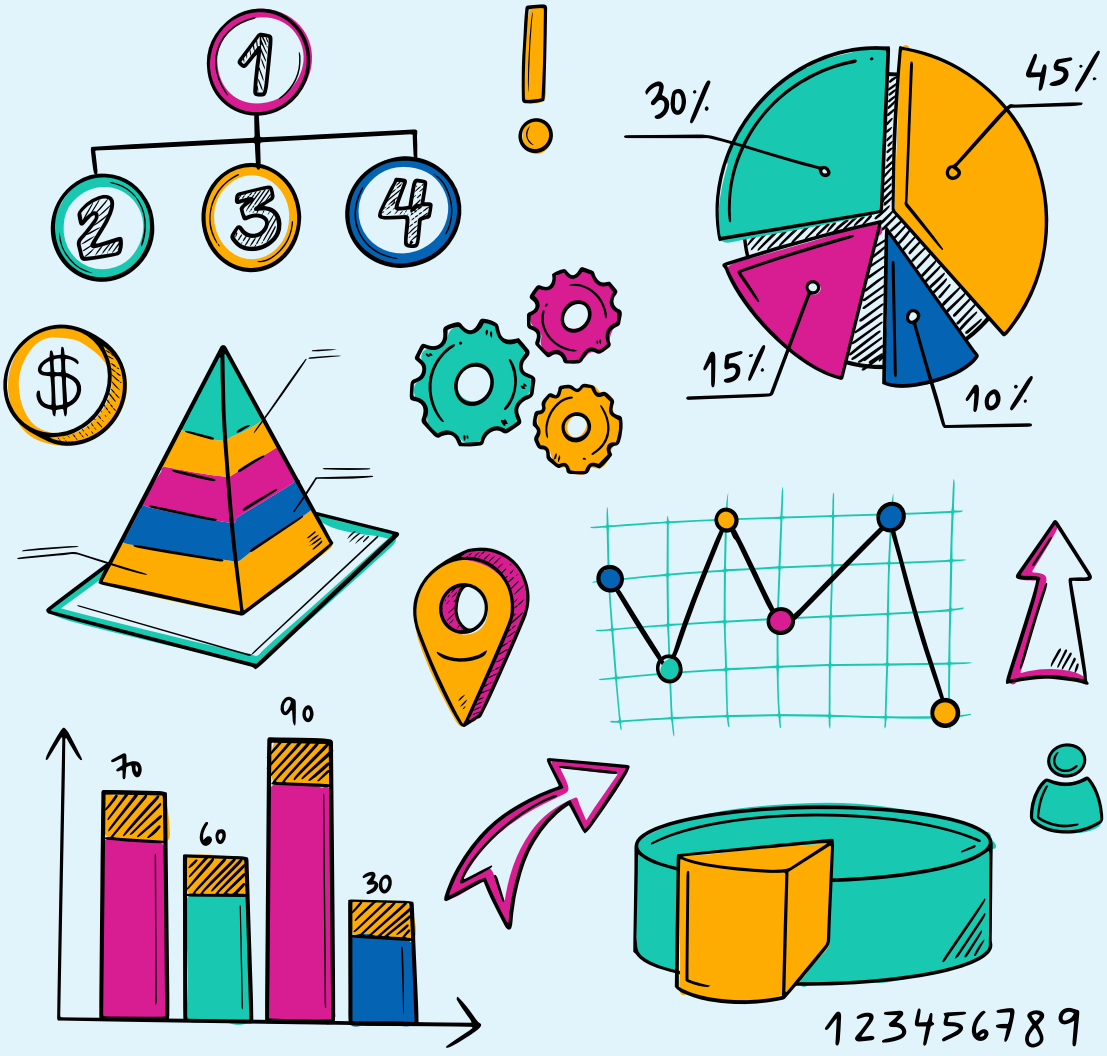
SECTION

8

DATA ORGANISATION, ANALYSIS AND PRESENTATION



MAKING SENSE OF AND USING DATA

Statistical Reasoning and Its Application in Real Life

INTRODUCTION

Data is important in today's world and it comes in different types like numbers, categories and rankings. We can collect and organise data in many ways, then illustrate it using graphs, charts or tables to make it easier to understand. To summarise data we use the mean, median and mode, which give us one number that shows the "average" of the data. These numbers help us compare and make conclusions. Doing real-life projects with data helps us learn how to collect, analyse and present information. These projects can be connected to different subjects, like science, history, geography and economics. Learning these skills helps us make smart decisions and prepares us for more advanced studies in areas like statistics and data science. In this section, we will explore the collection and study of data in order to get facts or information. This process is termed 'Data collection' or 'collection of data'. In JHS, you have learnt some data types and how to analyse them, so let us do a quick recall of some of the things you have learnt on data and data types.

At the end of this section, you will be able to:

- Classify data (primary and secondary) as quantitative (discrete and continuous), qualitative (nominal and ordinal), numerical, categorical, grouped, ungrouped, etc.
- Identify and validate quantitative data collection methods (Survey/ Questionnaire, Interviews, Observation, Existing Data, and Probability) and use it to collect everyday- life data.
- Identify and validate qualitative data collection methods (interviews, observations, s groups, oral histories, online tracking, social media monitoring, etc.) and use it to collect everyday-life data.
- Organise and present data (grouped/ungrouped) using frequency tables, line graphs, pie charts, multiple bar graphs, info graphics, etc., including generating 3D graphs/ charts with appropriate digital technology (where available) and solve problems on them.

- Analyse (include using appropriate computer applications) and interpret data using descriptive statistics (i.e., measures of central tendency/ location and minimum & maximum values) and justify which of the averages best represent the data.
- Use mathematical arguments to support personal choices as well as incorporate the views and perspectives of others to assess and make inferences of data presented in everyday life (including class discussions, school debates, textbooks, school clubs, etc.)
- Develop and execute a project with a team by collecting and analysing data within the school environment and give useful conclusions and recommendations (including the use of appropriate computer applications, e.g., excel).
- Present a project report to your class or at a school forum. Include the use of presentation software such as power point, infographics, etc., and publish the report in a school magazine, school notice board, school social media platforms

Key Ideas

- A variable is a characteristic or attribute that can take on different values or levels.
- Data collection is gathering data through surveys, experiments, or observations.
- Data analysis means using statistical methods to summarise, visualise and model data.
- Frequency refers to the number of times a particular value or category occurs in a dataset.
- A frequency distribution table is a table that displays the frequency of each value or category in a dataset.
- An interval refers to a range of values between two specific limits, inclusive or exclusive.
- Inference means drawing conclusions about a population based on a sample of data.
- Tally involves marking vertical lines for each occurrence of a value.

WHAT IS DATA?

Data refers to the collection/gathering of pieces of information or facts in order to analyse them, make decisions or gain insights. Data can be in various forms, such as numbers, text, images or audio. In a more technical sense, data is often raw and unprocessed (ie, it has yet to be organised or analysed) to reveal patterns or meaning. In short, data is information gathered from various sources through observation or from results of an experiment.

For example:

- Numbers like 24, 55 and 12 are data points that can represent ages, temperatures, or other quantities.
- A list of customer reviews and their ratings is data that can be analysed to understand customer satisfaction.

Data becomes more useful when it is processed and interpreted to generate information and knowledge. In fields like statistics, data science and information technology, data plays a crucial role in drawing conclusions and making informed decisions.

