



Rosary School \ Marj Elhamam Worksheet (4)

Name : _____

Date: / 11 / 2025

Subject: (Unit 6: Collecting and analyzing data)

Grade : 8 ()

➤ **Primary Data**

Primary data is information that **you collect yourself** for a specific purpose.

Examples:

- Asking your classmates about their favorite sport and recording the results.
- Measuring the heights of students in your class.
- Counting how many cars pass the school gate in 10 minutes.

➤ **Secondary Data**

Secondary data is information that was **already collected by someone else**, and you use it.

Examples:

- Using data from a website about temperature records.
- Reading statistics from a government report.
- Using information from a textbook or newspaper.

➤ **Methods of Collecting Data**

1. **Survey:** A survey is both the set of questions and the process of collecting, aggregating, and analyzing the responses from those questions.
2. **Questionnaire:** A questionnaire is any written set of questions.
3. **Observation:** You watch and record what happens without asking people.
4. **Data logging:** Is a process of collecting data using data logging devices.

Q6 page 122

6 For the following data:

- i decide whether it is **primary** or **secondary data**
- ii if it is primary data, give a suitable method for collecting it.

- a The total time spent in the Emergency Department in every hospital in Australia from April 2018 to June 2019.
- b Numbers given by people who were asked to choose their favourite number.
- c Percentage of the Indian population that used the internet in 2019.
- d IGCSE results of all pupils in New Zealand in 2019.
- e The percentage of people who write with their right hand.
- f The number of people in cars on a main road.

a) _____

b) _____

c) _____

d) _____

e) _____

f) _____

➤ Meaning and size of a sample

What is a Sample?

A **sample** is a small group taken from a larger group (the population) to collect data.

Example:

If the school has **500 students**, and you collect data from **50 students**, the 50 students are the *sample*.

Key point



The total number of items your survey relates to is called the **population**.

The group of items you test is called a **sample**.

Sampling can be time consuming and expensive but the bigger the sample, the more reliable it can be.

For a sample to be reliable and unbiased, it should

- be at least 10% of the population
- represent the population.

Q2+3 page 132

2 A school has 1800 students. It wants to investigate how its students travel to school. Select the most appropriate sample size for the survey.

A 1000 B 500 C 200 D 20

3 A library has 3860 members. How big should a sample of the library members be?

➤ When Is a Question Biased in a Survey?

A biased question is unfair and leads people to answer in a certain direction instead of giving their true opinion.

1. Leading Words

The question suggests the answer you want.

Example:

“Don’t you think the school cafeteria food is amazing?”

This pushes people to say **yes**.

2. Restrictive

The answer options are limited.

Example:

“How often do you exercise per week?

Every day 1 – 2 days 3 – 4 days

No option for “never” or “5 – 6 days.”

3. Unclear (Vague)

If the question is too long or unclear, people might misunderstand.

Example:

“How many times did you not forget to bring your homework last week?”

Q11 page 123

11 Explain why each of these **leading questions** is unsuitable to use in a survey.

For each question write a more suitable question to replace it.

a What methods of travel have you used in the last 12 months?

car bus train taxi none

<input type="checkbox"/>				
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

b Petrol prices keep rising, so do you intend to use your car less during the next year?

c What do you do at weekends?

d Should the inadequate bus service in our town be improved?

e The main cause of bad behaviour in lessons is because the

lesson is boring teacher is not strict enough lesson is too long

□ □ □

f Because killing animals is cruel, should more people become vegetarians?

g What kind of food do you eat most of? Key point

Key point



A question in a survey can be biased if it encourages people to give a particular answer.

A good survey question should not be

- unclear
- leading
- restrictive.

a) _____

b) _____

c) _____

d) _____

f) _____

g) _____

- A two-way table is a way of organizing data that shows the relationship between two different categories **at the same time**.

Q3 page 124

3 The table shows the driving test results of 50 people.

	Men	Women
Pass	11	10
Fail	15	14

- a What percentage of the drivers passed their driving test?
- b What fraction of the men failed their driving test?
- c Work out the pass : fail ratio.

