

Theme 7: Statistics



Prior Knowledge

It is recommended that you revise the following topics before you start working on this unit.

- Basics of statistics - data, tally marks, frequency, range, class interval, etc
- Plotting and reading from different graphs - bar graph, histogram and frequency polygon
- Measure of central tendency - mean, median and mode for ungrouped data



Body Mass Index

As the access to sugar and fat has become easier for humans, as a species, health concerns associated with it have increased over the last several decades. A larger awareness has emerged in the 21st century, with a whole host of diseases exacerbated by being overweight or obese.

One of the easiest - though not the most thorough - ways of knowing whether one is obese, overweight, normal or underweight is to measure one's Body Mass Index, or BMI. It's measured by a simple formula: one's mass (or weight, in kg) is divided by the square of one's height (in metres), and so is expressed in units of kg/m^2 .

Note that although BMI gives an indication of your weight status, it is not to be medically used as the basis on which to determine your risk or medication. That requires a professional medical practitioner to make a call with more thorough tests. However, BMI continues being an effective and non-invasive way for the lay person to determine whether they are overweight or not.

Case Study A - Body Mass Index

A BMI that's not normal can be an indicator to serious health risks, like heart disease, high BP, diabetes, etc. The table below categorises BMI values, based on risk, and can be used as an indicator by adults of any gender:

Weight status	BMI	Risk of developing health problems
Underweight	Below 18.5	Increased
Normal	18.5 to 24.9	Least
Overweight	25.0 to 29.9	Increased
Obese Class I	30.0 to 34.9	High
Class II	35.0 to 39.9	Very high
Class III	Above 40	Extremely high

Table 7.1, Weight, BMI range and the risk of developing health problems. Source: en.wikipedia.org - Body Mass Index (Three categories of underweight in the original table are clubbed together in order to make the context simpler).

Question 1

Researchers often use BMI at the population level to determine the general obesity levels in a certain group of people. A medical survey was conducted by a group of young dietitians to research more about body mass index and the graph in Fig. 7.1 shows the details.

[Note that all the graphs in this Case Study (Fig. 7.1 - 7.5) have been recreated using data available on the internet with minor modifications to suit the questions.]

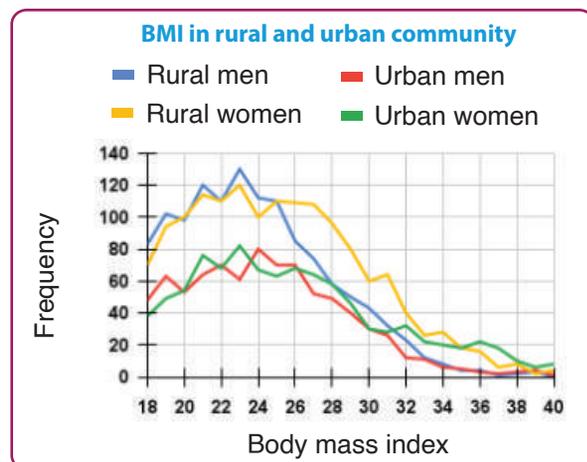


Fig. 7.1, BMI in rural and urban areas.

Which group do you think is the most obese, based on the graph shown in Fig. 7.1, with BMI ranging between 30 to 35?

a. Rural men	b. Urban men	Answer
c. Rural women	d. Urban women	

Question 2

As mentioned earlier, BMI can be used as an indicator of the health of individuals and populations. Being overweight, especially, is associated with an increased risk of numerous health issues, e.g., including type 2 diabetes, hypertension, high cholesterol, many types of cancer, etc. Based on the results of a directly measured body mass index (BMI), a graph (as shown in Fig. 7.2) was plotted for further analysis.

