O Level E Maths Tutorial 17: Data handling and analysis

Syllabus :
• simple concepts in collecting, classifying and tabulating data
• mean, mode and median as measures of central tendency for a set of data

---

1. Mean, mode and median are 3 ways to estimate a "centre value" of a group of numbers, like the age or height of a group of children. The value in each case may be different.

24 33 15 19 41 33 12 33

For this set of numbers, find the:

(i) mean
(ii) median
(iii) mode
(iv) range

- tables
- pie charts

---

2. The table shows how a group of people go to work.

	Frequency	Sector angle
Walk	23	
Bus	20	
Cycle	8	

Draw a pie chart for this. Find the sector angles in each case.

- stem-and-leaf diagrams
• drawing simple inference from statistical diagrams

---

3. The time, to the nearest minutes, that each of 23 students take to go to school in one day was recorded. The results are shown on the stem-and-leaf diagram.

7	5 6 8 9 9 9
8	0 0 1 1 2 3 4 7
9	3 4 4 5 6 8
10	0 1 2

Key: 7 | 5 represents 75 minutes

Figure 17-1

(a) Write down the median of the times.

(b) When the time for a 24th student is added to the diagram, the range increases to 33 minutes. Find the two possible times that this student took to go to school.

4. The masses, in grams of apples and bananas in a box are recorded. The results are given in this stem-and-leaf diagram.

Apples					Bananas			
		5	2	9				
8	2	1	0	10	3	3	5	5
	9	9	3	15	0	8	x	

Key (Apples)
2 | 9 means 92 grams

Key (Bananas)
10| 3 means 103 grams

Figure 17-2

(a) What is the median mass of the apples?

(b) The range of masses for the bananas is 45 g. Find x.

(c) Make two comments comparing the masses of he apples and the bananas.

• explaining why a given statistical diagram leads to misinterpretation of data

---

5.

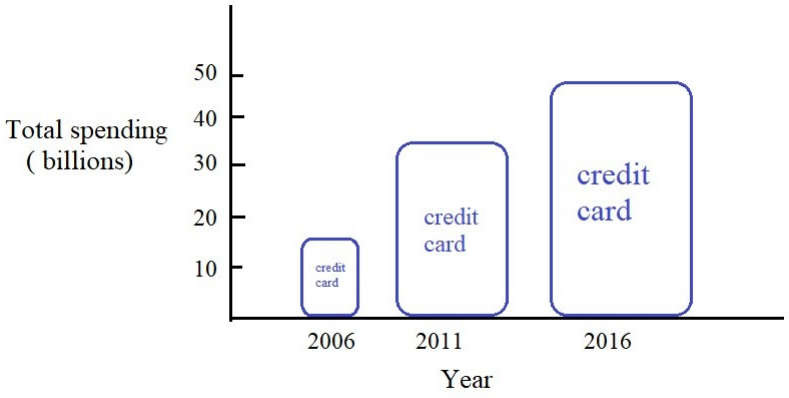


Figure 17-3

Explain how the chart above may be misleading.

- histograms with equal class intervals
- cumulative frequency diagrams
- range, interquartile range and standard deviation as measures of spread for a set of data
- calculation of the standard deviation for a set of data (grouped and ungrouped)

6.(a)

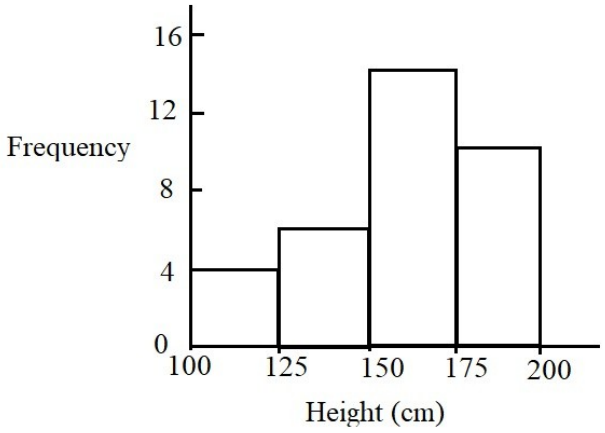


Figure 17-4

The histogram shows the distribution of the heights of a sample of trees.

(i) What is the total number of trees in the sample? What is the range of the heights?

(ii) Find the mean height of the trees.

(b) (i) Using the above histogram, tabulate the cumulative frequencies and plot a graph against height.

(ii) Use the graph to estimate the number of trees over 190 cm height.

(iii) Find the interquartile range of the heights.

(c) Standard deviation is number that tells us roughly how spread out most of the data is. It is given by this formula

$$s^2 = \frac{1}{n-1}\sum (x - \bar{x})^2$$

where s is the standard deviation, n the number

Calculate an estimate of the standard deviation. For each interval in the histogram, use the mid value of the height. E.g. for interval 100 to 125, use 112.5 cm.

[Note: There is a function in scientific calculators to calculate the standard deviation.]

