# Mathematics

## Edexcel IAL

# S1

## Worksheets

### Representations of Data

**Eng. Nagy Elraheb**

# Representations of data

## Exercise 1:

**1** The data in the table show the mass, in kilograms, of 50 adult puffer fish.

  **a** Draw a histogram for these data.

  **b** On the same set of axes, draw a frequency polygon.

| Mass, $m$ (kg) | Frequency |
|---|---|
| $10 \leqslant m < 15$ | 4 |
| $15 \leqslant m < 20$ | 12 |
| $20 \leqslant m < 25$ | 23 |
| $25 \leqslant m < 30$ | 8 |
| $30 \leqslant m < 35$ | 3 |

**2** Some students took part in an obstacle race. The time for each student to complete the race was noted. The results are shown in the histogram.

  **a** Give a reason to justify the use of a histogram to represent these data.

90 students took between 60 and 70 seconds to complete the race.

  **b** Find the number of students who took between 40 and 60 seconds.

  **c** Find the number of students who took 80 seconds or less.

  **d** Calculate the total number of students who took part in the race.
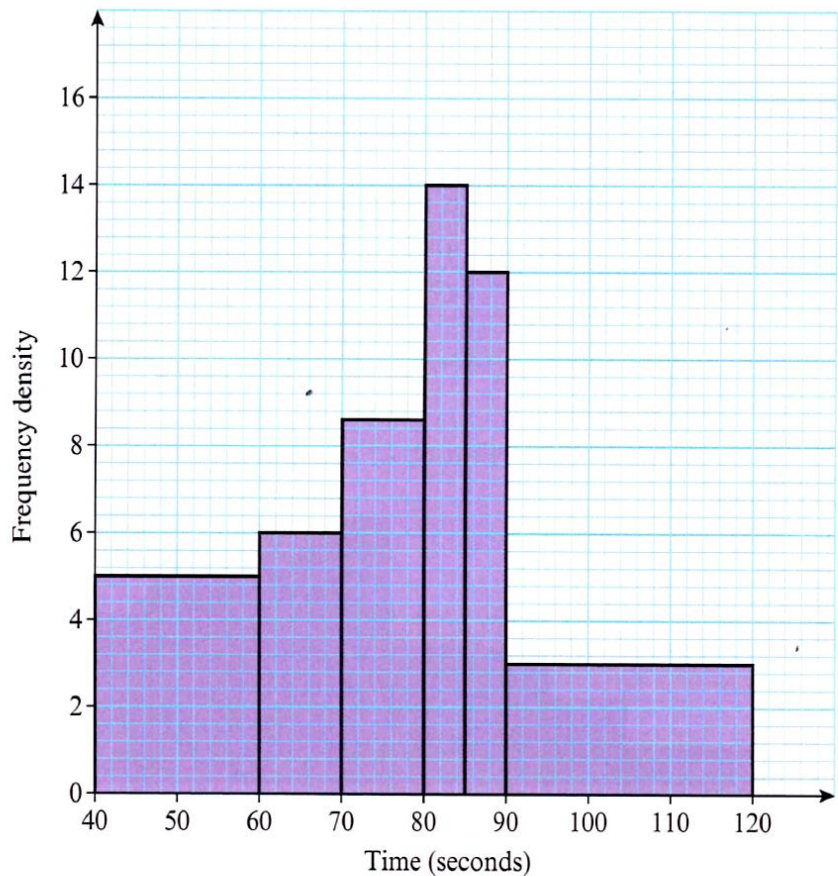


**Watch out** Frequency density × class width is always **proportional** to frequency in a histogram, but not necessarily **equal** to frequency.

