**1.** Sara is investigating the variation in daily maximum gust, $t$ kn, for Camborne in June and July 1987.

She used the large data set to select a sample of size 20 from the June and July data for 1987. Sara selected the first value using a random number from 1 to 4 and then selected every third value after that.

(a) State the sampling technique Sara used.

**(1)**

(b) From your knowledge of the large data set, explain why this process may not generate a sample of size 20.

**(1)**

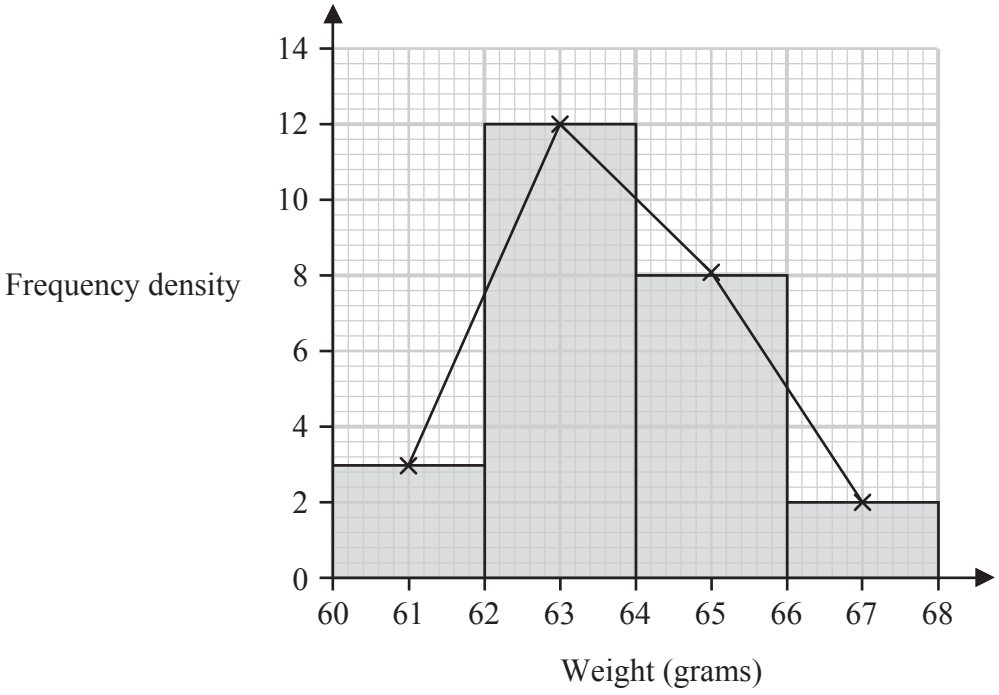
The data Sara collected are summarised as follows

$$n = 20 \qquad \sum t = 374 \qquad \sum t^2 = 7600$$

(c) Calculate the standard deviation.

**(2)**

**2.** The histogram and its frequency polygon below give information about the weights, in grams, of 50 plums.




(a) Show that an estimate for the mean weight of the 50 plums is 63.72 grams.

**(2)**

(b) Calculate an estimate for the standard deviation of the 50 plums.

**(2)**

Later it was discovered that the scales used to weigh the plums were broken.

Each plum actually weighs 5 grams less than originally thought.

(c) State the effect this will have on the estimate of the standard deviation in part (b). Give a reason for your answer.

**(1)**

**3.**

A lake contains three different types of carp.

There are an estimated 450 mirror carp, 300 leather carp and 850 common carp.

Tim wishes to investigate the health of the fish in the lake.

He decides to take a sample of 160 fish.

(a) Give a reason why stratified random sampling cannot be used.

**(1)**

(b) Explain how a sample of size 160 could be taken to ensure that the estimated populations of each type of carp are fairly represented.

You should state the name of the sampling method used.

**(2)**

As part of the health check, Tim weighed the fish.