Let us now consider some types of data

Types of Data

Primary and secondary data are two fundamental types of data used in research and analysis. They serve different purposes and have distinct characteristics, each contributing to the understanding of a given subject.

Primary data

Primary data refers to information collected first-hand, specifically for a particular research study or investigation. This data is original and directly obtained from the source. Primary data collection methods include surveys, interviews, experiments, observations or direct measurements.

Secondary data

Secondary data refers to existing data that has been collected by someone else for a different purpose but can be repurposed for new research or analysis. Secondary data sources include published literature, reports, databases, government records and data collected by other researchers.

Grouped and Ungrouped Data

When working with data, it can be classified as either grouped or ungrouped, depending on how the values are organised and presented.

Ungrouped data is a raw set of individual observations or values without any categorisation or grouping. Each data point is unique and stands alone. For instance, if you record the heights of a group of individuals without categorising them into ranges, you would have ungrouped data.

Grouped data involves categorising values into intervals or classes, creating a summary of data within each category. Grouping data helps simplify large data sets and provides a broader overview of the distribution.

Example 1

Height (in inches) – the raw data: 62, 65, 67, 62, 70, 65, 66, 64, 67, 71

Table 1: Ungrouped Frequency Distribution

Height (in inches)	Frequency
62	2
64	1
65	2
66	1
67	2
70	1
71	1

In the above frequency table, each data point is unique and stands alone, this makes it ungrouped.

Table 2: Grouped Frequency Distribution

Height Range (in inches)	Frequency
60 - 64	3
65 - 69	5
70 - 74	2

In the table above, the heights have been grouped into three categories: 60-64 inches, 65-69 inches and 70-74 inches. The frequency column represents the

number of observations falling within each category. This means the data is grouped.

Let us now look at and draw out the differences, with examples, between discrete and continuous data.

Discrete and Continuous Data

Discrete data

This typically consists of whole numbers or counts of items and does not include decimal values. However, in certain cases, discrete data may be represented as decimal values for practical purposes. It also represents items that can be counted and each item is unique and cannot be divided into smaller parts.

Discrete data includes discrete variables that are finite, numeric, countable and non-negative integers and rarely involve decimals.

E.g.:

- The number of students who have attended the class.
- The number of customers who have bought different products.
- The number of groceries people are purchasing every day.
- Population of a country

For instance, consider a situation where you are measuring the number of defective items produced in a factory per hour. While the number of defects is a discrete variable (you cannot have half a defect), it might be represented as a decimal to indicate the average number of defects per hour over time. In such cases, the decimal values are used for measurement or calculation purposes, but the actual data points remain discrete. Another example is that in the UK, shoe sizes come in half sizes. But nothing in between. For example, 5, 5.5, 6, 6.5 etc. – therefore, this is discrete data.

Continuous data

This is the unspecified number of possible measurements between two presumed points. Continuous data is a type of data that can take any value within a certain range or interval, including fractions and decimals. It represents measurable values that can be divided into smaller parts and can be expressed on a continuous scale. For example, time can be measured in hours, minutes, seconds and fractions of seconds, such as 12:30:45.123. Also, a person's height can be measured in

centimetres or inches and can include fractional values, like 175.5 cm or 68.3 inches etc.

Other examples are;

- The weather temperature
- The wind speed
- The weight of the children

With these definitions and examples, I hope you can state some differences between the various types of data. Discuss them with your classmates.

Activity 1

Suppose these are the exam scores of some of your classmates in Mathematics in the last semester.

70, 60, 87, 90, 65, 90, 65, 76, 87, 95, 71, 85, 90, 62, 90

Below are the steps to help organise these exams scores in an ungrouped frequency distribution table.

Step 1: Identify all the unique values in the dataset (in this case, exam scores) from the smallest to the largest. These values will form the basis of your table's categories. These are 60, 63, 65, …, 98

Step 2: For each unique value, count how many times it appears in the dataset.

Step 3: Draw a frequency distribution table with two columns: one for the unique values (exam scores) and one for their frequencies.

Step 4: Organise this information into the table.

Step 5: Ensure that the total of the frequencies equals the total number of data points. This step verifies that all data points have been accounted for.

Total Frequency: $2 + 1 + 2 + 1 + 1 + 1 + 1 + 2 + 3 + 1 = 15$

Step 6: Review the data entry by double-checking for any errors in the frequency counts.

Table 3: Frequency Distribution

Exam Scores (in %)	Frequency
60	2
62	1
65	2
70	1
71	1
76	1
85	1
87	2
90	3
95	1

Let us proceed to that of a grouped frequency distribution table using the same exam scores.

Activity 2

Raw Data of the exam scores: 70, 60, 87, 90, 65, 90, 65, 76, 87, 95, 71, 85, 90, 62, 98

Step 1: Determine the Number of Groups (categories or intervals) you want to create. Typically, 5-7 groups are sufficient.

Step 2: Determine the Group Size:

- Calculate the range of the data (maximum value - minimum value).
- Divide the range by the number of groups to determine the group size, tweak as necessary for it to make sensible group sizes.
- In this example, group size = (98 - 60) /4 = 9.5. Therefore, to the nearest whole is 10, which seems a sensible group size.

Steps 3: Create Group Boundaries

- Start with the minimum value and add the group size to create the upper boundary of the first group. Note, this is discrete data,

so it is OK to have a gap between the upper bound of one group and the lower bound of the next, but if the data is continuous this cannot happen so you would need groups such as $60 \leq x < 70$ and $70 \leq x < 80$.

- Repeat this process to create the boundaries for each group.
  - 60-69
  - 70-79
  - 80-89
  - 90-99

Step 4: Create the Frequency Distribution Table

- Write the group boundaries in the exams scores column.
- Write the frequency in the frequency column.

Step 5: Review the data entry by double-checking for any errors in the frequency counts.

Table 4: Grouped Frequency Distribution

Exam Scores (in %)	Frequency
60-69	5
70-79	3
80-89	3
90-99	4

Quantitative and Qualitative Data

Quantitative data

Quantitative data is anything that can be counted or measured. It refers to numerical data. Data collected about a numeric variable will always be quantitative.

Characteristics of Quantitative data:

- It is numerical: Quantitative data is represented by numbers.
- It is measurable: Quantitative data can be measured or counted.
- Quantitative data is objective: It is based on facts, not opinions.
- Quantitative data is reliable: It is consistent and accurate.
- It can be generalised: Quantitative data can be applied to a larger population.

Examples of data that can be considered as quantitative:

- Kwabena is 5 feet and 7 inches tall
- They have size 6 feet
- They weigh 63 kilograms
- Mariama has one older sibling and two younger siblings
- They have two cats
- Setor lives twenty miles away from the school
- They go swimming four times a week

Qualitative data

Qualitative data is descriptive, referring to things that can be observed but not measured, such as colours or emotions. Data collected about a categorical variable will always be qualitative.

Characteristics of qualitative data:

- It is non-numerical: Qualitative data is descriptive, not numerical.
- Qualitative data is based on personal opinions and experiences.
- Qualitative data is contextual, meaning it is influenced by the research context
- It is rich and detailed: Qualitative data provides in-depth insights.
- It is flexible: Qualitative data collection methods can be adapted during research.

Examples of data that can be considered as qualitative;

- Maxjane has curly brown hair.
- They have brown eyes.
- Yamusah is funny, loud and a good listener.
- They can also be quite impatient and impulsive at times.
- Tinyor Raymond drives a red car
- They have a very friendly face and a contagious laugh.