- box-and-whisker plots

---

7. The speeds of 80 cars passing a checkpoint one morning were recorded.
The cumulative frequency curve below shows the distribution of the speeds.

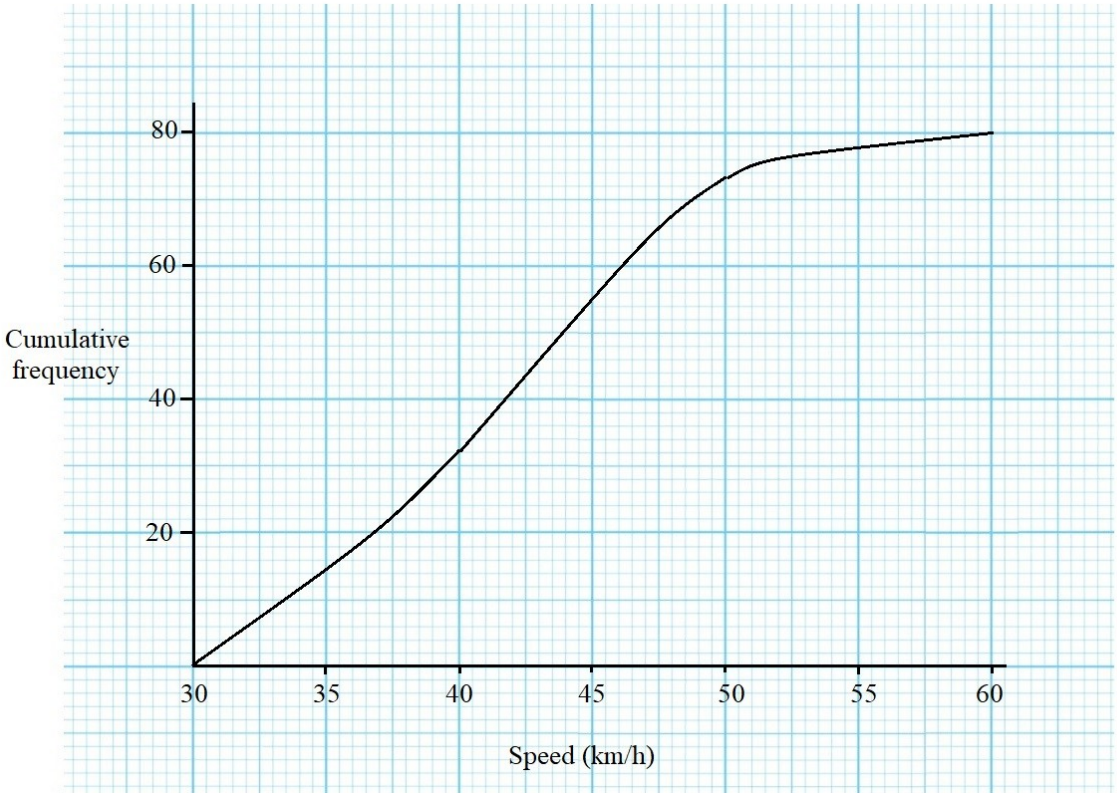


Figure 17-5

(i) Use the curve to estimate
(a) the median speed,
(b) the interquartile range of speeds.

(ii) The speed limit on the road is 50 km/h.
Estimate the percentage of the cars that exceeded the speed limit.

(iii) The speeds of 80 cars passing the same checkpoint in the afternoon were also recorded.
The box-and-whisker plot shows the distribution of the speeds.

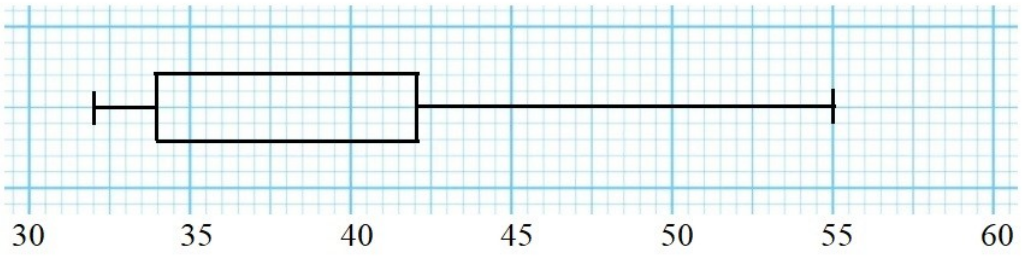


Figure 17-6

Make two comments comparing the speeds of the cars in the morning and in the afternoon.