His results are given in the table below.

| **Weight ($w$ kg)** | **Frequency (f)** | **Midpoint ($m$ kg)** |
|---|---|---|
| $2 \leqslant w < 3.5$ | 8 | 2.75 |
| $3.5 \leqslant w < 4$ | 32 | 3.75 |
| $4 \leqslant w < 4.5$ | 64 | 4.25 |
| $4.5 \leqslant w < 5$ | 40 | 4.75 |
| $5 \leqslant w < 6$ | 16 | 5.5 |

(You may use $\sum fm = 692$ and $\sum fm^2 = 3053$)

(c) Calculate an estimate for the standard deviation of the weight of the carp.

**(2)**

Tim realised that he had transposed the figures for 2 of the weights of the fish.

He had recorded in the table 2.3 instead of 3.2 and 4.6 instead of 6.4

(d) Without calculating a new estimate for the standard deviation, state what effect

(i) using the correct figure of 3.2 instead of 2.3

(ii) using the correct figure of 6.4 instead of 4.6

would have on your estimated standard deviation.

Give a reason for each of your answers.

**(2)**

**Question 3 continued**

**4.** Helen is studying the daily mean wind speed for Camborne using the large data set from 1987. The data for one month are summarised in Table 1 below.

| **Windspeed** | n/a | 6 | 7 | 8 | 9 | 11 | 12 | 13 | 14 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Frequency** | 13 | 2 | 3 | 2 | 2 | 3 | 1 | 2 | 1 | 2 |

**Table 1**

(a) Calculate the mean for these data.

**(1)**

(b) Calculate the standard deviation for these data and state the units.

**(2)**

The means and standard deviations of the daily mean wind speed for the other months from the large data set for Camborne in 1987 are given in Table 2 below. The data are not in month order.

| **Month** | *A* | *B* | *C* | *D* | *E* |
|---|---|---|---|---|---|
| **Mean** | 7.58 | 8.26 | 8.57 | 8.57 | 11.57 |
| **Standard Deviation** | 2.93 | 3.89 | 3.46 | 3.87 | 4.64 |

**Table 2**

(c) Using your knowledge of the large data set, suggest, giving a reason, which month had a mean of 11.57

**(2)**

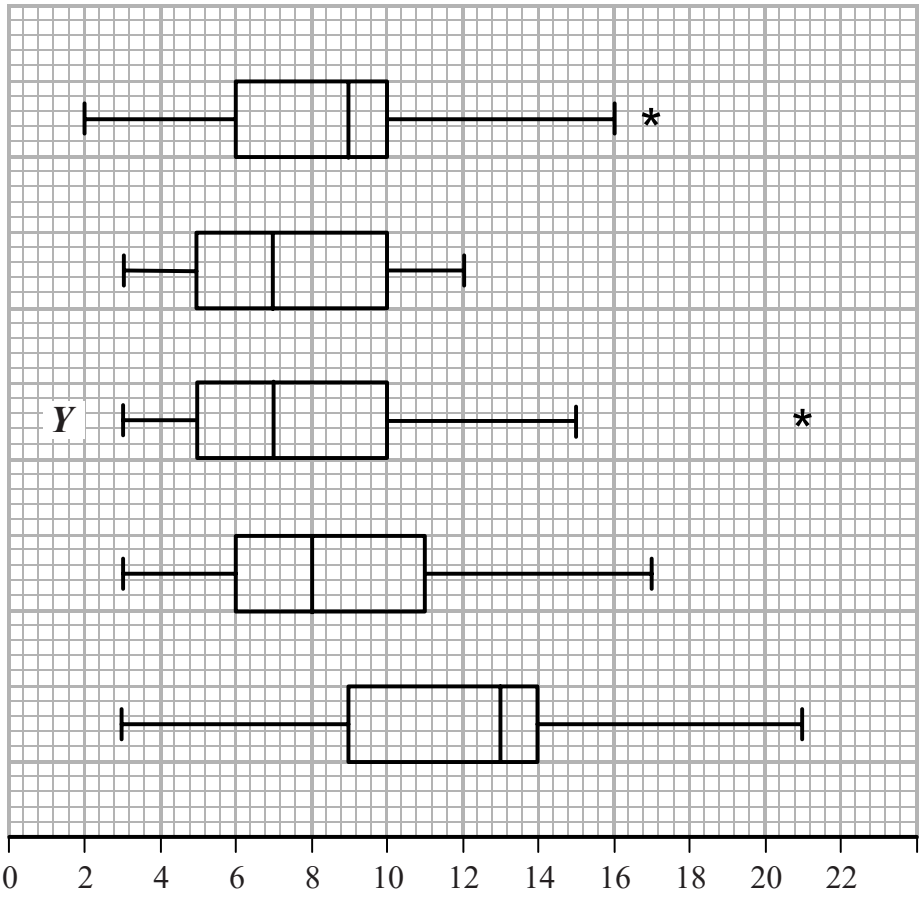
The data for these months are summarised in the box plots on the opposite page.
They are not in month order or the same order as in Table 2.