3 A Fun Day committee at a local sports centre organised a tennis ball throwing competition. The distance thrown by every competitor was recorded and the data is shown in the histogram. Forty people threw the ball less than 20 m.
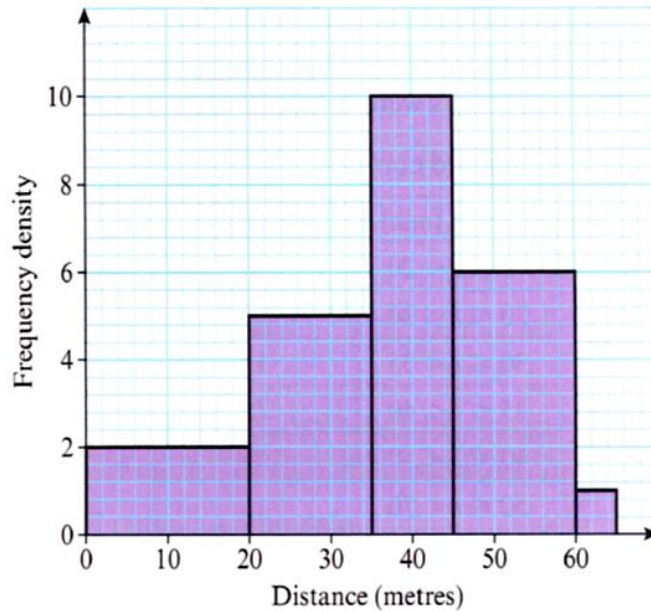
a Why is a histogram a suitable diagram to represent these data?

b How many people entered the competition?

c Estimate how many people threw between 30 and 40 metres.

d How many people threw between 45 and 65 metres?

e Estimate how many people threw less than 25 metres.



4 A farmer found the masses of a random sample of lambs. The masses were summarised in a grouped frequency table and represented by a histogram. The frequency for the class $28 \leqslant m < 32$ was 32.
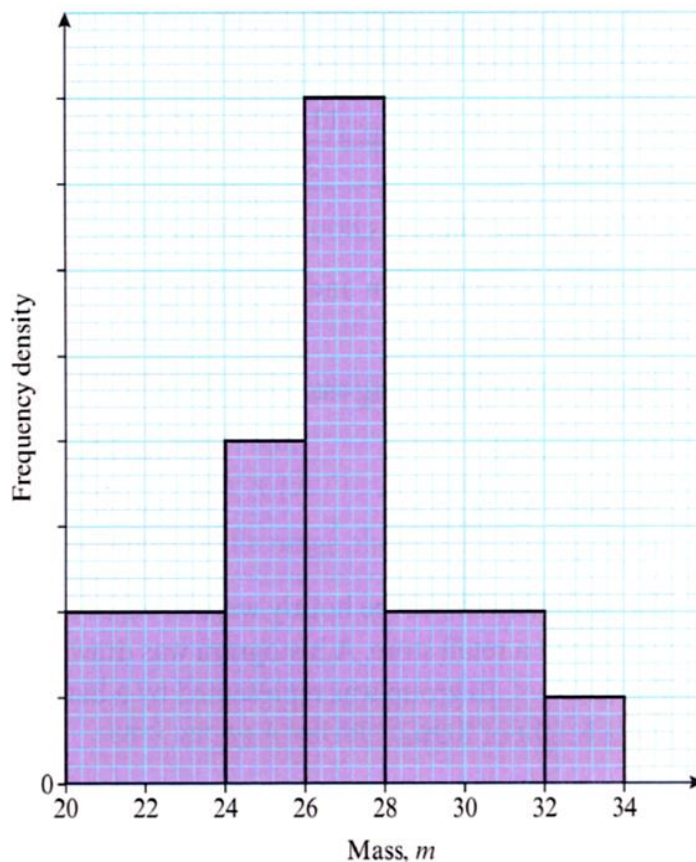
a Show that 25 small squares on the histogram represents 8 lambs.

b Find the frequency of the $24 \leqslant m < 26$ class.

c How many lambs did the farmer weigh in total?

d Estimate the number of lambs that had masses between 25 and 29 kg.



**Problem-solving**

You can use area to solve histogram problems where no vertical scale is given. You could also use the information given in the question to work out a suitable scale for the vertical axis.