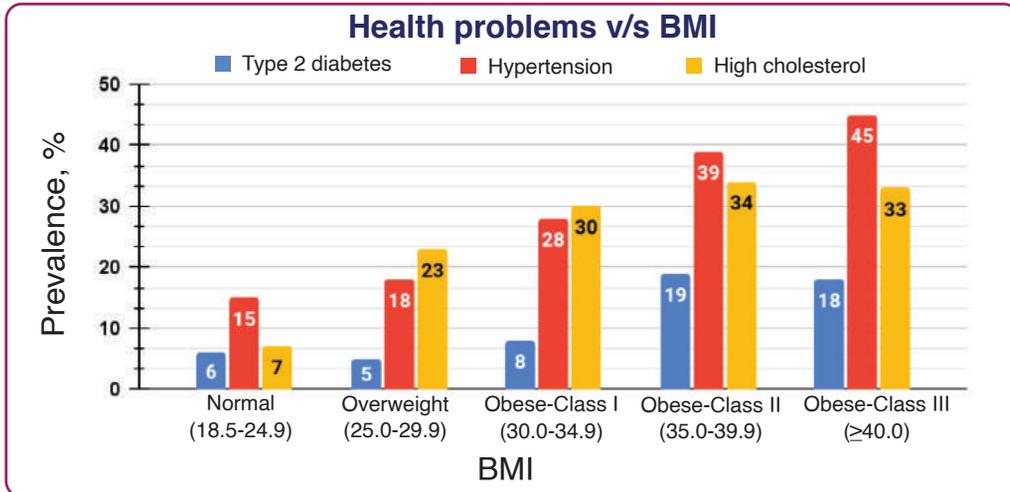


Fig. 7.2, Health problems associated with BMI.

Among Obese - Class I people, which disease is most prevalent? (Prevalence refers to the number of cases of a disease that are present in a particular population at a given time)

a. Type 2 Diabetes

b. Hypertension

c. High cholesterol

d. Both Type 2 Diabetes and Hypertension

Answer

Question 3

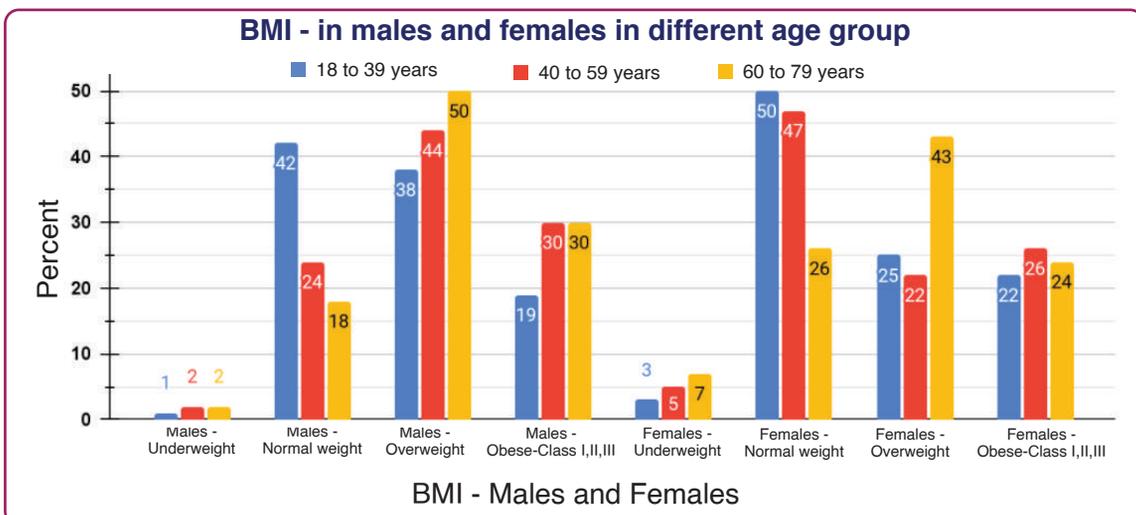


Fig. 7.3, BMI in males and females.

As per the graph in Fig. 7.3, which age and gender group has the least percentage of people in the obese category?

a. 18 to 39 years males	b. 18 to 39 years females	Answer
c. 40 to 59 years females	d. 60 to 79 years females	

Question 4

Referring to the graph in Fig. 7.3, more overweight females are found in the age group:

a. 18 to 39 years	b. 40 to 59 years	Answer
c. 60 to 79 years	d. More overweight is seen in 60 to 79 year males	

Case Study B - Climate Change

Emission of excessive amounts of carbon dioxide into the atmosphere by the past few generations is causing catastrophic and irreversible climate change. If we can make a conscientious effort to reduce the carbon footprint, future generations may have a better planet to live on. Climate Change is what will define the 21st century, and how our and the next couple of generations respond to it will define the future of life on Earth. Let us look at the carbon dioxide emission across countries worldwide over the last few decades.

Question 5

The graph given in Fig. 7.4 shows the CO₂ emission across the countries worldwide from 1960 to 2019 due to oil combustion and industrial processes.