(d) (i) State the meaning of the * symbol on some of the box plots.

(ii) Suggest, giving your reasons, which of the months in Table 2 is most likely to be summarised in the box plot marked *Y*.

**(3)**

**Question 4 continued**



**5.**

Joshua is investigating the daily total rainfall in Hurn for May to October 2015

Using the information from the large data set, Joshua wishes to calculate the mean of the daily total rainfall in Hurn for May to October 2015

(a) Using your knowledge of the large data set, explain why Joshua needs to clean the data before calculating the mean.

**(1)**

Using the information from the large data set, he produces the grouped frequency table below.

| **Daily total rainfall ($r$ mm)** | **Frequency** | **Midpoint ($x$ mm)** |
|---|---|---|
| $0 \leqslant r < 0.5$ | 121 | 0.25 |
| $0.5 \leqslant r < 1.0$ | 10 | 0.75 |
| $1.0 \leqslant r < 5.0$ | 24 | 3.0 |
| $5.0 \leqslant r < 10.0$ | 12 | 7.5 |
| $10.0 \leqslant r < 30.0$ | 17 | 20.0 |

You may use $\sum fx = 539.75$ and $\sum fx^2 = 7704.1875$

(b) Use linear interpolation to calculate an estimate for the upper quartile of the daily total rainfall.

**(2)**

(c) Calculate an estimate for the standard deviation of the daily total rainfall in Hurn for May to October 2015

**(2)**

(d) (i) State the assumption involved with using class midpoints to calculate an estimate of a mean from a grouped frequency table.

(ii) Using your knowledge of the large data set, explain why this assumption does not hold in this case.

