Outliers, Box Plots and Cumulative Frequency Diagrams

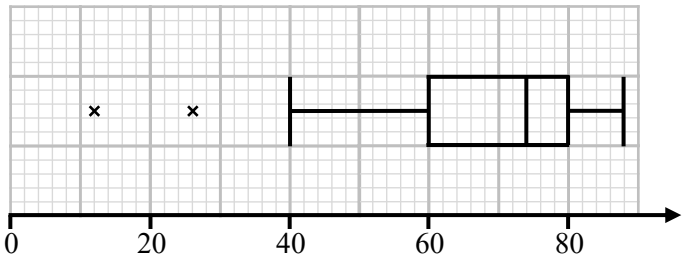
SCAN ME

REVISE THIS TOPIC

CHECK YOUR ANSWERS

SCAN ME

1 180 students completed a maths assessment. The maximum mark for the assessment was 100. The results are summarised in the box plot below.



(a) Write down the median test score. **(1)**

(b) Find the range of the test scores. **(1)**

(c) Find the interquartile range of the test scores. **(1)**

(d) State the meaning of the × symbols shown on the box plot. **(1)**

One of the students is selected at random. Eric says

"the probability that the student scored between 40 and 60 marks is the same the probability that the student scored between 60 and 80 marks."

(e) Explain how you know that Eric must be incorrect. **(1)**

(Total for Question 1 is 5 marks)

2 The table below shows the race times for the 100 metre final at the 2009 World Championships.

Race Position	1	2	3	4	5	6	7	8
Time (seconds)	9.58	9.71	9.84	9.93	9.93	10.00	10.00	10.34

(a) Calculate the interquartile range of the race times. **(1)**

(b) Calculate, to 3 decimal places, the mean race time. **(1)**

(c) Calculate, to 3 decimal places, the standard deviation of the race times. **(1)**

Hannah defines a race time an outlier if it falls either

more than 1.5 × (interquartile range) above the upper quartile or
more than 1.5 × (interquartile range) below the lower quartile.

(d) Determine if any of the race times are considered outliers using Hannah's definition. **(2)**

Ross defines a race time an outlier if it falls either

more than 2 × (standard deviation) above the mean or
more than 2 × (standard deviation) below the mean.

(e) Determine if any of the race times are considered outliers using Ross' definition. **(2)**

(Total for Question 2 is 7 marks)

3 The table below shows the ages of 30 players in a football squad.

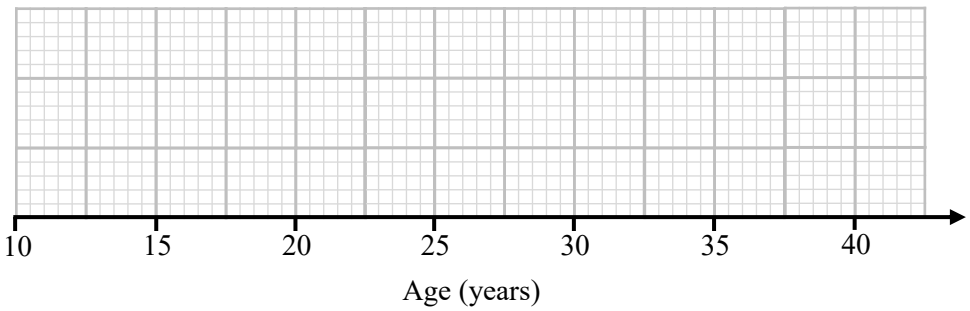
15	17	19	19	20	21	21	22	23	23	23	23	23	23	24
24	24	24	24	25	25	26	27	28	29	29	31	33	35	36

An outlier is any value that falls either

more than 1.5 × (interquartile range) above the upper quartile or
more than 1.5 × (interquartile range) below the lower quartile.

(a) Show that the ages two of the players in the football squad are considered outliers. **(3)**

(b) Draw a box plot for the ages of the players. **(2)**
You should indicate outliers with a ×



(Total for Question 3 is 5 marks)

Solutions

4 A college tracks lateness to school to the nearest minute.
Students that are early or on time are considered to have 0 minutes of lateness.

A sample of 20 students from Year 12 and 20 students from Year 13 were taken on Monday.