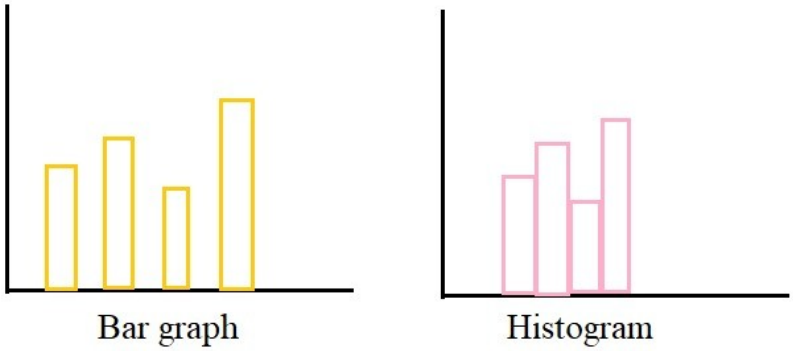
[N17/II/9(a)]

- bar graphs -
- pictograms -
- line graphs
- pie charts
- dot diagrams -

Notes on bar graphs, pictograms, line graphs, dot diagrams

Bar graph and histogram difference

AI Overview
Learn more
A bar graph displays data by comparing values of different categories, while a histogram shows the distribution of numerical data within specific ranges or bins. Bar graphs use discrete categories on the x-axis, with bars representing the values for each category, while histograms use continuous numerical ranges on the x-axis, with bars showing the frequency of data within those ranges.



Here's a more detailed breakdown:

Bar Graph:

- Purpose: To compare values across different categories or groups.

- **X-axis:** Represents discrete categories or groups (e.g., different brands, countries, or months).
- Y-axis**:** Represents the value being compared for each category (e.g., sales figures, population, or temperature).
- Bars**:** Have spaces between them, emphasizing that each bar represents a separate category.
- Example**:** Comparing the sales of different products in a store.

Histogram:

- **Purpose:**

  To visualize the distribution of a continuous variable, showing how frequently data points fall within specific ranges or bins.

- **X-axis:**

  Represents continuous numerical data, typically divided into intervals or bins.

- **Y-axis:**

  Represents the frequency or count of data points that fall within each bin.

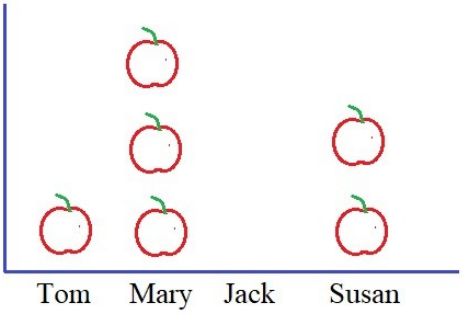
- **Bars:**

  Are adjacent, without spaces, to emphasize the continuity of the data.

- **Example:**
  Showing the distribution of test scores in a classroom, grouped into ranges like 0-20, 20-40, etc.

Pictogram

A pictogram is a graphical symbol, icon, or picture that represents a concept, word, instruction, or data. They are used in various forms, including visual communication, signs, and labels, and also as a type of graph to represent data. Pictograms are designed to convey information quickly and effectively.



Key aspects of pictograms:

- **Visual Representation:**

  Pictograms rely on images or symbols to represent ideas or data instead of words or numbers.

- **Data Representation:**

  In the context of graphs, pictograms use images or symbols to show the relationship between different sets of data, similar to a bar chart but using pictures instead of bars.

- **Meaningful Symbols:**

  Each symbol or picture in a pictogram has a defined meaning, often indicated by a key that explains what the symbol represents.

- **Examples:**

  Common examples include hazard pictograms on chemical containers, signs in public places like "no smoking" or "no diving", and pictograms used in educational materials to help children understand concepts.

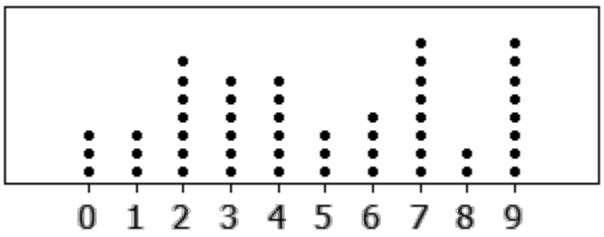
- **Purpose:**
  Pictograms are used for various purposes, including communicating warnings, instructions, or representing data in a visually appealing and easily understandable way.

  For example, a pictogram could show the number of students who chose different flavors of ice cream, with each ice cream picture representing a certain number of students.

Dot Diagram

AI Overview

Learn more



A dot diagram, also known as a dot plot, is a simple statistical chart that represents data points as dots along a number line. Each dot represents a single data point, and if there are multiple data points with the same value, the dots are stacked on top of each other. Dot diagrams are particularly useful for visualizing the distribution of small to medium-sized datasets and for identifying clusters, gaps, and outliers.

Here's a more detailed explanation:

- **Data Representation:**

  Each dot on the diagram represents a specific data point or value.

- **Number Line:**

  The data values are plotted along a horizontal number line, which serves as the axis.

- **Frequency:**

  The number of dots stacked above a particular number on the number line indicates the frequency or how often that value appears in the data set.

- **Data Distribution:**

  Dot diagrams help visualize the spread and shape of the data, allowing for easy identification of patterns such as clusters, gaps, and skewness.

- **Outliers:**

Values that are significantly different from other values in the data set can be easily spotted as outliers on the dot diagram**.**

- **Small Datasets:**
  Dot diagrams are best suited for visualizing small to medium-sized datasets, as they become crowded and less readable for large datasets.