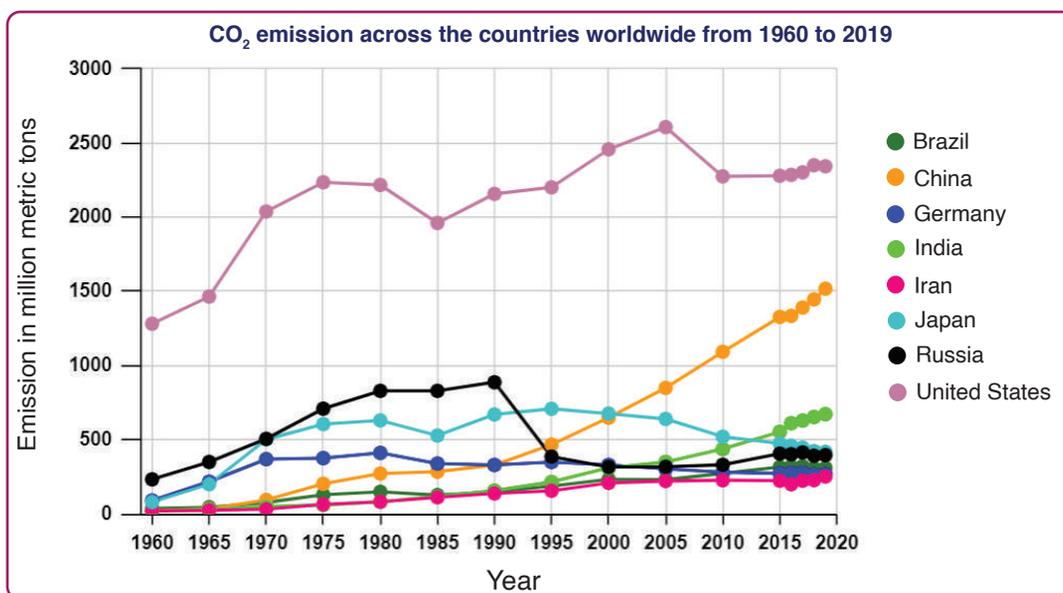


Fig. 7.4, Emissions from oil use in selected countries worldwide from 1960 to 2019 (in million metric tons of carbon dioxide).

From the graph in Fig. 7.4, three inferences are written as below:

- A. The United States is the biggest producer of emissions, and it has been in that position for a number of decades.
- B. The second biggest emitter of CO₂ is now China, which has seen a steep rise in emissions since the 1990s.
- C. Considerable decrease in CO₂ emissions is seen from Russia after the 1990s.

Which of the options below holds good about the inference drawn from the graph?

a. Only A and B are true	b. Only A is true	Answer
c. Only A and C are true	d. All 3 statements are true	

Question 6

Let us look at the data for India alone over the years. Based on the data for the year 2010 and later, India is the largest emitter of carbon dioxide, after the United States and China. The graph in Fig. 7.5 shows the CO₂ emissions per capita in India from 2001 to 2019.

