PES Univeristy, Bangalore

In-Semester Assessment (ISA)

UE17CS203 – Introduction to Data Science

Quiz –1

Instructions:

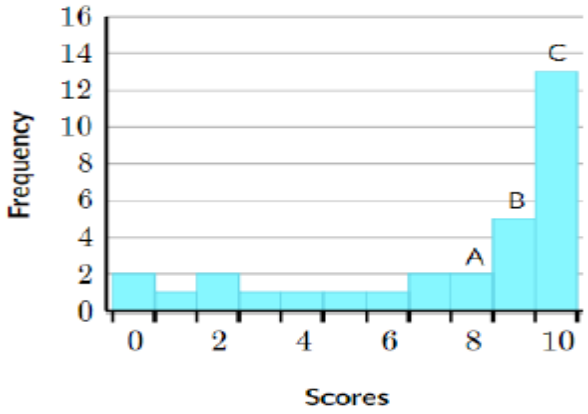
1. *Read all the questions carefully*
2. *Max Marks: 20*
3. *Time: 30 min*
4. *Make efficient use of calculator if necessary.*

1. Gupta works at a small petting zoo with 8 animals. He was looking at some data showing the masses of the animals. Each animal had a different mass between 2 and 160 kg. The zoo then buys a horse that weighs 900kg as their 9th animal. **How does buying the horse affect the mean and median?**

A. Both the mean and median will increase, but the median will increase by more than the mean.
B. Both the mean and median will decrease, but the median will decrease by more than the mean.
C. Both the mean and median will increase, but the mean will increase by more than the median.
D. Both the mean and median will decrease, but the mean will decrease by more than the median.

Answer: OPTION C

2. Graph is obtained based on agility test. Here are their scores: **(2 Marks)**



Fill in the blanks:

- The approximate location of the median is in interval ____
- The approximate location of the mean is in interval ____

Answer: B, A

Explanation: There are 31 data points in this distribution, so the median will be the 16th data point. Adding up the heights of the bars, we can find that the point is in interval B, so that is the approximate location of the median. The distribution is skewed to the left, so the mean will likely be lower than the median.

3. The following is a sample distribution of number of subscribers (in million) for all channels on a video posting platform. There are 44 channels.

No of	No of

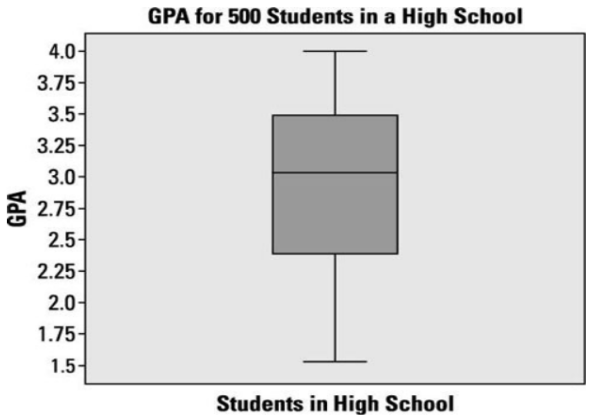
subscribers (xi)	channels (fi)
3	8
9	10
17	12
23	9
27	5

Find:
A. Mean number of subscribers B. Standard deviation

Answer:

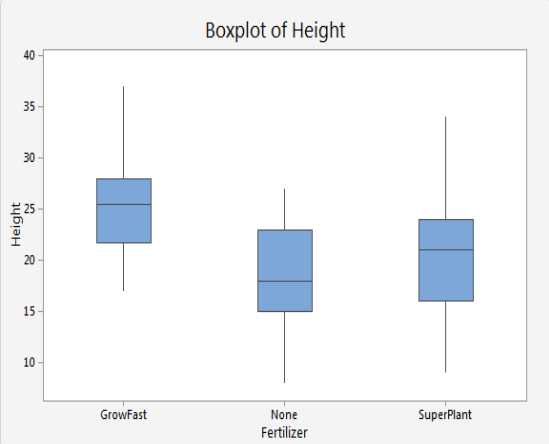
A. Mean = fixi/N = 660/44 = 15
B. Std Deviation= sqrt(2856/43)= 8.15

4. The following box plot represents data on the GPA of 500 students at a high school. (2.5 marks)



A. What is the range of GPAs in this data? **Ans:** 2.5
B. What is the median of the GPAs? **Ans:** 3.0
C. What is the IQR for this data? **Ans:** 1.125 [IQR = 3.5 – 2.375 = 1.125]
D. What is the approximate shape of the distribution of this data? **Ans:** skewed left
E. What percentage of students has a GPA below the median in this data? **Ans:** 50%

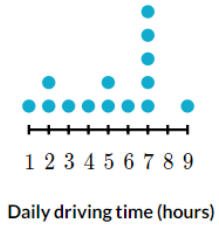
5. The following box plots represent the effect of two fertilizers- GrowFast and SuperPlant on plant height. **(2 marks)**



Which fertilizer, according to you, produces a greater and more consistent increase in plant height and why?

