Theme 7: Statistics

Prior Knowledge

It is recommended that you revise the following topics before you start working on these questions.

- Mean, Median and Mode for grouped frequency data.
- Representation and interpretation of the cumulative frequency distribution graph.
- Finding the median of grouped data graphically from the less than type and the more than type ogive.

Statistics - A Handy Tool

Statistics pervade almost every facet of our life: from collecting data about population growth, to looking at Covid-19 numbers, to analysing Virat Kohli's "average", we use numbers to justify a whole host of opinions and crucially, also use them to make all sorts of decisions. Statistics, and its close cousin Probability, are useful in research, economic planning, collecting and interpreting public data etc. Statistics is also one of those subjects where 'bigger' or more is better! The more data you collect, the more accurate your analysis will be. Statistics can sometimes be misleading. For example, Virat Kohli's current average in Test cricket may be 50 runs per innings, however, when he goes out to bat, he is most likely to get out for a score under 20. He crosses 50 only once in 3 innings. So which statistic you use for what purpose, and to make what point, also becomes crucial.

Case Study A - Chillies per Plant

A study of the yield of 200 chilli plants is recorded by a farmer and is given in Table 7.1.

Fig. 7.1, Chilli Plant

Chillies per plant	1-5	6-10	11-15	16-20	21-25
Number of plants	15	14	75	87	9

Table 7.1, The number of plants for chillies per plant

To avoid excessive usage of chemicals in his farm, he wants to buy organic manure, only if the overall number of chillies per plant is less than 12. Analyse this data and answer the question given below.

Question 1

i. Estimate the mean and median for the given data without actual calculations?

Answer

ii. To help the farmer make his decision, which central tendency (mean, median or mode) will you find out. Calculate its actual value.

Answer

Question 2

How many plants are there that yield less than 21 chillies per plant?

a. 87	b. 200 - 9
c. 15+14+75	d. 9

Answer

Question 3

He uses organic manure only on those chilli plants which are yielding less than 5 chillies now, and with this the new modal class changes to 11-15. Which of the following statements cannot be true? (There may be more than one correct answer)

a. All the chilli plants that were giving 1-5 chillies now give 11-15 chillies

b. At least 13 chilli plants that were giving 1-5 chillies now give 11-15 chillies

c. Half the chilli plants that were giving 1-5 chillies now give 11-15 chillies

d. All the chilli plants that were giving 1-5 chillies now give 21-25 chillies

Answer

Question 4

If the farmer decides to plant more chilli plants, rather than using organic manure, to increase the mean value of crop production, then

a. Mean will further reduce if those new plants yield less than 10 chillies

b. Mean will increase just by planting more chilli plants, irrespective of chillies per plant

c. Mean is not dependent on the number of plants; it depends only on the total number.

d. Mean is not dependent on the total number of chillies; it just depends on the number of plants.

Answer

Case Study B - Cycling Race

In a marathon cycling race, the organisers had arranged resting stalls at regular intervals where the cyclists can stop for a quick bite, water, glucose, etc. Each resting stall had a person appointed to track the number of cyclists crossing that point and their speed. He had to relay back these informations to the organisers. One such resting stall, just after the toughest section of the cycling path, which is near the halfway point, was used to analyse this speed data collected. The core management team of the organisers were stationed at the starting point and they continued to plot the cumulative frequency graph of the speed of the cyclists, using a software. The team continued to analyse this graph during the race because it had an impact on the day's schedule.

Question 5

To conduct the event smoothly and in a timely manner, the management team wanted to know: (1) When will the first 3 cyclists reach the venue, so that they can start the award ceremony and (2) The time at which the last cyclist reaches, so that they can close the event.
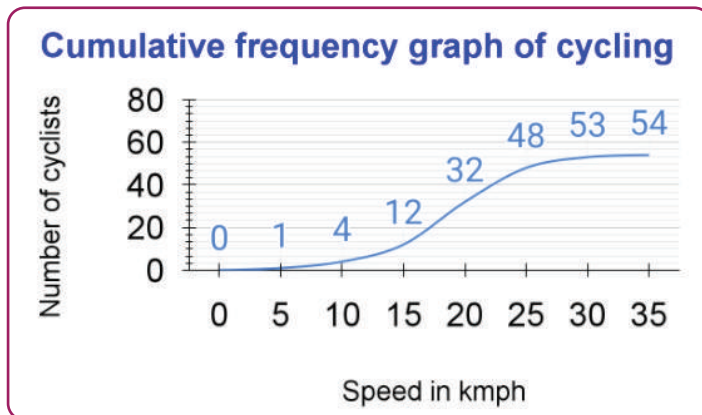

Fig. 7.2, Cumulative frequency graph of cycling speed

i. Based on the current cumulative graph (Fig. 7.2), what is the speed (in kmph) of the cyclists who will help them answer the first question? In the options given below, assume that both numbers of the range are included.

a. 48 to 54	b. 30 to 35	Answer
c. 25 to 35	d. 52 to 54	

ii. What is the speed (in kmph) of the cyclist who will help them answer the second question?

a. $0 < speed < 1$	b. $0 < speed < 5$	Answer
c. 5	d. $0 < speed < 10$	

Question 6

Find the number of cyclists who are travelling faster than 25 kmph.

a. 10	b. 48	Answer
c. 6	d. 16	

Table 7.2 lists the speed of cyclists and the number of cyclists. Find out how many cyclists are there with speeds 10-15 kmph and 20-25 kmph?

Speed of cyclist in kmph	Number of cyclist
0-5	1
5-10	3
10-15	x
15-20	20
20-25	y
25-30	5
30-35	1

Table 7.2, Speed of cyclist and the number of cyclists

Answer

Case Study C - Flight Delay

On a particular day, due to bad weather conditions, all the flights were delayed in an airport and the maximum delay time was nearly an hour. The details about the flight delay times were analysed by referring to the frequency of delayed flights as given in Table 7.3. It was noted that a total of 45 flights were delayed in the period of 6 hours after which the weather conditions came back to normal.

Time delayed (minutes)	Frequency
0-10	8
10-20	10
20-30	17
30-40	3
40-50	6
50-60	1

Table 7.3, Time delayed and frequency

Question 8

What is the average delay time, in minutes?

a. 23.22	b. 30	Answer
c. 19.22	d. 8	

Question 9

How many flights were delayed beyond half an hour?

a. 10	b. 7	Answer
c. 27	d. 8	

Question 10

What is the delay time, in minutes, which has an equal number of flights that are delayed below and above this time?

a. 17.36	b. 22.64	Answer
c. 30	d. 25	

Question 11

What is the most frequent delay time?

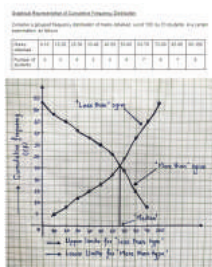| a. 55	b. 33.33	Answer
c. 17	d. 23.33	

Exploration Pathway

Ogive and Median

For the collected dataset, make a less-than type and more-than type cumulative frequency table. Choose a proper scale, based on the data. Plot the lower limit of the class interval along the x-axis and cumulative frequency along the y-axis to get a more-than type ogive. Similarly, on the same graph plot the upper limit of the class interval along the x-axis and cumulative frequency along the y-axis to get a less-than type ogive. Drop a perpendicular from the intersection point of these two graphs to read the median value. Understand the significance of the median value for the given data.