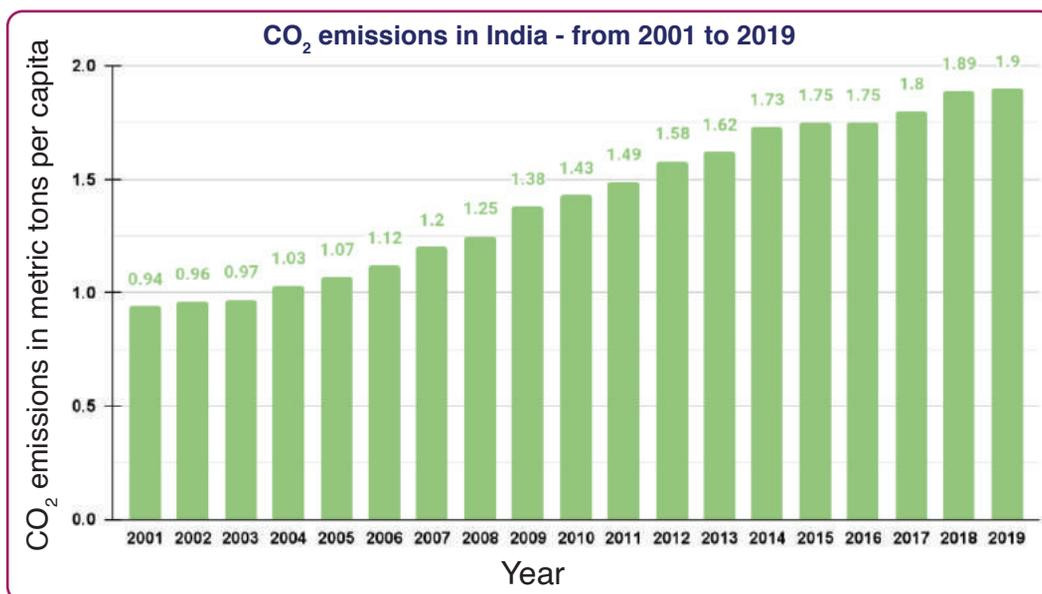


Fig. 7.5, CO₂ emissions in India.

Calculate the percentage increase in CO₂ emissions (per capita) from 2001 to 2010 (10 years) and from 2010 to 2019 (10 years). Is the percentage increase in CO₂ emissions constant, reduced or raised?

Case Study C - Runs Scored by Sachin in IPL

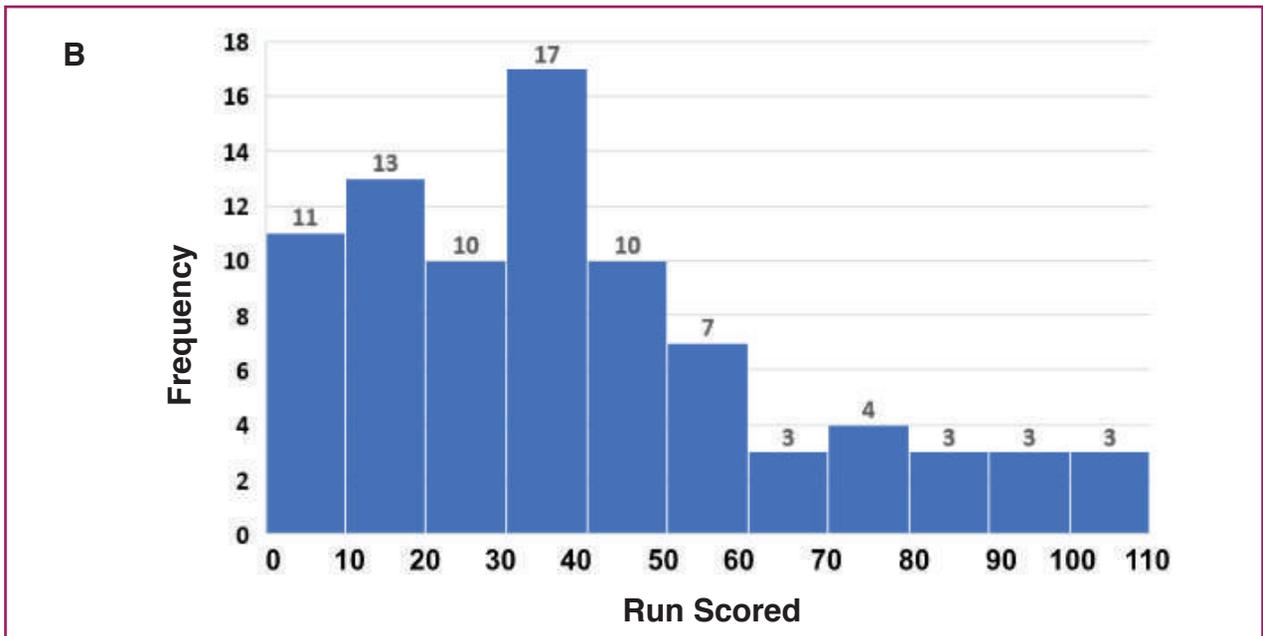
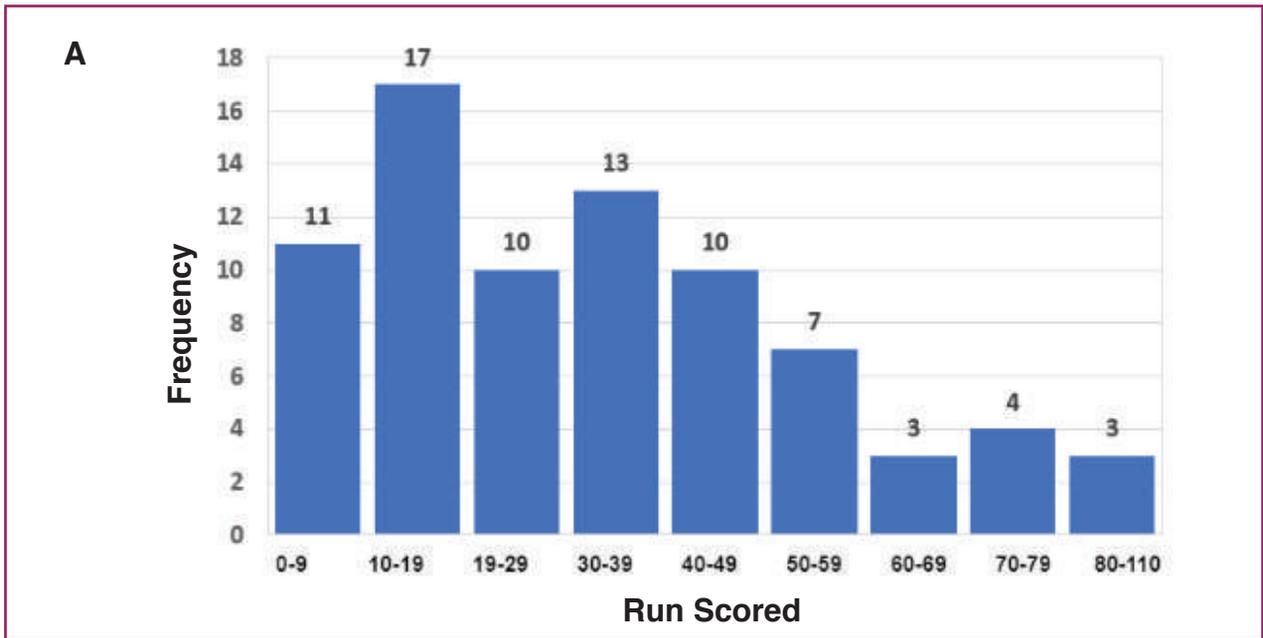
The former Indian international cricketer, Sachin Tendulkar, an all-time great, played in the Indian Premier League (IPL) from 2008 to 2013, scoring a total of 2334 runs from 78 matches that he played. Table 7.2 gives the runs scored by him grouped into the class intervals 0-10, 10-20 and so on till 110.