- A **grouped frequency table** is a way of organizing **large sets of numerical data** by putting the numbers into **groups (intervals)** instead of listing each value separately.
- There are two types of data look at the key point to the right.

Q6 page 125

6 The numbers of people using a supermarket each day in one month were

123, 179, 235, 189, 207, 199, 145, 154, 198, 132, 201, 99, 134, 245, 207, 198, 164, 157, 149, 183, 172, 175, 188, 192, 184, 167, 203, 201, 188, 181

- a Is the data **discrete** or **continuous**?
- b Design a grouped frequency table to record this data.
- c Which group has the highest frequency?

a) _____

b)

c) _____

Key point



Discrete data can only take particular values. For example, dress sizes are usually only even numbers. For discrete data you can use groups like 1–10, 11–20, ...

Continuous data is measured and can take any value. Length, mass and capacity are continuous. For continuous data there are no gaps between the groups. You must use the \leq and $<$ symbols.

Q9 page 126

9 a Explain what is wrong with each question and rewrite them.

i How many portions of fruit or vegetables do you eat a day?

0–2 4–6 6–8

ii Do you eat a healthy diet?

Yes No

iii Do you agree that fruit and vegetables are good for you?

Yes No

i) _____

ii) _____

iii) _____

➤ **A hypothesis:** is a statement that you can test by collecting data.

Examples of Hypotheses

- “Students in Grade 8 spend more time on math homework than students in Grade 7.”
- “Most students prefer pizza over burgers.”
- “Boys are more likely to play sports during break than girls.”

Investigation

Problem-solving

Choose a topic that you are interested in. It could be sport, music, school activities – anything you like.

- 1 Write a hypothesis related to your topic.
- 2 Design a questionnaire to collect the information you need for testing your hypothesis.
- 3 Test your questionnaire on a friend.
- 4 Collect data from a suitable sample.
- 5 Record your findings in a two-way table or spreadsheet.



You can refer back to question 4 in the book page 139 to help you with the questionnaire.

1 The frequency table shows how many siblings (brothers or sisters) each pupil in a class has.

Number of siblings	0	1	2	3	4
Frequency	5	13	7	4	1

a Work out the mean. b How many pupils have 2 siblings?

➤ **Estimated mean and range for grouped-frequency table**

Key point



When data has been grouped, you cannot work out the exact mean but you can work out an **estimate**.

Worked example



The table shows the results of a survey into the lengths of long rivers in the United Kingdom.

Work out an **estimate** for a the mean b the range.

a	Length of river, L (km)	Frequency	Midpoint of class (km)	Midpoint \times frequency
	$110 \leq L < 140$	17	125	2125
	$140 \leq L < 170$	5	155	775
	$170 \leq L < 200$	3	185	555
	$200 \leq L < 230$	2	215	430
	$230 \leq L < 260$	1	245	245
	$260 \leq L < 290$	0	275	0
	$290 \leq L < 320$	1	305	305
	$320 \leq L < 350$	1	335	335
	$350 \leq L < 380$	1	365	365
	Total	31	Total	5135

You don't know the exact value of each length, so estimate it as the midpoint of each class. Draw a column for the midpoints. Calculate an estimate of the total length for each class (midpoint \times frequency).

Calculate the total number of rivers and an estimate for the sum of their lengths.

$$\text{Estimated mean} = \text{Estimated sum of lengths} \div \text{total number of rivers}$$

$$= 5135 \div 31$$

$$= 165.65 \text{ km (to 2 d.p.)}$$

From the frequency table, the smallest possible value is 110 km and the largest possible value is 380 km.

$$\text{b An estimate of the range is } 380 - 110 = 270 \text{ km}$$

Q2 page 128

2 The table shows the results of a survey into the lengths of river systems in Europe.

Work out an estimate for

- a the range
- b the mean.