Quantitative Data Collection Methods

1. **Surveys**: This involves using close-ended questions with fixed response options to gather numerical data from a large number of respondents.
2. **Experiments**: This is the method of conducting controlled experiments to collect quantitative data on the effects of variables on outcomes.

3. **Tests and Assessments**: This involves administering standardised tests or assessments to measure specific characteristics or abilities quantitatively.
4. **Observational Studies**: Observational studies involve collecting data through observation, without intervening or manipulating variables. Systematically observing and recording behaviours or events to gather numerical data is called observational studies.
5. **Sensor Data Collection**: Sensor data collection involves gathering data from sensors. These are devices that detect and measure physical or environmental parameters. Thus, using sensors and devices to collect quantitative data on physical phenomena such as: Temperature, Humidity, Light, Sound, Motion etc.

Qualitative Data Collection Methods

1. **Interviews**: An interview is a qualitative data collection method that involves a conversation between the researcher and the participant. It is a powerful tool for gathering in-depth, detailed information about people's experiences, attitudes and behaviours. In short, it is a method of conducting one-on-one or group interviews to gather in-depth information and insights from participants.
2. **Focus Groups**: Focus groups are a qualitative research method used to gather information from a small, diverse group of people. It involves bringing together a small group of people to discuss specific topics and gather opinions and perceptions.
3. **Observations**: Observation is a research method used to collect data by systematically observing (watching) and recording behaviours, interactions and phenomena in natural or controlled settings.
4. **Surveys**: A survey is a method of collecting data from a group of people through a series of questions. Surveys aim to gather information, opinions or feedback from a sample of individuals, often to understand their attitudes, behaviours or characteristics. Using open-ended questions to gather detailed responses and opinions from participants is termed a survey.
5. **Document Analysis**: This is a method of analysing written or visual materials such as documents, photographs or videos to understand a specific topic or phenomenon.

I hope you have noticed that survey and observation can be both quantitative and qualitative methods of data collection depending on the nature of the questions/ observations made.

Take note of the differences between all these different types of data collection and discuss them with your classmates. Then have a look at the following samples of survey questionnaires, observation checklist and interview guide in the activity below.

Activity 3

Identify some features of each of the sample data collection tools below. Share these with your classmates.

A Sample Survey Questionnaire

Instructions: Please indicate your level of agreement or disagreement with each of these statements regarding QRZ Family Restaurant. Place an "X" mark in the box of your answer.

Q1: How many times per year do you visit QRZ Family Restaurant?

_____

Q2: Do you visit QRZ Family Restaurant with family or friends?

☐ **Yes** ☐ **No**

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. The store is accessibly located.					
2. Store hours are convenient for my dining needs.					
3. Advertised dish was in stock.					
4. A good selection of dishes was present.					
5. The meals sold are good value for the money.					
6. Store has the lowest prices in the area.					
7. Meals sold are of the highest quality.					
8. Store atmosphere and decor are appealing.					

Q9: How would you rate your overall experience at the QRZ Family Restaurant?

- ☐ Highly satisfactory
- ☐ Satisfactory
- ☐ Neutral
- ☐ Unsatisfactory
- ☐ Highly Unsatisfactory

Q10: What could we do to make your restaurant dining experience better?

_____

_____

_____

Q11. Demographic Data

Name (optional):_____

Age: ____

Gender: _____

Number of Family Members:

- ☐ 1-2
- ☐ 3-5
- ☐ 6-10
- ☐ **more than 10**

Email Address (optional): _____

Notes: This section is optional. The questions asking for demographic data should be relevant to the survey goal and must point to the characteristics of the target population.

IV. Thank you for sharing your thoughts with us. We hope you enjoy dining at QRZ Family Restaurant and look forward to seeing you back with us soon.

Note: This section may also include further information regarding on how to claim the incentive that you wish to provide to the respondent.

A Sample Observation Guide

Date: _____

Time: _____

Observer's Name: _____

Location: _____

Observer's Role: _____

1. **Context and Environment**
   - **Classroom Setting:**
     - Layout (e.g., rows, groups, seating arrangement)
     - Resources available (e.g., computers, books, materials)
     - General atmosphere (e.g., quiet, noisy, relaxed)
   - **Time of Day:**
     - (e.g., Morning, Midday, Afternoon)
   - **Type of Activity:**
     - (e.g., Group Work, Lecture, Independent Study)
2. **General Behaviour**
   - **Student Engagement:**
     - Active/Passive/Neutral
     - Participation level (e.g., frequently participates, seldom participates)
   - **Attention:**
     - Focused/Distracted/Varies
   - **Emotional State:**
     - Positive/Negative/Neutral
3. **Specific Behaviours**
   - **Interactions with Peers:**
     - Cooperative/Competitive/Isolated
     - Examples: [e.g., Student A helps Student B with a problem]

A Sample Interview Guide

Interview Guide

First name:	Last name:
Interview date:	Allowed to work in the country: Yes ☐ No ☐
Position sought:	Criminal record: Yes ☐ No ☐

Availability

On what date will you be available to start working?: _____

Minimum number of hours of availability per week: _____

Maximum number of hours of availability per week: _____

How long will you remain available for work?: _____

Availability breakdown per day

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
From							
To							

[The following questions are provided as a reference; adapt them to your situation as needed.]

Introduction

Question	Notes
What can you tell us about yourself and your background?	
What are your main fields of interest, your favorite activities, and your passions?	
Where do your relatives live? How long do you plan to stay in the area?	

Look for other samples and discuss the features of each of them with your classmates.

Activity 4

Your school is wondering about increasing the contact hours to provide students with more instructional time to interact with their teachers. If the school management wants to gather the students' opinions on this, which data collection method(s) would you recommend and why? Organise yourselfs in small groups to discuss your ideas and come together as a class. There are many different ways to go about this. Which one would work best in your school?

DATA PRESENTATION METHODS: FREQUENCY DISTRIBUTION TABLE

In Junior High School, you learned about various methods for presenting data, such as bar charts, frequency tables, pie charts and line graphs. These methods have many real-world applications for both categorical and numerical data. Frequency tables can be used for datasets of any size, while pie charts can show part-to-whole relationships but they can be misleading if not used carefully. Line graphs can be used for representing continuous data.

Frequency Tables

Ungrouped data

Example 2

A jar contains beads of different colours. Here is the raw data for this: red, green, blue, black, red, green, blue, yellow, red, red, green, green, green, yellow, red, green, yellow.

Organise the data in a frequency distribution table.

Solution

- Create the frequency distribution table with **respective** columns
- **fill in the** data, tally (count) and frequency column
- Compare your final result to the table below

Table 5: Frequency Distribution

Categories	Tally Marks	Frequency
Red	𝍸	5
Green	𝍸 |	6
Blue	||	2
Black	|	1
Yellow	|||	3

Grouped data

Example 3

The raw data below represents the number of trees that survived in a survey by 100 schools who decided to plant 100 tree saplings in their gardens on world environment day.