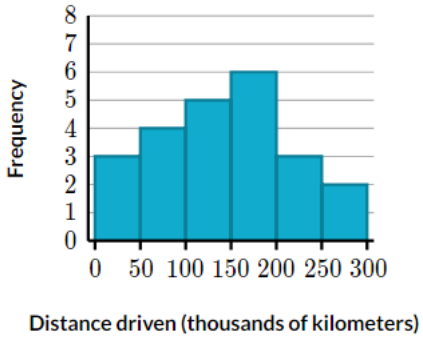
Ans: *GrowFast* produces the tallest plants overall. *SuperPlant* also increases plant height, but its variability is greater, and *SuperPlant* does not have a positive effect on a large proportion of the seedlings. The graph shows that *GrowFast* causes a greater and more consistent increase in plant height.

6. The dot plot shows the number of hours of daily driving time for 14 school bus drivers. Each dot represents a driver.



What is the closest estimate to the Percentile for the driver with a daily driving time of 6 hours?
Ans: 50th percentile

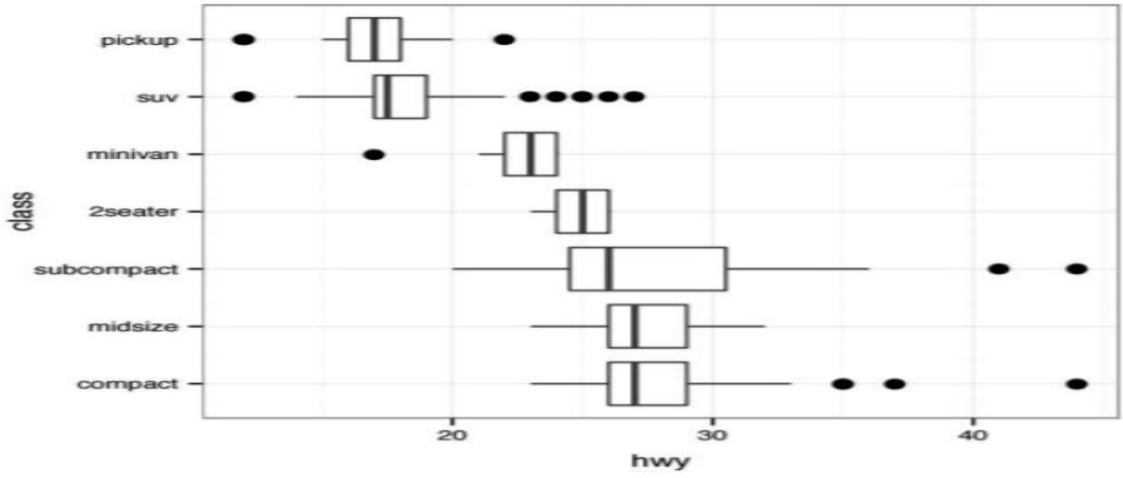
7. An auto repair shop records how far each vehicle they receive has been driven. The data for their most recent 23vehicles is shown in the histogram below.



What interval contains the 83rd percentile of this data?
Ans: 200 to 250
(83*23)/100 = 19

Falls in the 200 to 250 interval

8. The following are box plots of highway miles per gallon of each type of vehicle. Rank the vehicle classes by fuel efficiency (1 is the most efficient, 7 is the least efficient) **(2.5 Marks)**



a) Pickup: **7**
b) SUV: **6**
c) Minivan: **5**
d) 2 seater: **4**
e) Subcompact: **3**
f) Mid size: **2**
g) Compact: **1**

Explanation:
(i) Higher the median, better the highway miles per gallon
(ii) Between subcompact and minivan, the minivan has a lower median miles per gallon and an outlier with a much lower median miles per gallon
(iii) Between SUV and pickup, the SUV has several outliers that are higher

(In general, look at the median, next look at the spread (Are the two ends of the box symmetric about the median? How much do the extend about the median on either side?), next look for number and direction of outliers)

9. When each member of a population has an equally likely chance of being selected, this is called:
A. A nonrandom sampling method
B. A quota sample
C. A snowball sample
D. An Equal probability selection method
Ans: OPTION D

10. Which of the following techniques yields a simple random sample?
A. Choosing volunteers from an introductory psychology class to participate
B. Listing the individuals by ethnic group and choosing a proportion from within each ethnic group at random.
C. Numbering all the elements of a sampling frame and then using a random number table to pick cases from the table.
D. Randomly selecting schools, and then sampling everyone within the school.

Ans: OPTION C

11. Which of the following will give a more “accurate” representation of the population from which a sample has been taken?
A. A large sample based on the convenience sampling technique
B. A small sample based on simple random sampling
C. A large sample based on simple random sampling
D. A small cluster sample
Ans: OPTION C

12. People who are available, volunteer, or can be easily recruited are used in the sampling method called _____.
A. Simple random sampling
B. Cluster sampling
C. Systematic sampling
D. Convenience sampling
Ans: OPTION D

13.If central tendency is found by using whole population as input data then this is classified as

A. sample statistic

B. population statistic

C. population tendency

D. population parameters
Ans: OPTION D

14. Measure which describes detailed characteristic of whole data set is classified as

A. average or central value

B. positive skewed value

C. negative skewed value

D. positive extended value
Ans: OPTION A

15. A medical researcher wants to determine whether exercising can lower blood pressure. At a health fair, he measures the blood pressure of 100 individuals, and interviews them about their exercise habits. He divides the individuals into two categories: those whose typical level of exercise is low, and those whose level of exercise is high. Is this a controlled experiment or an observational study?
Ans: observational study