**5** The partially (not entirely) completed histogram shows the time, in minutes, that passengers were delayed at an airport.

**a i** Copy and complete the table.

| Time, $t$ (min) | Frequency |
|---|---|
| $0 \leqslant t < 20$ | 4 |
| $20 \leqslant t < 30$ | |
| $30 \leqslant t < 35$ | 15 |
| $35 \leqslant t < 40$ | 25 |
| $40 \leqslant t < 50$ | |
| $50 \leqslant t < 70$ | |
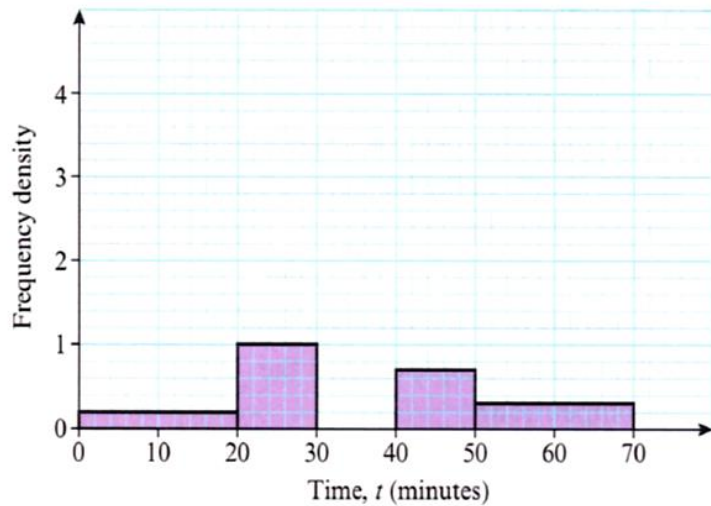


**ii** Copy and complete the histogram. **(4 marks)**

**b** Estimate how many passengers were delayed between 25 and 38 minutes. **(2 marks)**

**6** The **variable** $y$ was measured to the nearest whole number. 60 observations were taken and are recorded in the table below.

| $y$ | 10–12 | 13–14 | 15–17 | 18–25 |
|---|---|---|---|---|
| Frequency | 6 | 24 | 18 | 12 |

**a** Write down the class boundaries for the 13–14 class. **(1 mark)**

A histogram was drawn and the bar representing the 13–14 class had a width of 4 cm and a height of 6 cm.

For the bar representing the 15–17 class, find:

**Problem-solving**

**b i** the width **(1 mark)**

**ii** the height. **(2 marks)**

Remember that area is proportional to frequency.

**7** The table here shows the hourly wage in euros for 31 employees working at a retail store. A histogram was drawn using this data. The $8 \leqslant w < 10$ group was represented by a bar of width 1 cm and a height of 8 cm.

| Hourly wage, $w$ (€) | Frequency |
|---|---|
| $4 \leqslant w < 8$ | 4 |
| $8 \leqslant w < 10$ | 8 |
| $10 \leqslant w < 11$ | 6 |
| $11 \leqslant w < 12$ | 7 |
| $12 \leqslant w < 15$ | 5 |
| $15 \leqslant w < 16$ | 1 |

**a** Find the width and height of the bar representing the $10 \leqslant w < 11$ group. **(2 marks)**

**b** Estimate the mean and standard deviation of the employees working at the store. **(3 marks)**

**c** Use linear **interpolation** to find an estimate for the lower quartile of wages. **(2 marks)**

**d** Estimate how many employees had an hourly wage higher than the mean plus one standard deviation. **(2 marks)**

**1** Some data are collected. $Q_1 = 46$ and $Q_3 = 68$.

A value greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or less than $Q_1 - 1.5 \times (Q_3 - Q_1)$ is defined as an outlier.

Using this rule, work out whether or not the following are outliers:

**a** 7          **b** 88          **c** 105

**2** The masses of male and female turtles are given in grams. For males, the lower quartile was 400 g and the upper quartile was 580 g. For females, the lower quartile was 260 g and the upper quartile was 340 g.