95, 67, 28, 32, 65, 65, 69, 33, 98, 96, 76, 42, 32, 38, 42, 40, 40, 69, 95, 92, 75, 83, 76, 83, 85, 62, 37, 65, 63, 42, 89, 65, 73, 81, 49, 52, 64, 76, 83, 92, 93, 68, 52, 79, 81, 83, 59, 82, 75, 82, 86, 90, 44, 62, 31, 36, 38, 42, 39, 83, 87, 56, 58, 23, 35, 76, 83, 85, 30, 68, 69, 83, 86, 43, 45, 39, 83, 75, 66, 83, 92, 75, 89, 66, 91, 27, 88, 89, 93, 42, 53, 69, 90, 55, 66, 49, 52, 83, 34, 36

Represent the given data in the form of a grouped frequency distribution table.

Solution

- Create the frequency distribution table with **respective** columns
- **fill in the** data, tally (count) and frequency column
- Compare your final result to the table below

Table 6: Grouped Frequency Distribution

Number of plants survived	Tally Marks	Number of schools (frequency)
20 - 29	𝍸	3
30 - 39	𝍸 𝍸 ||||	14
40 - 49	𝍸 𝍸 ||	12
50 - 59	𝍸 |||	8
60 - 69	𝍸 𝍸 𝍸 |||	18
70 - 79	𝍸 𝍸	10
80 - 89	𝍸 𝍸 𝍸 𝍸 |||	23
90 - 99	𝍸 𝍸 ||	12
Total		100

Line Graphs

A **line graph** is a type of chart used to display data points connected by straight line segments

Steps in drawing a line chart

1. Choose a suitable scale for the x and y axes.
2. Draw two perpendicular lines and label the axes:
   x-axis (horizontal): independent variable (e.g., time, categories)
   y-axis (vertical): dependent variable (e.g., values, amounts)
3. Mark each data point on the graph with a small dot or cross.
4. Ensure accurate placement according to x and y values.
5. Connect the data points with straight lines using a ruler.
6. Add a title to the graph.

Example 4

The temperature recorded in a city from 5 a.m. to 8 p.m. on a day was recorded in the form of a line graph as shown below. Study the graph and answer the following questions.

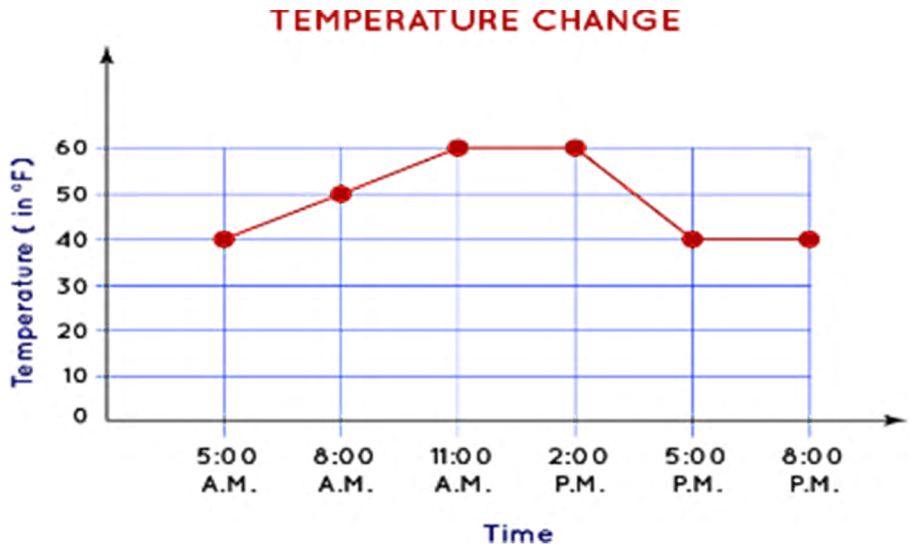


Figure 1: Line graph

a. At which time(s) of the day was the temperature 40° F?
b. What was the maximum recorded temperature?

Solution

a. 5:00 a.m. and 5-8 pm.
b. 60° F

Bar Graphs

A **bar graph** (or bar chart) is a visual representation of data using rectangular bars, where the length or height of each bar corresponds to the value it represents.

Steps in drawing a bar graph

1. Choose a suitable scale for the x and y axes.
2. Draw two perpendicular lines and label them:
   x-axis (horizontal): categories
   y-axis (vertical): numerical values
3. Decide on a uniform bar width.
4. Leave equal spaces between bars (at least 1/2 bar width).
5. Measure and draw each bar's height according to its value.
6. Ensure bars are aligned with corresponding x-axis labels
7. Label each bar with its corresponding value (optional).
8. Add a title to the graph.

Example 5

The following table shows the percentage of monthly salary saved by each employee in a 100-person company. Create a vertical bar graph to represent it

Table 7: Distribution table

Savings (in percentage)	Number of Employees
15	20
25	45
35	25
45	10
Total	**100**

Solution

Figure 2: Bar graph

Example 6

The bar chart below is the summary of cases and recoveries of corona virus pandemic recorded in various regions in Ghana from March 2020 to August 2022.

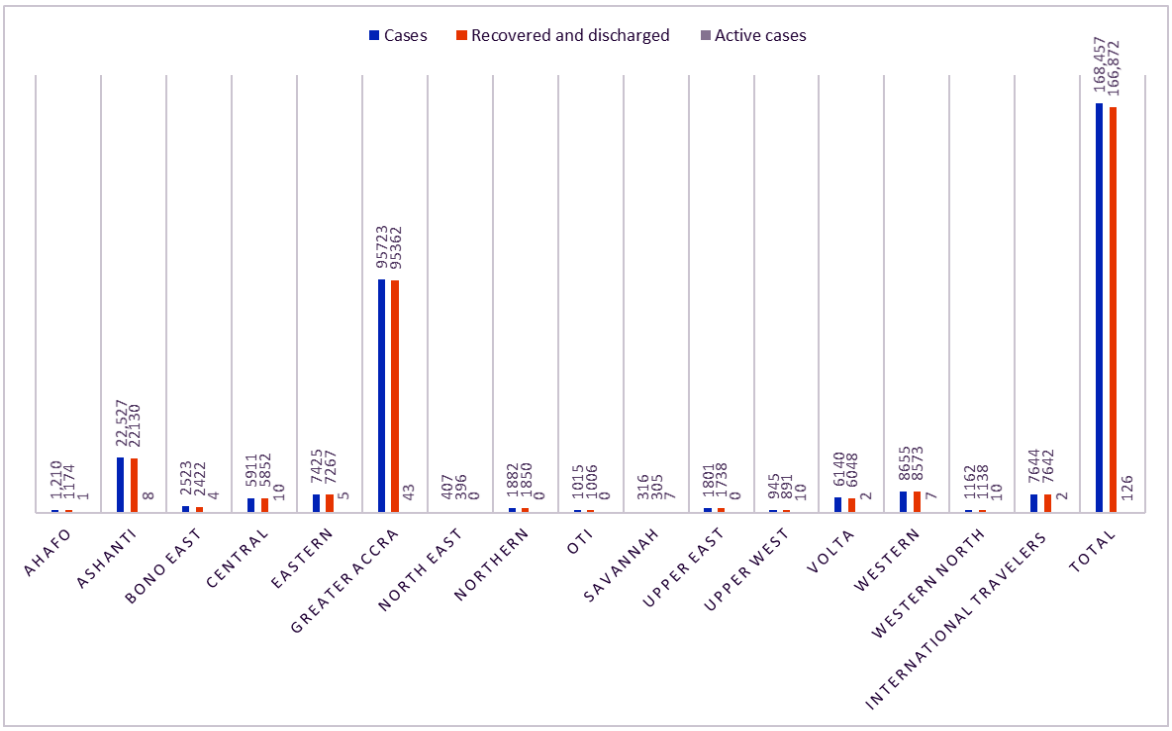


Figure 3: Double bar graph

Using the data in the graph answer the following questions:

1. Which region recorded the least cases over the period?
2. What percentage of the total cases recorded were in Greater Accra?