The number of minutes lateness to school for the Year 12 students (x) is summarised below.

$$\sum x = 72 \qquad \sum x^2 = 1048$$

(a) Calculate, to 1 decimal place, the mean and standard deviation of the lateness for Year 12 students. **(2)**

An outlier is any value that falls either

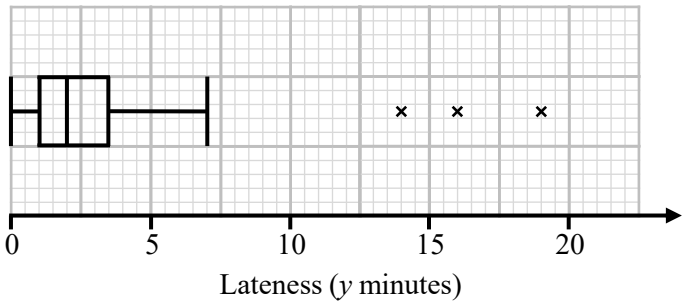
more than 2 × (standard deviation) above the mean or
more than 2 × (standard deviation) below the mean.

The students in Year 12 who were the latest to school are

Merry (13 minutes late) and Perry (26 minutes late)

(b) Show that Perry's lateness is the **only** outlier from the sample of 20 students in Year 12. **(3)**

The box plot below shows the lateness to school for the Year 13 students (y)



(c) State the meaning of the × symbols shown on the box plot. **(1)**

(d) Find the range of the lateness to school for the Year 13 students. **(1)**

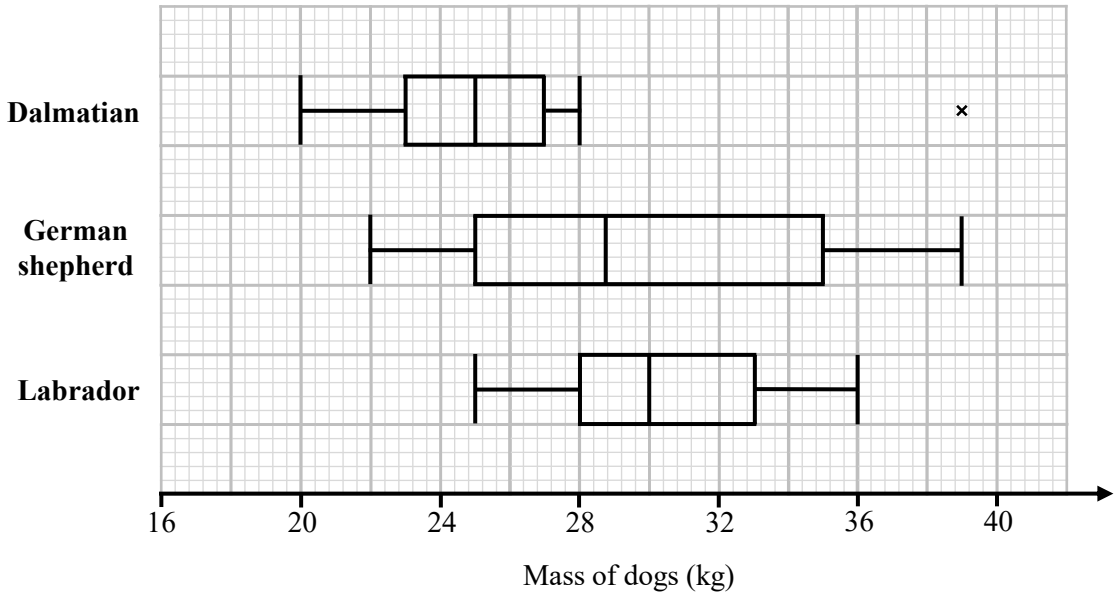
Students who arrive to school 6 or more minutes late receive a detention.
A teacher claims that 5 of the Year 13 students received a detention on Monday for lateness.

(e) Show that the teacher must be incorrect. **(3)**

(Total for Question 4 is 10 marks)