Length of river, L (miles)	Frequency		
$0 \leq L < 300$	5		
$300 \leq L < 600$	13		
$600 \leq L < 900$	8		
$900 \leq L < 1200$	1		
$1200 \leq L < 1500$	2		
$1500 \leq L < 1800$	1		
$1800 \leq L < 2100$	0		
$2100 \leq L < 2400$	1		

Q3 page 128

3 **Problem-solving** Class 9Y carried out a survey to find out how many miles people drive each year. The table shows their results.

Work out an estimate for

- a the range
- b the mean.

Distance driven each year, d (thousands of kilometres)	Frequency		
$0 < d \leq 5$	9		
$5 < d \leq 10$	26		
$10 < d \leq 15$	35		
$15 < d \leq 20$	20		
$20 < d \leq 25$	9		
$25 < d \leq 30$	1		

Instead, you can find out which group the median value is in .

Steps to find the group the median value is in.

1. Add all the frequencies to find the **total number of values (N)**.
2. Find $N \div 2$.
3. Work out the cumulative frequency (go down the frequencies, adding them up as you go along).
4. Use the cumulative frequency to find where the position is.

Q4 page 128

4 The table shows the results of a survey into the lengths of rivers in North America.

Length of river, L (miles)	Frequency			
$0 < L \leq 400$	21			
$400 < L \leq 800$	18			
$800 < L \leq 1200$	4			
$1200 < L \leq 1600$	3			
$1600 < L \leq 2000$	3			
$2000 < L \leq 2400$	2			
$2400 < L \leq 2800$	1			
$2800 < L \leq 3200$	2			

- a How many items are there in the sample?
- b Which number item is the median?
- c In which group will the median value be?
- d What is the modal class?
- e Work out an estimate for
 - i the mean
 - ii the range.

Q5 page 139

5 In a survey into eating habits, students recorded the number of portions of fruit and vegetables they ate for a week.
The results are shown in the table.

Number of portions per week	Frequency			
0–9	6			
10–19	17			
20–29	15			
30–39	48			
40–49	21			
50–59	3			

a Work out an estimate for the mean number of portions eaten each week.
b Which class contains the median number of portions?

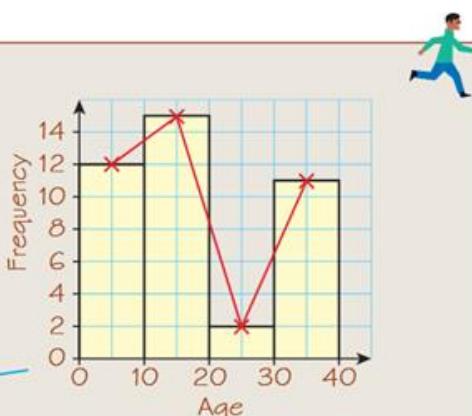
Frequency polygon

Worked example

Draw a frequency polygon to represent this data.

Age, a	Frequency
$0 \leq a < 10$	12
$10 \leq a < 20$	15
$20 \leq a < 30$	2
$30 \leq a < 40$	11

First draw a frequency diagram. Then join the midpoints of the tops of bars.



Key point

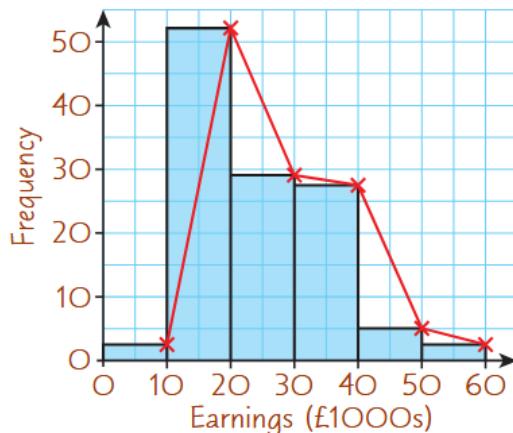
You can draw a frequency polygon by joining the midpoints of the tops of the bars in a frequency diagram.