An outlier is an observation that falls either $1 \times$ the interquartile range above the upper quartile or $1 \times$ the interquartile range below the lower quartile.

**a** Which of these male turtle masses would be outliers?

    400 g    260 g    550 g    640 g

**b** Which of these female turtle masses would be outliers?

    170 g    300 g    340 g    440 g

> **Hint** The definition of an outlier here is different from that in question 1. You will be told which rule to use in the exam.

**c** What is the largest mass a male turtle can be without being an outlier?

**3** The masses of arctic foxes are found and the mean mass was 6.1 kg. The variance was 4.2.

An outlier is an observation which lies $\pm 2$ standard deviations from the mean.

**a** Which of these arctic fox masses are outliers?

    2.4 kg    10.1 kg    3.7 kg    11.5 kg

**b** What are the smallest and largest masses that an arctic fox can be without being an outlier?

**4** The ages of nine people at a children's birthday party are recorded. $\Sigma x = 92$ and $\Sigma x^2 = 1428$.

**a** Calculate the mean and standard deviation of the ages. **(3 marks)**

An outlier is an observation which lies $\pm 2$ standard deviations from the mean.

One of the ages is recorded as 30.

**b** State, with a reason, whether or not this is an outlier. **(2 marks)**

**c** Suggest a reason why this age could be a legitimate data value. **(1 mark)**

**d** Given that all nine people were children, clean the data and recalculate the mean and standard deviation. **(3 marks)**

> **Problem-solving**
>
> After you clean the data you will need to find the new values for $n$, $\Sigma x$ and $\Sigma x^2$.

## Exercise 3:

**1** A group of students took a test. The summary data are shown in the table.

| Lowest mark | Lower quartile | Median | Upper quartile | Highest mark |
|:---:|:---:|:---:|:---:|:---:|
| 5 | 21 | 28 | 36 | 58 |

Given that there were no outliers, draw a box plot to illustrate these data.

**2** Here is a box plot of marks in an examination.

  **a** Write down the upper and lower quartiles.

  **b** Write down the median.

  **c** Work out the interquartile range.

  **d** Work out the range.



Marks
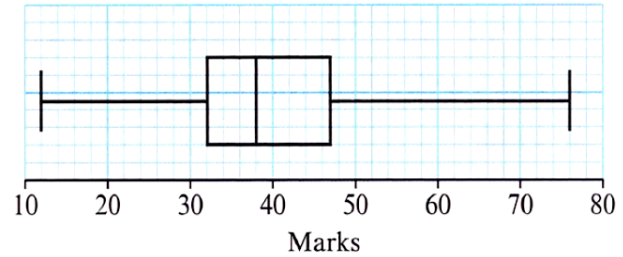
**3** The masses of male and female turtles are given in grams. The data are summarised in the box plots.

  **a** Compare and contrast the masses of the male and female turtles.

  **b** A turtle was found to have a mass of 330 grams. State whether it is likely to be a male or a female. Give a reason for your answer.

  **c** Write down the size of the largest female turtle.



Mass (g)
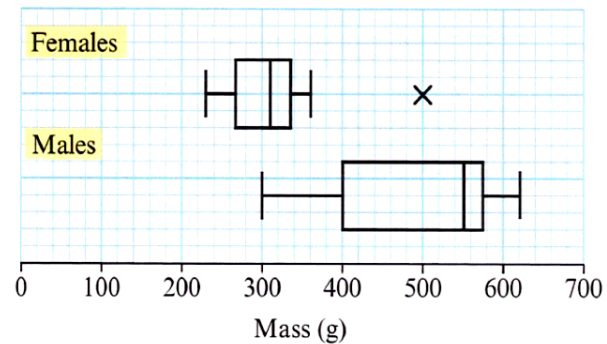
**4** The average weight (in kg) for 30 different breeds of dog are shown below.