Solution

1. Savannah recorded the least cases.
2. Approximately 57% $\left(\frac{95723}{168457} \times 100\right)$

Pie Charts

Example 7

The data below shows the amount of money that Akosua spent on buying some items:

- Pepper - GH4
- Onions - GH4
- Salt - GH2
- Fish - GH10

Draw a pie chart for the data.

Solution

To draw a pie chart for a given data

Step 1. Find the sum of the frequencies i.e. $4 + 4 + 2 + 10 = 20$. Since a circle contains a total of 360°, we need to express each fraction for a sector as an angle per 360°.

Step 2. Calculate the angle for a data value of each sport as follows:

i. Find the angle for a data value $\frac{360}{20} = 18°$

ii. Next, we find the angle for each sector by multiplying the number of pupils (frequency) for that sector by 18 as shown in the table below.

Alternatively, express each frequency over the total frequency and multiply by 360°.

Table 8: Calculating Angle of Sector

Items	Amount Spent (GHc)	Angle of sector	Alternatively, Angle of sector
Pepper	4	$4 \times 18 = 72°$	$\frac{4}{20} \times 360° = 72°$
Onions	4	$4 \times 18 = 72°$	$\frac{4}{20} \times 360° = 72°$
Salt	2	$2 \times 18 = 36°$	$\frac{2}{20} \times 360° = 36°$

Items	Amount Spent (GHc)	Angle of sector	Alternatively, Angle of sector
Fish	10	$10 \times 18 = 180°$	$\frac{10}{20} \times 360° = 180°$
Total	**20**	**360°**	**360°**

Step 3. Use a pair of compasses to draw a circle with a radius of at least 5cm.

Step 4. Use a protractor to draw (measure) the angles for each sector in the circle.

Step 5. Write and label each sector.

Step 6. It is usually appropriate to write the angles (magnitudes) inside the sectors.

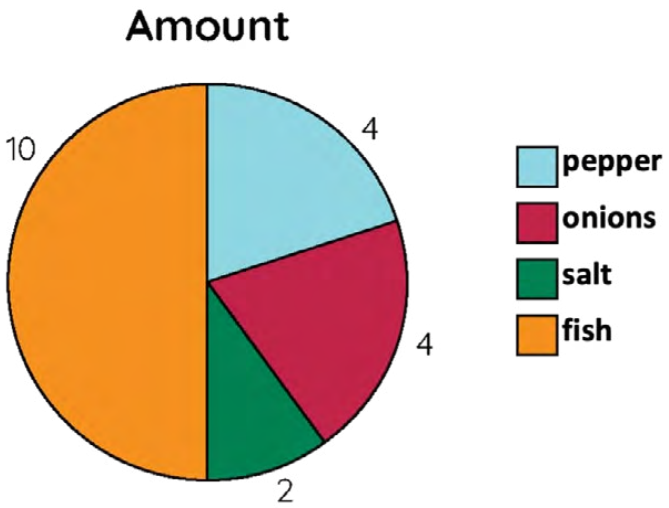


Figure 4: Pie chart

Measures of Central Tendency

As data analysts, statisticians or curious learners, understanding the central tendencies of a dataset is crucial in gaining a meaningful understanding of its characteristics. These statistical measures provide us with a concise summary of the central or typical values within a dataset. They help us identify the centre around which the data clusters and give us insights into the overall distribution.

We will look at three primary measures of central tendency: the mean, the median, and the mode.

Mode, Median and Mean

The mode refers to the value or values which appear most frequently in a dataset. It helps identify the most common or popular value(s) within a dataset and is especially useful in categorical or discrete data analysis.

The median represents the middle value of a dataset when arranged in ascending or descending order. It is not affected by very high or very low numbers and is especially helpful when the data is uneven or has values that are very different from the rest.

The mean, often referred to as the average, is computed by summing all the values in the dataset and dividing by the total number of observations. It represents the arithmetic centre and provides a measure of the typical value in the dataset.

Let us consider the steps in finding the mean, median and mode of a raw data set.

Calculating mean, median and mode for raw data

Example 8

The marketing department of a company has collected the following sales figures (in thousands of Ghana cedis) for its top 10 sales representatives for the previous quarter:

11, 18, 15, 22, 19, 14, 17, 20, 16, 22

Calculate the following measures of central tendency for the given sales figures:

1. Mode
2. Mean
3. Median

Solution

1. Calculating the **Mode**: The mode is the value that appears most frequently in the data set. In the given data, 22 appears twice, everything else only appears once, so the mode in this data set is **22**.
2. Calculating the **Mean**: The mean is the sum of all the values divided by the total number of values.

   Sum of all values $= 11 + 18 + 15 + 22 + 19 + 14 + 17 + 20 + 16 + 22$
   $= 174$

   Total number of values $= 10$

   Mean $= \frac{174}{10} = \mathbf{17.4}$

3. Calculating the **Median**: The median is the middle value when the data is arranged in ascending or descending order. Arranging the data in ascending order:

   11, 14, 15, 16, 17, 18, 19, 20, 22, 22

   Since there are 10 values, the median is the average of the 5th and 6th values.

   $\text{Median} = \frac{(17 + 18)}{2} = \mathbf{17.5}$

Therefore, the measures of central tendency for the given sales figures are:

- Mode: 22
- Mean: 17.4
- Median: 17.5

Now that you know how to calculate measures of central tendencies with a raw data set, let us now look at how to find mean, median and mode from a frequency distribution table.

Calculating the mean, median and mode from an *ungrouped* frequency distribution table

Example 9

The data below shows the marks obtained in a short science test.

Organise the data in a frequency distribution table and find the mode and median.

2	2	4	3	1	4	1	0
1	2	3	1	2	1	3	0
4	2	3	3	3	4	3	1
1	0	1	4	0	0	1	4

Solution

Construct your frequency distribution table:

Table 9: Frequency Distribution

Marks	Tally	Frequency
0	𝍸	5
1	𝍸 ////	9
2	𝍸	5
3	𝍸 //	7
4	𝍸 /	6

To obtain the **mode**, look for the mark with the highest frequency. From our frequency table "mark 1" had the highest frequency of 9, hence it is the mode of the data set.

To obtain the **median** from the frequency table, follow these steps:

1. Find the total frequency. To do this, sum up all the frequencies.
2. Divide the total frequency by 2.
3. Find the item that correspond to the frequency from your answer in point two above. In the example, total frequency = 32
   $\therefore \frac{32}{2} = 16$

Table 10: Frequency Distribution

Marks	Tally	Frequency
0	𝍸	5
1	𝍸 ////	9
2	𝍸	5 16th & 17th
3	𝍸 //	7
4	𝍸 /	6

From the frequency table the 16^{th} and 17^{th} frequencies fall within 2. Therefore, the median is 2.

N.B.: When the total frequency is an even number, it means there are two numbers in the middle. It means the median will halfway between the 16th and the 17th number.