Class Interval (Run scored)	Frequency (Number of times)
0-10	11
10-20	17
20-30	10
30-40	13
40-50	10
50-60	7
60-70	3
70-80	4
80- above	3
Total	78

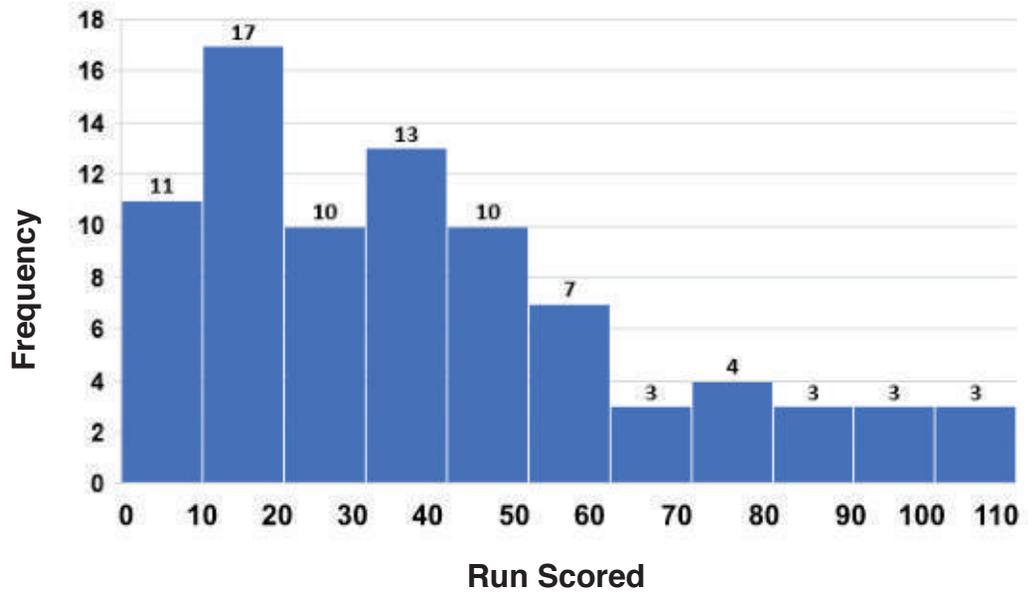
Table 7.2, Run scored by Sachin in IPL matches via howstat.com