Solutions

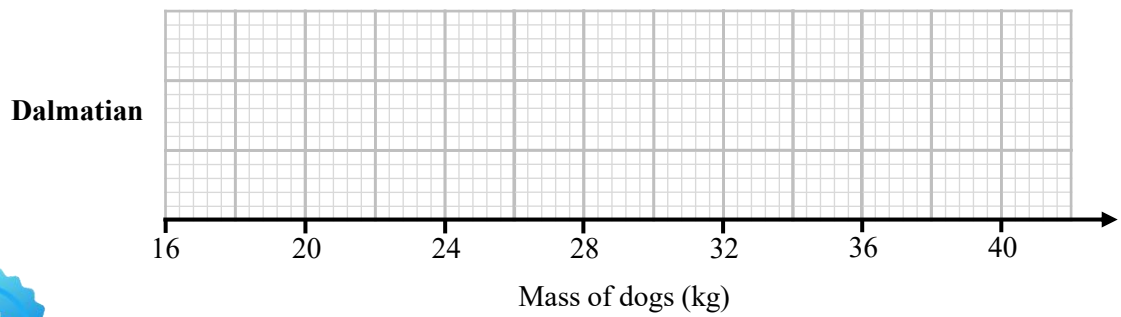
5 A vet measured the masses of 40 dogs of three different breeds.
The results are summarised in the box plots below.



(a) Compare the masses of the German shepherds to the labradors. **(2)**

The heaviest dalmatian had a mass of 29 kg but was incorrectly logged at 39 kg.

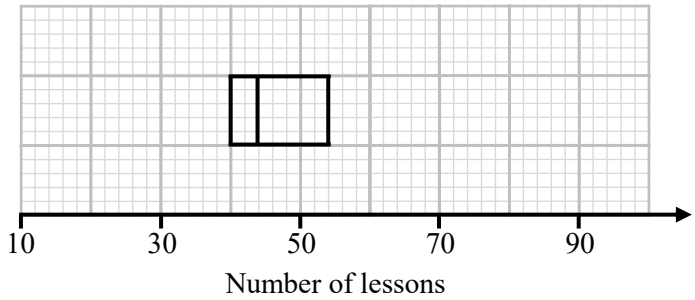
(b) Given that the mass of 29 kg dalmatian is not considered an outlier, draw a corrected box plot for the masses of the 40 dalmatians on the grid below. **(2)**



(Total for Question 5 is 4 marks)

6 A driving instructor recorded the number of lessons before each of their learners were ready to take their driving test in 2023.

The box plot below shows the lower quartile, median and upper quartile for this data.



An outlier is any value that falls either

more than 1.5 × (interquartile range) above the upper quartile or
more than 1.5 × (interquartile range) below the lower quartile.

The learners with the greatest and least number of lessons are shown below.

Learner	A	B	C	D	E	F
Number of Lessons	14	24	25	72	80	85

(a) (i) Determine which of the values above are outliers. **(2)**

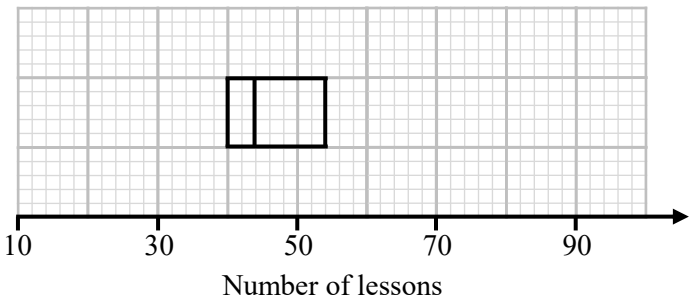
(ii) Hence, complete the box plot on the next page for the number of lessons. **(2)**

The driving instructor also recorded the number of lessons their learners had before they were ready to take their test in 2024. The 2024 data is summarised below.

Q_1	Q_2	Q_3
38	41	50

(b) Compare the number of lessons before learners were ready to take their driving test in 2023 and 2024. **(2)**

Solutions



(Total for Question 6 is 6 marks)

Solutions

7 A student recorded the sound level (S, dB) in a town centre every minute between 7am and 7pm. The results are summarised in the table below.

Sound level (S, dB)	Frequency (f)	Midpoint (S, dB)
$40 \leq S < 50$	99	45
$50 \leq S < 60$	162	55
$60 \leq S < 70$	132	65
$70 \leq S < 80$	129	75
$80 \leq S < 90$	120	85
$90 \leq S < 100$	78	95

You may use $\sum fS = 49230$ and $\sum fS^2 = 3544800$

(a) Calculate, to 3 decimal places, an estimate for the mean sound level. **(1)**

(b) Calculate, to 3 decimal places, an estimate for the standard deviation of the sound levels. **(2)**

An outlier is any value that falls either

more than 2 × (standard deviation) above the mean or
more than 2 × (standard deviation) below the mean.

The minimum sound level recorded was 41 dB.
The maximum sound level recorded was 98 dB.