| | | | | |
|:---:|:---:|:---:|:---:|:---:|
| 13 | 15 | 16 | 19 | 20 |
| 21 | 22 | 22 | 24 | 24 |
| 25 | 25 | 26 | 26 | 26 |
| 27 | 29 | 29 | 30 | 30 |
| 33 | 33 | 38 | 46 | 48 |

  **a** Calculate $Q_1$, $Q_2$ and $Q_3$.         **(3 marks)**

An outlier is defined as a value which lies either $1.5 \times$ the interquartile range above the upper quartile, or $1.5 \times$ the interquartile range below the lower quartile.

  **b** Show that 46 kg and 48 kg are outliers.         **(1 mark)**

  **c** Draw a box plot for this data.         **(3 marks)**

**1** Thirty college students were asked how many movies they had in their collection.
The results are as follows:

| 12 | 25 | 34 | 17 | 12 | 18 | 29 | 34 | 45 | 6 | 15 | 9 | 25 | 3 | 29 |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 22 | 20 | 32 | 15 | 15 | 19 | 12 | 26 | 27 | 27 | 32 | 35 | 42 | 26 | 25 |

Draw a stem and leaf diagram to represent these data.

**a** Find the median.

**b** Find the lower quartile.

**c** Find the upper quartile.

**2** The following stem and leaf diagram shows some information about the marks gained by a group of students in a statistics test.

| Stem | Leaf | | Key: 2\|3 means 23 marks |
|------|------|---|------|
| **0** | 8  9 | | (2) |
| **1** | 2  5  5  9 | | (4) |
| **2** | 3  6  6  6  7 | | (5) |
| **3** | 4  4  5  7  7  7  7  7  9 | | (9) |
| **4** | 5  8  8  9 | | (4) |

**a** Write down the highest mark.

**b** Write down the lowest mark.

**c** How many students scored 26 marks?

**d** What is the modal mark?

**e** Find the median.

**f** Find the lower quartile.

**g** Find the upper quartile.

**3** A class of 16 boys and 13 girls completed a Physics test. The test was marked out of 60.
Their marks are shown below:

| | Boys | | | | Girls | | |
|----|----|----|----|----|----|----|----|
| 45 | 54 | 32 | 60 | 26 | 54 | 47 | 32 |
| 28 | 34 | 54 | 56 | 34 | 34 | 45 | 46 |
| 32 | 29 | 47 | 48 | 39 | 52 | 24 | 28 |
| 44 | 45 | 56 | 57 | 33 | | | |

**a** Draw back to back stem and leaf diagrams to represent these data.

**b** Comment on your results.

**4** The stem and leaf diagram below shows the median age, in years, of a selection of African elephants in Tanzania.

| Stem | Leaf |
|------|------|
| 1 | 5  6  6  6  7  7  7  7  7  7  7  8  8  8  8  8  8  8  8  8  9  9  9  9  9  9  9  9  9  9  9  9 |
| 2 | 0  0  0  0  0  0  0  1  1  1  1  1  3  3  3  4  5  7 |
| 3 | 4  4 |
| 4 | 1 |

Find:

**Key: 1 | 8 = 18 years**

**a** the median

**b** the interquartile range and any outliers.

## Exercise 5:

**1** In a survey of the earnings of some college students who worked weekend jobs, the median wage was $36.50. The 75th percentile was $45.75 and the interquartile range was $30.50. Use the quartiles to describe the skewness of the distributions.

**2** A group of estate agents recorded the time spent on the first meeting with a random sample of 120 of their clients. The mean time spent with their clients is 31.1 minutes and the variance is 78.05. The median time is 29.7 minutes and $Q_1$ and $Q_3$ values are 25.8 minutes and 34.8 minutes.

One measure of skewness is found using $\dfrac{3(\text{mean} - \text{median})}{\text{standard deviation}}$

**a** Evaluate this measure and describe the skewness of the data

The estate agents are undecided whether to use the median and quartiles, or the mean and standard deviation to summarise the data.