To calculate the mean we use the formula below:

The mean formula for ungrouped data is

$Mean = \bar{X} = \sum \frac{fx}{\sum f}$

Where:

$\sum$ = is the summation sign, meaning total sum of…

$\bar{X}$ = the mean value of the set of given data

f = frequency of the individual data

x = the data point

Therefore, for our data:

Table 11: Frequency Distribution

Marks / x	Tally	Frequency / f	f × x
0	𝍸	5	$5 \times 0 = 0$
1	𝍸 ////	9	$9 \times 1 = 9$
2	𝍸	5	$5 \times 2 = 10$
3	𝍸 //	7	$7 \times 3 = 21$
4	𝍸 /	6	$6 \times 4 = 24$
Total		**$\Sigma f = 32$**	**$\Sigma fx = 64$**

$\therefore Mean = \Sigma \frac{fx}{\Sigma f} = \frac{64}{32} = 2$

Example 10

The data below are marks obtained in a social studies test in JHS 1. Calculate the mean for this data.

3	2	2	3	4
2	5	4	2	2
2	2	1	4	2
4	5	3	2	7
2	3	2	6	3

Solution

To obtain the mean from the frequency table, continue as follows:

Step 1. Organise the data into a frequency distribution table.

Step 2. Find the total frequency (f). To do this, sum all the frequencies.

Step 3. Multiply the frequency (f) by the test score (x)

Step 4: Sum all the values under the (fx) column.

Step 5. Divide the summation of fx by the summation of f

Table 12: Frequency distribution table

Marks (x)	Tally	Frequency (f)	fx
1	/	**1**	**1**
2	~~////~~ ~~////~~ /	**11**	**22**
3	~~////~~	**5**	**15**
4	////	**4**	**16**
5	//	**2**	**10**
6	/	**1**	**6**
7	/	**1**	**7**
Total		**25**	**77**

$\sum f = 25, \sum f\text{x} = 77$

$$Mean\ (\bar{x}) = \frac{\sum fx}{\sum f} = \frac{77}{25} = 3.08$$

$\therefore$ the mean score is 3.08

Calculating the Mean, Median and Mode from a grouped frequency distribution table

Example 11

You grew fifty carrots using special soil. You dig them up and measure their lengths (to the nearest mm) and group the results.

Calculate the mean, median and mode of the lengths of carrots.

Table 13: Frequency Distribution

Length (mm)	Frequency
150 - 154	5
155 - 159	2
160 - 164	6
165 - 169	8
170 - 174	9
175 - 179	11
180 - 184	6
185 - 189	3

Solution

Step 1: Calculate the **mean** from the table below. In this case x is the **mid-point** of the group.

Table 14: Frequency Distribution

Length (mm)	Midpoint (x)	Frequency (f)	fx
150 - 154	152	5	760
155 - 159	157	2	314
160 - 164	162	6	972
165 - 169	167	8	1336
170 - 174	172	9	1548
175 - 179	177	11	1947
180 - 184	182	6	1092
185 - 189	187	3	561
	Totals:	**50**	**8530**

$$\text{Estimated Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{8530}{50} = 170.6 \text{ mm}$$

Note, it is only an 'estimated' mean as we have not used each exact carrot length, but only estimated by taking the mid-point of the group.

Step 2: Find the **median** group.

The median is the mean of the 25th and the 26th length, so it occurs in the **170 – 174** group.

Step 3: Find the **mode**, or, **modal** group

The modal group is the one with the highest frequency, which is **175 – 179**:

Establish that the mean is a form of average

The term average is frequently used in everyday life to denote a value that is typical for a group of quantities. For example, *average* rainfall in a month or the *average* age of employees of an organisation are typical examples.

- Average is the value that indicates what is most likely to be expected.
- They help to summarise large data into a single value.

Table 14: Differences between mean and median

Mean	Median
The mean, also known as the average, is calculated by summing all the values in a dataset and dividing by the total number of observations.	The median represents the middle value in a dataset when arranged in ascending or descending order.
It is influenced by extreme values or outliers in the dataset.	It is not affected by extreme values or outliers in the dataset.
The mean represents the arithmetic centre or the typical value of the dataset.	The median is particularly useful when dealing with skewed distributions or datasets with outliers.
It is widely used and provides a measure of the central tendency that takes into account all values in the dataset.	It provides a measure of the central value that is robust to extreme values.

Justifying which of the averages best represents given data

Example 12

A department of an organisation has 5 employees which include a supervisor and four executives. The executives draw a salary of GH 10 000 per month while the supervisor gets GH 40 000.

Calculate the mean and median for the data.

Which do you think better represents the data set?

Solution

Step 1: Find the **mean** salary

$$\text{Mean} = \frac{10000 + 10000 + 10000 + 10000 + 40000}{5} = \frac{80000}{5} = 16\,000$$

Thus, the mean salary is GH 16 000 per month.

Step 2: Find the **median** by arranging the given data in ascending order:

10000, 10000, 10000, 10000, 40000.

There are 5 pieces of data, so the middle piece of data is the 3rd piece.

Thus, the median is = GH10 000 per month

Step 3: Compare the two measures of central tendencies:

We can observe that the mean salary of GH 16,000 does not give even an estimated salary of any of the employees whereas the median salary represents the data more effectively. One of the weaknesses of the mean is that it gets affected by extreme values, which, in this case, is that of the supervisor. The median is not skewed by this data, but the mean is.

Effect of an extreme value on the mean

The following graphs show how extreme values affect mean and median:

- Symmetric Data
  - Data sets whose values are evenly spred around the centre
- Skewed Data
  - Data set that are not symmetric

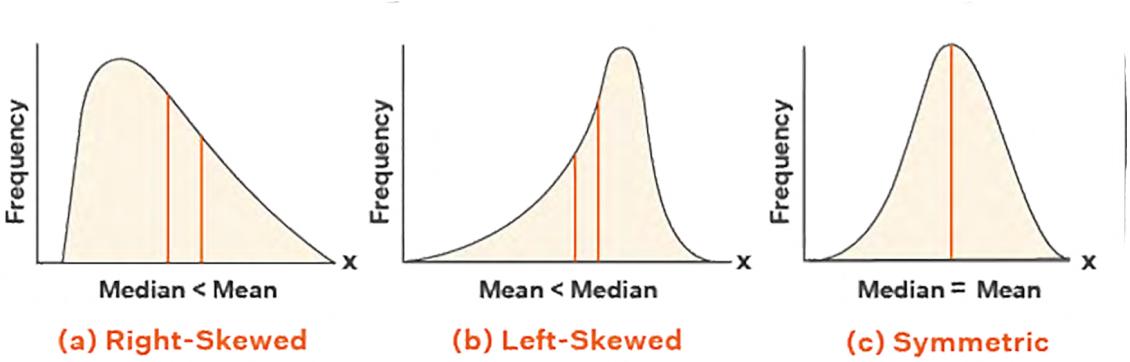


Figure 5: Effect of an extreme value on the mean

So, the mean can be used when we do not have extreme values in the data but the median is better if we have extreme values.

Revision Activity: Mean, Median and Mode

To review and reinforce the concepts of mean, mode, and median from raw data sets and frequency distribution tables follow these instructions.