Question 7

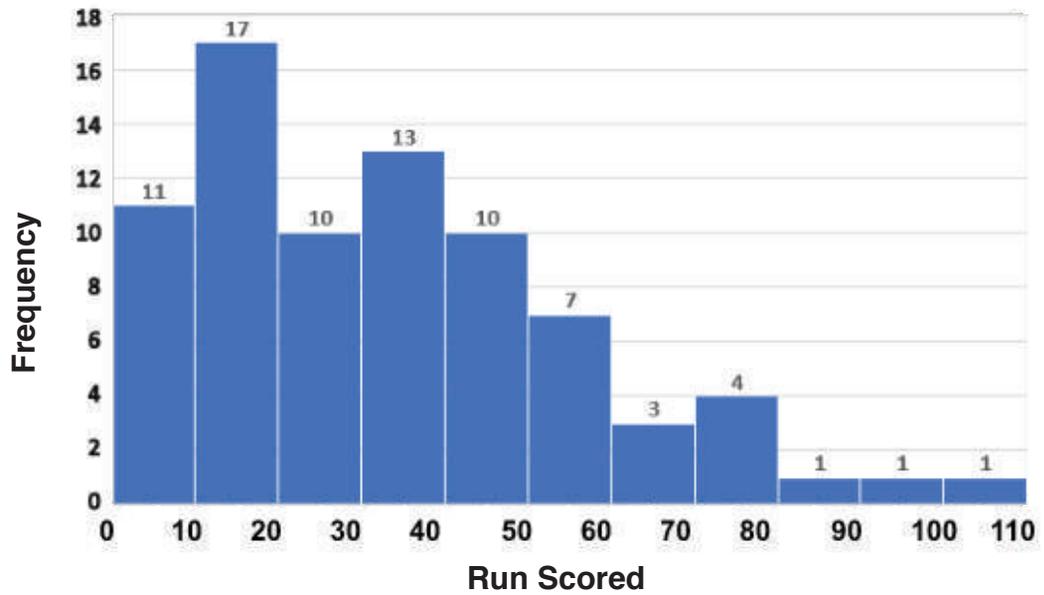
- i. A histogram was plotted to analyse the data given in Table 7.2. Identify the correct representation.



C



D



Answer

ii. Since the data given in Table 7.2 has one class interval with varying size, the corresponding length of the rectangle is calculated by using the formula,

- a. $\frac{\text{Frequency}}{\text{Width of the class}} \times \text{Minimum class size}$
- b. $\frac{\text{Width of the class}}{\text{Frequency}} \times \text{Minimum class size}$
- c. $\frac{\text{Frequency}}{\text{Width of the class}}$
- d. $\frac{\text{Frequency}}{\text{Minimum class size}}$

Answer

iii. Which other graphical representation can be considered to represent the data in Table 7.2.

- I. Bar graph
- II. Double bar graph
- III. Frequency polygon
- IV. Pie chart

Choose the correct answer.

- | | |
|-------------|--------------------|
| a. Only II | b. Both I and III |
| c. Only III | d. Both III and IV |

Answer

Question 8

i. The data for the class interval 0-10 indicate that Sachin has scored _____.

- a. 5 runs 11 times.
- b. 0 or 10 runs 11 times.
- c. 0 to 9 runs (including both) 11 times.
- d. 0 to 10 runs (excluding both) 11 times.

Answer

ii. Which of the following statements is false about the histogram plotted for the data given in Table 7.2?

- a. He most frequently scored in the range of 10 to 20 runs.
- b. He has scored zero runs 11 times and a century thrice.
- c. We cannot identify the score for any particular match.
- d. At least 50% of the time he has scored 30 or more runs.

Answer

Case Study D - Cycling Race

To spread awareness about the benefits of cycling, a famous sports club organised a cycling event open for participation in two categories. A junior category for 14 to 18 years and a senior category for 19 to 25 years. The graph in Fig. 7.6 gives the data obtained from the timings of the two categories.