**b** State, giving a reason, which you would recommend for them to use.

**3** The following stem and leaf diagram summarises the wing length, to the nearest mm, of a random sample of 67 birds.

| Wing length | | | | | | | | | | | | | Key: 5\|0 means 50 mm |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **5** | 0 0 0 1 1 2 2 3 3 3 4 4 | | | | | | | | | | | | (12) |
| **5** | 5 5 6 6 6 7 8 8 9 9 | | | | | | | | | | | | (10) |
| **6** | 0 1 1 1 3 3 4 4 4 4 | | | | | | | | | | | | (10) |
| **6** | 5 5 6 7 8 9 9 | | | | | | | | | | | | (7) |
| **7** | 1 1 2 2 3 3 | | | | | | | | | | | | (6) |
| **7** | 5 7 9 9 | | | | | | | | | | | | (4) |
| **8** | 1 1 1 2 2 3 3 4 | | | | | | | | | | | | (8) |
| **8** | 7 8 9 | | | | | | | | | | | | (3) |
| **9** | 0 1 1 2 | | | | | | | | | | | | (4) |
| **9** | 5 7 9 | | | | | | | | | | | | (3) |

**a** Write down the mode.

**b** Find the median and quartiles of the data.

**c** Construct a box plot to represent the data.

**d** Comment on the skewness of the distribution.

**e** Calculate the mean and standard deviation for the data.

**f** Use another method to shows that the data is skewed.

**g** State, giving a reason, which of **b** or **e** would you recommend using to summarise the data in the diagram.

## Exercise 6:

**1** The box plots below show the distribution of speeds of cars on two motorways.



Speed (kilometres per hour)
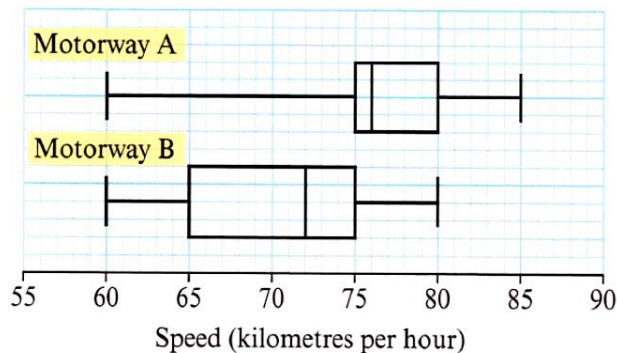
Compare the distributions of the speeds on the two motorways.

**2** Two classes of primary school children complete a puzzle. Summary statistics for the times the children took, in minutes, are shown in the table.

|  | $n$ | $\Sigma x$ | $\Sigma x^2$ |
|---|---|---|---|
| **Class 2B** | 20 | 650 | 22 000 |
| **Class 2F** | 22 | 598 | 19 100 |

Calculate the mean and standard deviation of the times and compare the distributions.

**3** The stem and leaf diagram below shows the age, in years, of the members of a sports club.

```
        Male                        Female
    8 8 7 6 | 1 | 6 6 6 7 7 8 8 9
7 6 5 5 3 3 2 1 | 2 | 1 3 3 4 5 7 8 8 9 9
    9 8 4 4 3 | 3 | 2 3 3 4 7
        5 2 1 | 4 | 0 1 8
          9 0 | 5 | 0
```

**Key: 1|4|0** represents a male aged 41 and a female aged 40

**a** Find the median and interquartile range for the males.

**b** The median and interquartile ranges for the females are 27 and 15 respectively.

Make two comparisons between the ages of the males and females.

**4** In the box plots here, the marks for a group of students taking their Mathematics exam for the first time are shown on the top.

The marks for a group of students who are retaking their Mathematics exam are shown on the bottom.

Compare and contrast the marks between the two groups taking the exam. **(3 marks)**



Students taking the exam for the first time

Students retaking their exam