Instructions:

1. **Watch the Videos**: Click on the following links to watch videos that demonstrate how to find the mean, median and mode from both raw data sets and frequency distribution tables

   - Video on Finding Mean, Median and Mode from Raw Data Set (https://youtu.be/A1mQ9kD-i9I)

   - Video on Finding Mean, Median and Mode from Frequency Distribution Tables (https://youtu.be/zjHfAhcU6kE)

2. **Take Notes**: While watching the videos, take notes on the key steps and methods used to calculate the mean, median and mode. Pay special attention to any examples provided in the videos.
3. **Practice Problems**: Think about what raw data/frequency distribution tables you could use to make your own problems on which to practise. Use these to challenge your classmates. Ensure all steps are shown, with calculations clearly written.
4. **Review and Discuss**: After completing the practice problems, review your answers with a classmate or in a group. Discuss any difficulties or questions you might have. If needed, refer back to the videos, or ask your teacher for clarification.

Use of mathematical language in making inferences of data

Example: 13

The table below is the summary of cases and recoveries of coronavirus recorded in various regions in Ghana from March 2020 to 11th August 2022. Use the information from the table to make recommendations to the government on resource allocation to the various regions.

Table 15: Cases and recoveries of coronavirus in Ghana from March 2020 to 11th August 2022

Regions	Cases	Recovered and discharged	Active cases
Ahafo	1,210	1174	1
Ashanti	22,527	22130	8
Bono east	2523	2422	4
Central	5911	5852	17
Eastern	7425	7267	5
Greater Accra	95723	95362	43
North East	407	396	0
Northern	1882	1850	0
Oti	1015	1006	0
Savannah	316	305	7

Regions	Cases	Recovered and discharged	Active cases
Upper east	1801	1738	0
Upper west	945	891	10
Volta	6140	6048	2
Western	8655	8573	7
Western north	1162	1138	10
International Travellers	7644	7642	2
Total	**168,457**	**166,872**	**126**

Examples of recommendations and their justifications

Recommendation 1: More resources should be allocated to Greater Accra region and Ashanti region since they recorded the highest cases, i.e., Greater Accra 95723 (56.8%) and Ashanti 22,527 (13.4%).

Recommendation 2: More resources should be allocated to Greater Accra region and Central region since they have the highest number of active cases, i.e., Greater Accra 43 (34%) and Central 17 (13.5%).

The two recommendations are based on the data. Each of the recommendation has a good basis from the given data. You must learn to make such recommendations based on data and accept alternative views of others since it can also be useful.

Real-life Project on Data collection, analysis and presentation

Project 1

- In convenient groups, obtain the West Africa Secondary School Certificate Exams (WASSCE) result of your school for the past five years and analyse it by looking at the overall differences in the performance by years, by programme/courses/subject.
- As part of the analysis, obtain the frequencies and percentages, then draw charts (line, pie, bar etc.) for the data.
- Also, from the analysis make conclusions and give recommendations to the school.

Project 2

- Design Student Course Evaluation Questionnaires and use them to collect data from students in your school.
- Analyse the data and make conclusions and recommendations based of the results.

Publishing Data Findings

Example 14

i. Make summaries of your results, conclusions and recommendations of your project and present them using power point, infographics design, Microsoft word or handwritten notes to the class or at a mini forum (including school clubs) in the school.

ii. Make oral presentations on the project by explaining the choice of project topic and its relevance, choice of data collection method(s) and the analysis and talk about the challenges faced.

iii. Publish the summary of the findings and recommendations of the project on school notice board or magazines, school social media platforms, etc.

REVIEW QUESTIONS

1. Classify each of the following data types as primary or secondary data;
   a) interviews
   b) reports
   c) experiments
   d) observations
   e) published literature
   f) databases
2. The data below are the scores of 21 learners in a creative arts test.

   Organise it into a grouped frequency distribution table

   59, 65, 61, 62, 53, 55, 60, 70, 64, 56, 58, 58, 62, 62, 68, 65, 56, 59, 68, 61, 67
3. Design an interview guide for your school to be used to collect qualitative data on students who are aspiring for House Prefect's positions in the schools.
4. A non-governmental organization (NGO) wants to embark on a developmental project in your community but would like to know the most preferrable project the community needs. Suggest the data collection tool(s) that will be appropriate in seeking the views of the members of the community.
5. A survey is conducted to ask 15 households how many pets they have in their home.

   The results are as follows:

   1, 1, 1, 1, 2, 2, 2, 3, 3, 4, 5, 5, 6, 7, 8

   Construct both an ungrouped and a grouped frequency table for the data.

6. The table shows Russel's height at 3-year intervals.
Make a line graph to display the data.

Age	Height(ft)
3	1.9
6	2.4
9	2.7
12	3.0
15	4.6
18	5.5
21	6.1
24	6.2

7. A survey of 145 people asked them "Which is your favourite fruit?" Construct a bar chart for the information below:

Fruit	People
Apple	35
Orange	30
Banana	10
Kiwifruit	25
Blueberry	40
Grapes	5

8. The table below shows the number of boys and girls in all JHS classes at Winneba Don Bosco B JHS.

Construct a bar chart for the data.

SEX	JHS1	JHS2	JHS3
BOYS	14	19	23
GIRLS	20	18	21

9. Find the mode of the data {14, 16, 16, 16, 17, 16, 18}.

10. The ages of the members of a community have been listed below:

{42, 38, 29, 37, 40, 33, 41}

Calculate the median of the given data.

11. Find the mean, median and mode for the following list of values: 13, 18, 13, 14, 13, 16, 14, 21, 13

12. Estimate the mean, median and mode for the data set.

Seconds	Frequency
51 - 55	2
56 - 60	7
61 - 65	8
66- 70	4

13. The ages of the 112 people who live on a tropical island are grouped as follows:

Age	Number
0 - 9	20
10 - 19	21
20 - 29	23
30 - 39	16
40 - 49	11
50 - 59	10
60 - 69	7
70 - 79	3
80 - 89	1

(a) Estimate the mean, median and mode for the data.

(b) Analyse the three central scores (mean, median and mode) calculated and justify why a particular one best represents the data

14. Estimate the mean, median and mode for the data set.

Length (mm)	Frequency
150 - 154	5
155 - 159	2
160 - 164	6
165 - 169	8
170 - 174	9
175 - 179	11
180 - 184	6
185 - 189	3

Analyse the three central scores (mean, median and mode) calculated and justify why a particular one best represents the data.

15. State and explain, with practical examples two differences between mean and median.

16. Explain, with practical examples the effect of an extreme value on the mean

ANSWERS TO REVIEW QUESTIONS

1. Primary data: interview, experiment, observation,

   Secondary data: reports, published literature, databases.