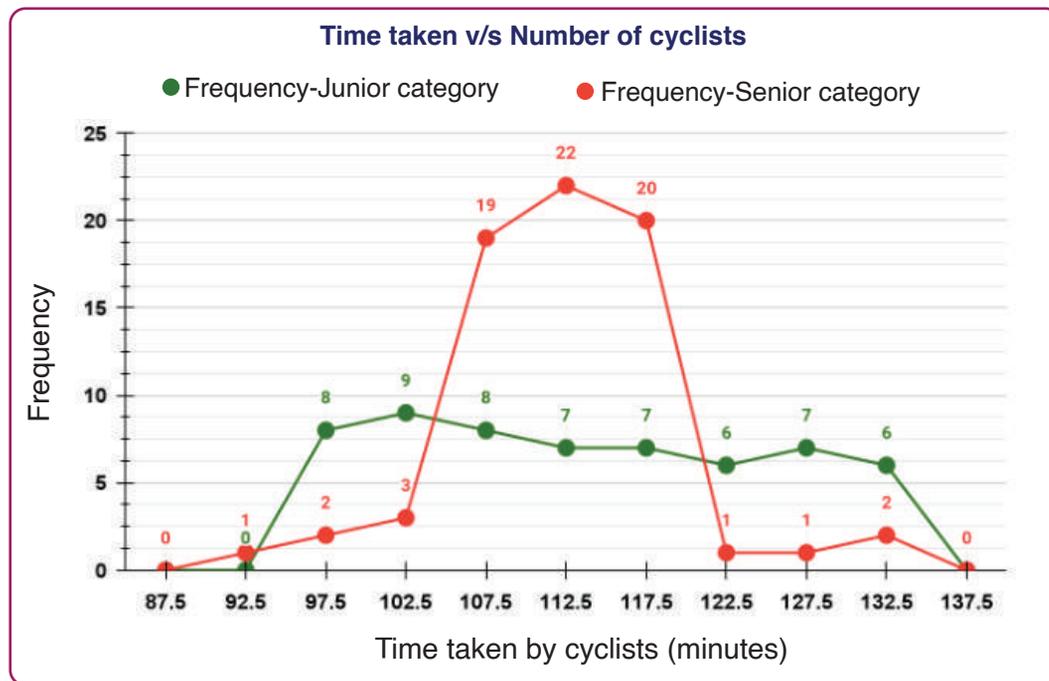


Fig. 7.6, Frequency graph of time taken by cyclists

Question 9

Look at the graph in Fig. 7.6 and validate the assertion statements made about this graph.

Assertion 1 (A1): Given the age of a cyclist, the time taken by her/him to complete the race can be predicted more easily for the senior category than for the junior category, since a majority of the seniors complete the race within 107.5 to 117.5 minutes.

Assertion 2 (A2): Frequency polygons are used when the data is continuous and large. They are useful to compare two different sets of data of the same nature.

- a. A1 is true, but A2 is false
- b. A1 is false, but A2 is true
- c. Both A1 and A2 are false
- d. Both A1 and A2 are true

Answer

Question 10

As per the graph in Fig. 7.6, which of the statements below is wrongly stated?

- a. There are more participants in the junior category than in the senior category
- b. There are a total of 9 class intervals
- c. Size of each class interval is 5 minutes
- d. Both juniors and seniors category have three class intervals in each, which have the same number of participants from their respective group.

Answer

Question 11

Apply your knowledge about mean, median and mode to analyse each of the statements below. Write YES if it is the correct measure for the given situation and write NO if it is the wrong measure. Justify your answer in each case.

- i. A swimming pool has a mean water depth of 130 cm. A boy who is 140 cm tall can stand safely anywhere in this pool.

Answer

- ii. To know the central tendency related to salaries in a factory, calculate the mean salary of all workers, including the least paid employee and the most paid employee. (10% of the factory employees are managers and supervisors and most of them have salaries which are 5 times more than the highest paid from the remaining 90%.)

Answer

iii. To know the usual price with apartment rates in a city, ask the broker for a modal value from the price list.

Answer

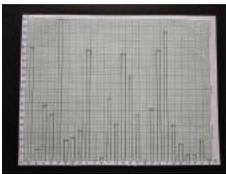
iv. Plan your monthly budget, listing the expenditure and finding its mode (E.g.: ₹350 water bill; ₹900 electricity bill; ₹12000 house rent so on...).

Answer

v. A shoe seller places an order for the required number of shoes of each size, based on the modal value of shoe size from his sales data.

Answer

Exploration Pathway



Letter Count Bar Chart

Cryptanalysis: Take a paragraph with a minimum of 100 words and create a frequency table for the letters in the paragraph. Plot a bar graph and identify the most frequently occurring letter and the least occurring letter.

