac{N}{2} - CF}{f} * W$$

$$= 15 + \frac{\frac{50}{2} - 9}{12} * 5$$

$= 17.3$

First Quartile = 17.3

(K/U) cont'd

Outliers of the data

$IQR = Q_3 - Q_1$

$= 28.5 - 17.3$

$\boxed{IQR = 11.2}$

Lower limit $= Q_1 - 1.5\,IQR$

$= 17.3 - 1.5(11.2)$

$= 17.3 - 16.8$

$\boxed{\text{Lower limit} = 0.5}$

Upper limit $= Q_3 + 1.5\,IQR$

$= 28.5 + 1.5(11.2)$

$= 28.5 + 16.8$

$\boxed{\text{Upper limit} = 45.3}$

Any value outside the limits (0.5 - 45.3) is regarded as outliers

Yes, there are outliers in the dataset. They are values from the interval 0-5 and 40-45

(T/I)
(2)

a) The type of sample in this scenario is a **convenience sample**. It's convenient because you're only surveying the people on your street, which might not represent the broader population of your town.

b) The type of sample in this scenario is a **stratified random sample**. The university is dividing its student population into different strata based on department enrollment proportion and then randomly selecting students from each stratum to ensure representation from all departments.

c) The type of sample in this case is a **cluster sample**. Swim Ontario is dividing the swim clubs into clusters (groups) and then randomly selecting a few clusters (10 swim clubs) to survey all members within each selected cluster. This method is more practical than surveying all members of all 139 swim clubs.

d) The type of sample is called **random sample**. Each basketball player has an equal chance of being selected because their names are all placed in the hat and the coach randomly draws one name without ~~of~~ any bias or stratification.

e) The type of sample in this scenario is a **Systema Sample**. The store is selecting every ninth person from an alphabetical list of credit card customers after choosing the first person randomly. This method ensures a systemati and organized approach to sampling while still providi representation from the customer list.

f) The type of sample in this case is called **Voluntar responsive sample**. The student council is inviting all students to provide ideas voluntarily. This method relies on individuals choosing to participate, so the sample ma not be representative of all students, as those who partici might have different interests or ~~innovation~~ motivations.

g) The type of sample in this scenario is a **multistag cluster sample**. The marketing firm first divides the city into neighbourhoods, then selects cluster of streets within each neighbourhoods, and then finally randomly selects households with each street cluster. This method allows for efficient sampling while ensuring representation from different areas of the city.

(A)

Day	1	2	3	4	5	6	7
Temperature (°C)	27	29	32	29	45	29	31

(a) Mean

To find the mean $= \frac{\sum_{i=1}^{n} x}{n} = \frac{27+29+32+29+45+29+31}{7}$

$= \frac{222}{7}$

$= 31.714$

Mean = 31.71°c

Median

To find the median, arrange the temperature in ascending values to find the middle value

Arranged temperature :- 27, 29, 29, (29), 31, 32, 45

Median = 29°c

Mode

The mode is the temperature that appear most frequently

Mode = 29°c -> occurs 3 times

(b) Comparing the predicted temperature of 36°c with the measure of central tendency

Mean = 31.71°c , Mode = 29°c , Median = 29°c

The predicted temperature of 36°c is higher than all the measures of central tendency. This suggests that the weather report may overestimate the temperature for the day of arrival.

c) Yes, there is an outlier in the data; 45°c. An Outlier can skew the mean significantly, pulling it away from the center. In this case, it affects the mean more than the median and mode, making mean less representative of typical temperatures.

d) The median and mode would best represent temperature in this Mexican location. The median is less affected by outliers, making it a robust measure of central tendency. The mode also provides insight into the most common temperatures experienced, which is valuable for understanding typical temperature conditions.

(C)

a) The question might introduce **response bias**, particularly if respondents feel pressured to provide a higher reading speed than they actually possess.

b) This situation could involve **selection bias** if only parents with strong opinions, either for or against bus safety laws, respond to the survey, rather than a representative.

© cont'd

This could lead to non-response bias if the survey is only sent to customers who have recently interacted with the company, potentially excluding those who are le engaged or dissatisfied with the services.

) This question is prone to framing bias, as it suggests a connection between the city debt and the mayors chances of winning the next election, potentially influencing responden opinions.