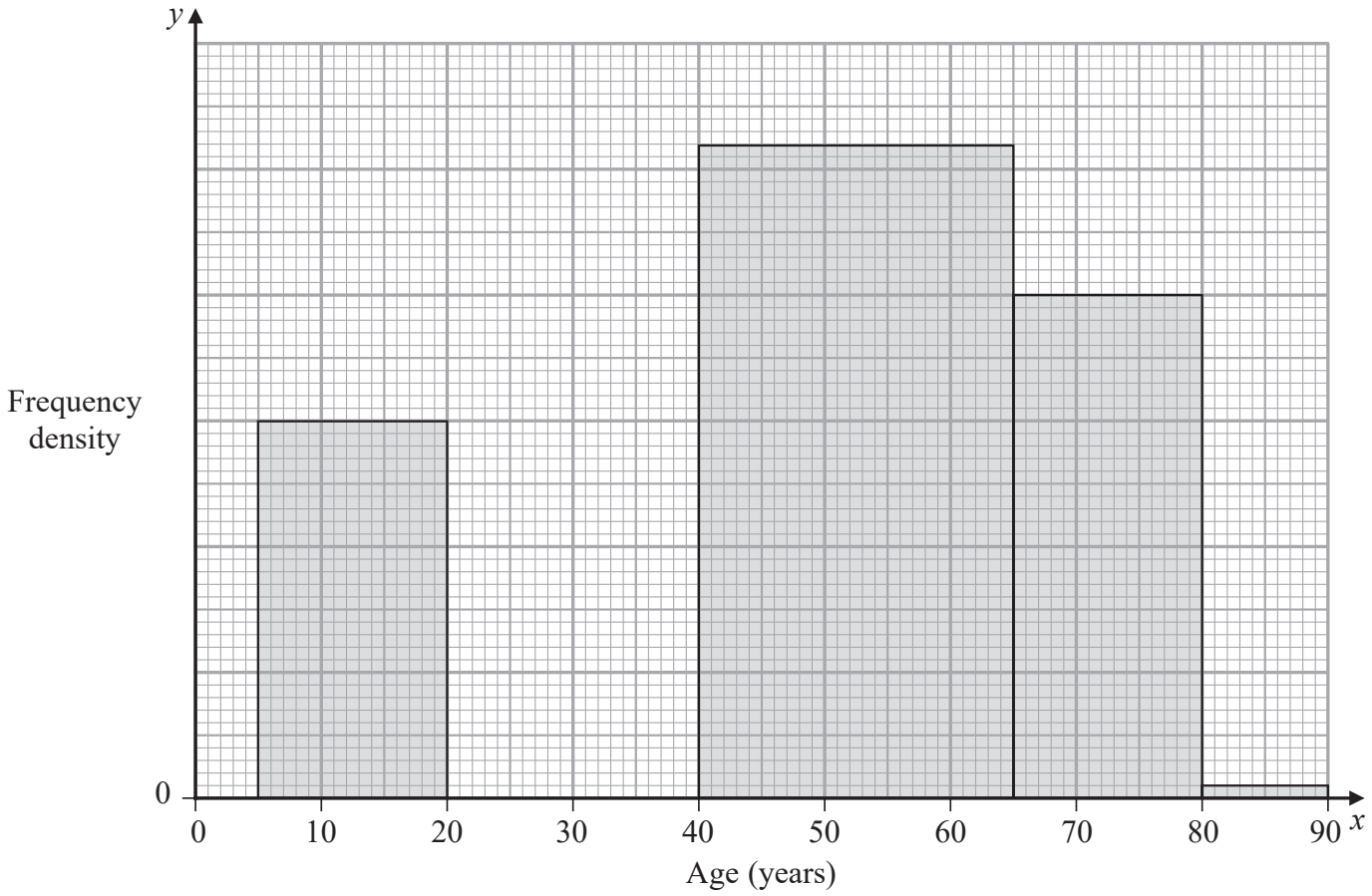
(iii) State, giving a reason, whether you would expect the actual mean daily total rainfall in Hurn for May to October 2015 to be larger than, smaller than or the same as an estimate based on the grouped frequency table.

**(3)**

**6.** The partially completed table and partially completed histogram give information about the ages of passengers on an airline.

There were no passengers aged 90 or over.

| **Age ($x$ years)** | $0 \leqslant x < 5$ | $5 \leqslant x < 20$ | $20 \leqslant x < 40$ | $40 \leqslant x < 65$ | $65 \leqslant x < 80$ | $80 \leqslant x < 90$ |
|---|---|---|---|---|---|---|
| **Frequency** | 5 | 45 | 90 | | | 1 |




(a) Complete the histogram.

**(3)**

(b) Use linear interpolation to estimate the median age.

**(4)**

An outlier is defined as a value greater than $Q_3 + 1.5 \times$ interquartile range.

Given that $Q_1 = 27.3$ and $Q_3 = 58.9$

(c) determine, giving a reason, whether or not the oldest passenger could be considered as an outlier.

**(2)**