2. For example:

Test Score	Frequency
50-54	1
55-59	7
60-64	7
65-69	5
70-74	1

3. Compare your answers with those of your classmates. Between you, can you come up with a perfect interview guide?

4. Again, compare your suggestions with those of your classmates to make a plan for the best route for this.

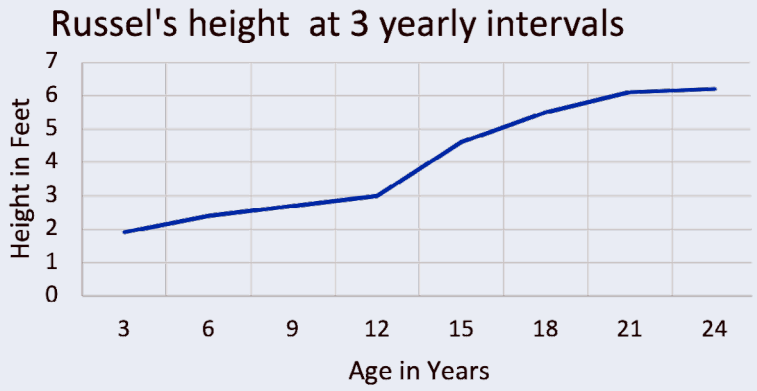
5. Ungrouped Frequency Table:

Number of pets	Frequency
1	4
2	3
3	2
4	1
5	2
6	1
7	1
8	1

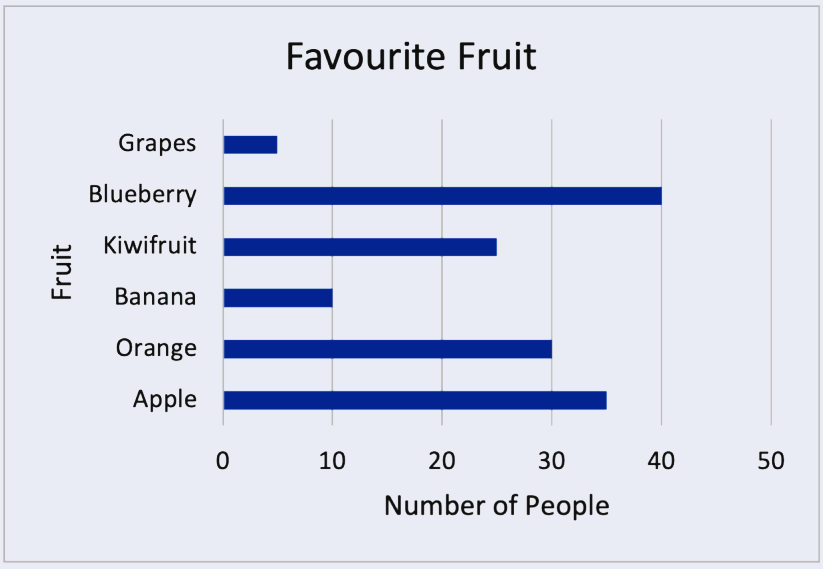
Grouped Frequency Table:

Number of pets	Frequency
1 - 2	7
3 - 4	3
5 - 6	3
7 - 8	2

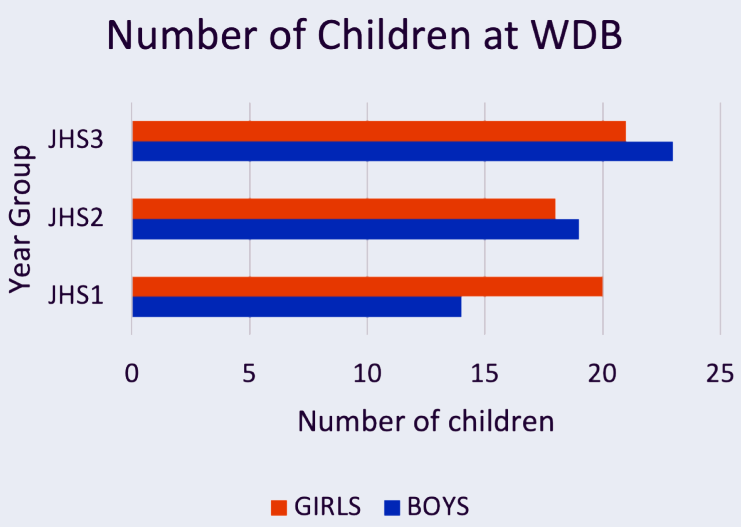
6.



7.



8.



9. Mode is 16

10. Median is 38

11. Mean: 15, Median: 14, Mode: 13

12. Mean: approximately 61.3

Median group: 61 – 65

Modal group: 61 – 65

13. **(a)** Mean is estimated at 30. The median group is 20 – 29, the modal group is 20 – 29.

(b) The mean is being skewed slightly by the people over 70 so the median group may be a better value but this is only a range of values, so we could stick with the mean.

14. The mean is estimated to be 170.6, the median group is 170 – 174 and the modal group is 175 – 179. The mean value lies within the median group. Therefore, the mean value seems reasonable to take.

15. The mean is calculated by summing all the values in a dataset and dividing by the total number of pieces of data. The median represents the middle value in a dataset when it is arranged in order. The mean lies within the median group, so this seems like a good central value. The mean is influenced by extreme values, or outliers, in the dataset, but the median is not. The mean takes into account all the values within a dataset, the median does not.

16. A sample answer:

Consider the monthly salaries (in thousands) of five employees at a small company:

Salaries: \$3, \$4, \$4, \$5, \$50

Calculate the Mean:

$$\text{Mean} = \frac{3+4+4+5+50}{5} = 665 = 13.2$$

In this dataset, the salary of \$50 is an extreme value compared with the other salaries. Without this extreme value, the calculation would be:

- **Salaries (without \$50)**: \$3, \$4, \$4, \$5

  $$\text{Mean} = \frac{3+4+4+5}{4} = \frac{16}{4} = 4$$

Effect: The mean increases from \$4 to \$13.2 due to the extreme value of \$50. This extreme value distorts the mean, making it appear that the average salary is higher than it is for the majority of employees. Hence, in such cases a better measure of central tendency would be the median.

GLOSSARY

- **Data:** Raw facts or information.
- **Data Organisation:** Arranging data systematically.
- **Data Analysis:** Examining and interpreting data.
- **Data Presentation:** Showing data visually (charts, graphs).
- **Qualitative Data:** Descriptive, non-numerical data.
- **Quantitative Data:** Numerical, measurable data.
- **Frequency Distribution:** How often values occur in data.
- **Mean:** Average of data values, found by the sum of data pieces divided by the total number of pieces of data.
- **Median:** Middle value in ordered data.
- **Mode:** Most frequent value in data.
- **Range:** Difference between highest and lowest values.
- **Pie Chart:** Circular chart showing parts of a whole.
- **Bar Chart:** Bars representing different data values.
- **Line Graph:** Graph showing changes over time.
- **Outliers:** Data points far from others.
- **Trend:** General direction of data over time.
- **Survey:** Method to collect data via questions.
- **Census:** Data collected from an entire population.
- **Sample:** Subset of a population for analysis.

EXTENDED READING

- Akrong Series: Core mathematics for Senior High Schools New International Edition (Pages 645 – 698)
- Aki – Ola series : Core Mathematics for Senior High Schools in West Africa, Millennium edition 5 (Pages 267 – 294)
- Baffour Asamoah, B. A. (2015). *Baffour BA series: Core mathematics*. Accra: Mega Heights, (Pages 259 - 298)

REFERENCES

- Akrong Series: Core mathematics for Senior High Schools New International Edition (Pages 645 – 698)
- Aki – Ola series: Core Mathematics for Senior High Schools in West Africa, Millennium edition 5 (Pages 267 – 294)
- Baffour Asamoah, B. A. (2015). *Baffour BA series: Core mathematics*. Accra: Mega Heights, (Pages 259 - 298)
- Basic mathematics for TTC volume one by Rev.Kusi Appau, October 2001

ACKNOWLEDGEMENTS

Ghana Education Service (GES)

List of Contributors

NAME	INSTITUTION
Janet Waponade	OLA SHS
Yamusah Issahaku	Kumbungu SHS
Joseph Bakpil Nagbija	Nchumuruman SHS, Chinderi
Modzaka Godfred	Keta SHTS, Keta