Q5 page 130

5 Samarah constructs a frequency polygon for this data.

Earnings (per year), e	Number of employees
$0 < e \leq £10000$	3
$£10000 < e \leq £20000$	52
$£20000 < e \leq £30000$	29
$£30000 < e \leq £40000$	27
$£40000 < e \leq £50000$	5
$£50000 < e \leq £60000$	3

a Explain what she has done wrong.
 b Construct an accurate frequency polygon for the data.

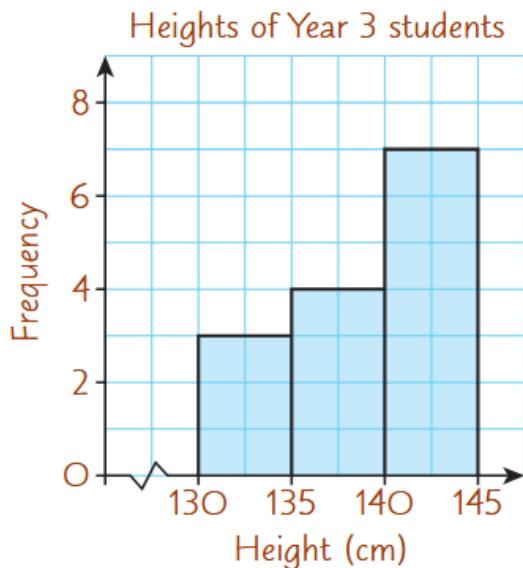


Q1+ 2 page 136

1 A nurse recorded the heights of 14 children in Year 3.

Height, h (cm)	Frequency
$130 < h \leq 135$	3
$135 < h \leq 140$	4
$140 < h \leq 145$	7

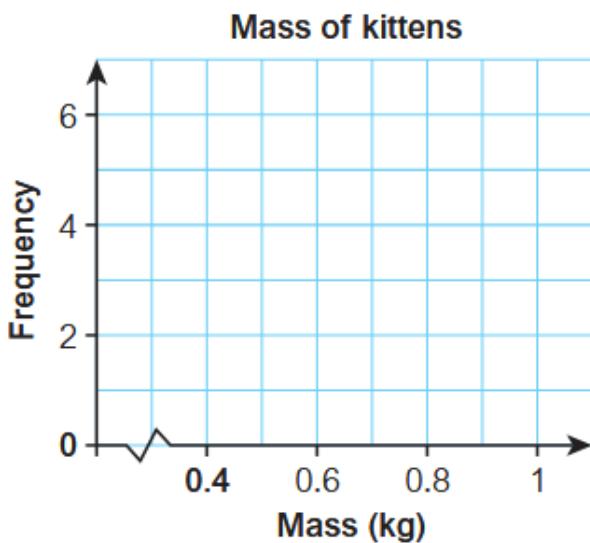
Polly is drawing a frequency polygon for this data.
 She starts by drawing the bars.
 What does Polly need to do next?



2 A vet recorded the masses of some kittens in a litter.

Mass, m (kg)	Midpoint	Frequency
$0.4 \leq m < 0.6$	0.5	1
$0.6 \leq m < 0.8$	□	6
$0.8 \leq m < 1$	□	2

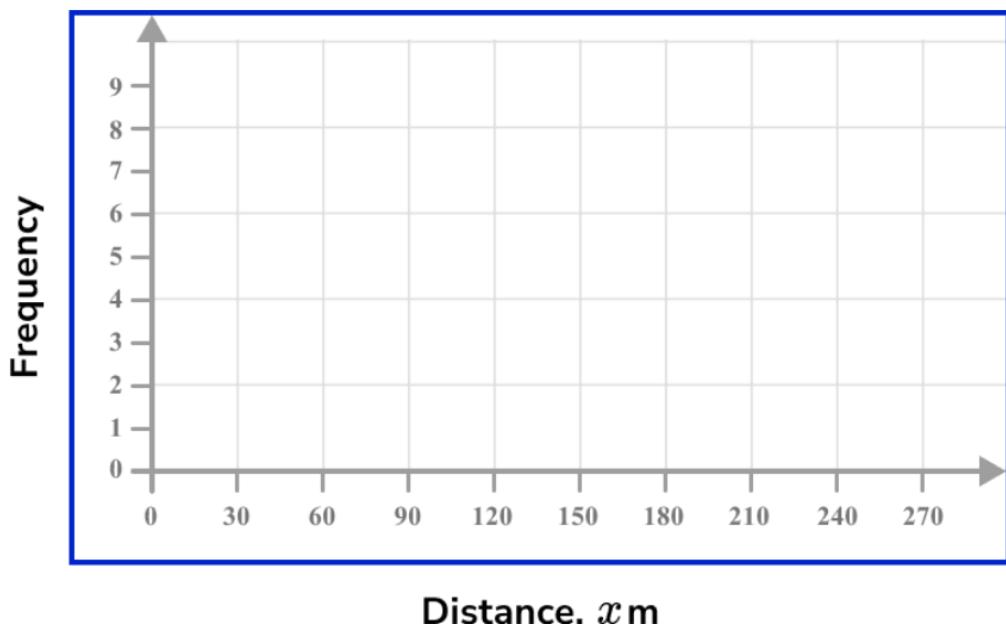
- Copy the axes.
- Draw a frequency polygon for this data.



Example:

18 students were asked to hop on one leg as far as they can in 3 minutes. The results were grouped into the table below. Draw a frequency polygon on the axes provided to represent this data.

Distance, x metres	Frequency
$0 < x \leq 60$	5
$60 < x \leq 120$	8
$120 < x \leq 180$	1
$180 < x \leq 240$	4

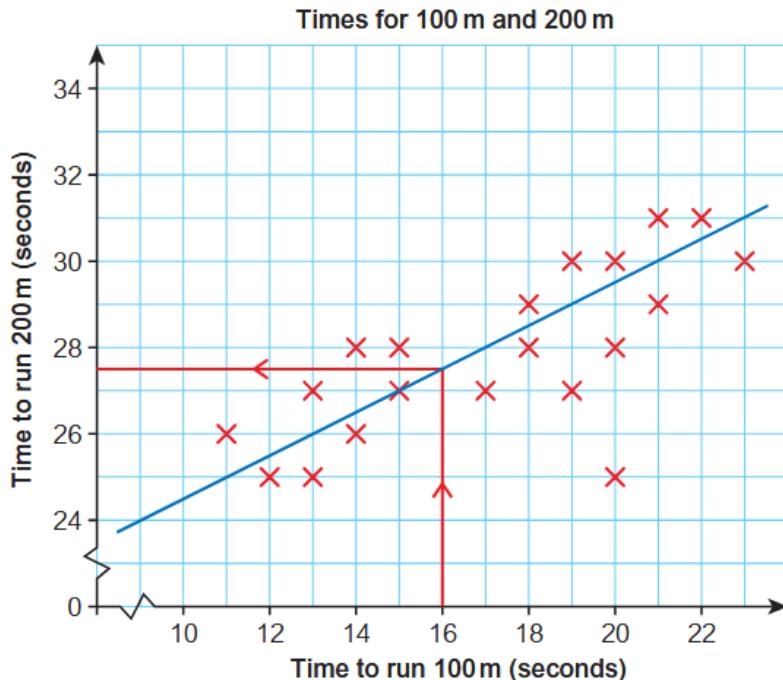


The line of best fit

You can find the explanation and details of this topic in (**Unit 10 grade 7 worksheet (1)**).

Q3page 136

3 The scatter graph shows the times taken for 20 students to run 200 m plotted against their times taken to run 100 m.



a Describe the type of correlation.

b **Modelling** Use the line of best fit to predict the 200 m time for a student who ran the 100 m in

- 16 seconds
- 19 seconds
- 17 seconds

c One point is an outlier. What were the times for 100 m and 200 m for this student?

Merry Christmas

Teacher: Sally Serkisian