(c) Use your answers to parts (a) and (b) to show that none of the sound levels are outliers. **(2)**

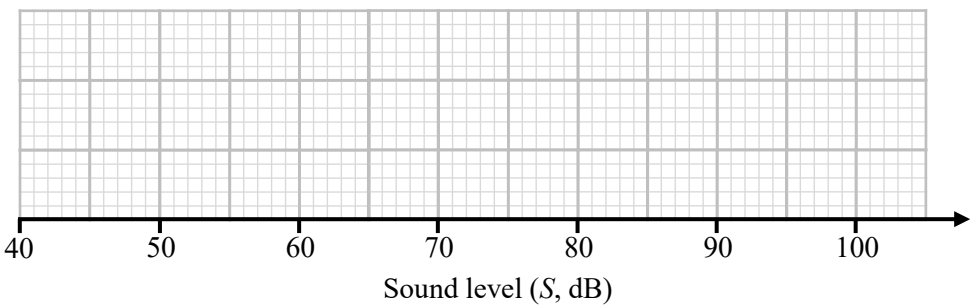
(d) Use linear interpolation to calculate estimates for the lower quartile, median and upper quartile for the sound levels. **(4)**

Another way to define an outlier is any value that falls either

more than 1.5 × (interquartile range) above the upper quartile or
more than 1.5 × (interquartile range) below the lower quartile.

(e) (i) Show that none of the sound levels were outliers using this definition. **(2)**

(ii) Draw a box plot for the sound levels. **(2)**



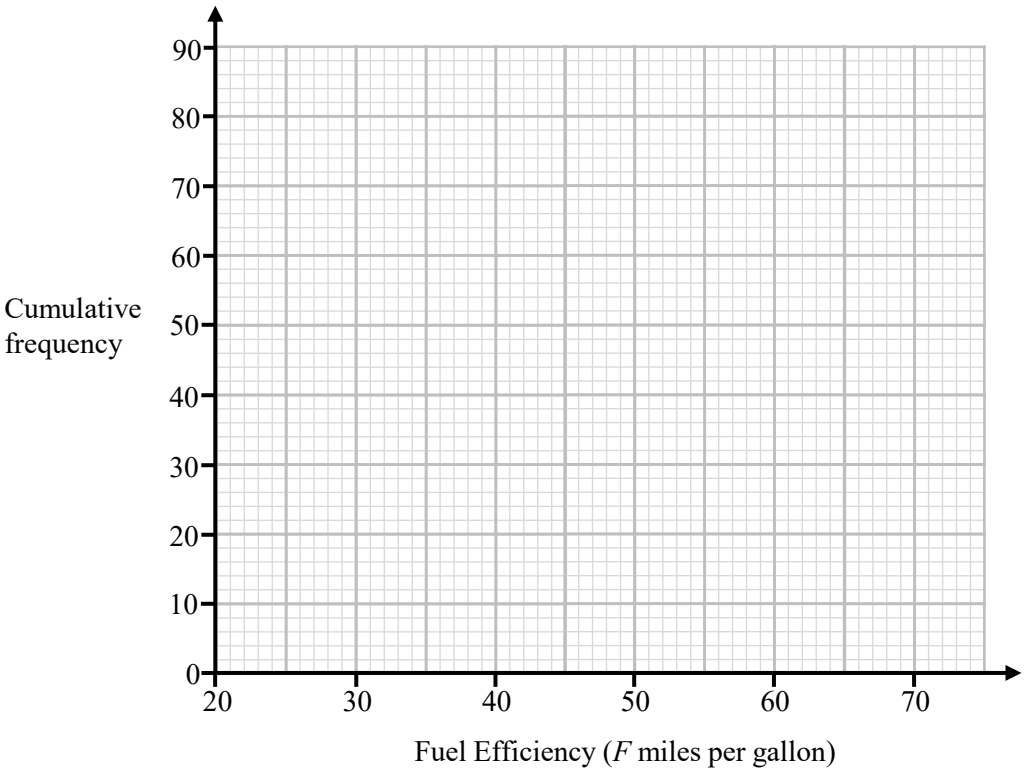
(Total for Question 7 is 13 marks)

Solutions

8 The grouped frequency table gives information about the fuel efficiency (F, miles per gallon) of 80 different vehicles.

Fuel Efficiency (F miles per gallon)	$20 \leq F < 30$	$30 \leq F < 40$	$40 \leq F < 50$	$50 \leq F < 60$	$60 \leq F < 70$
Frequency	15	21	19	17	8

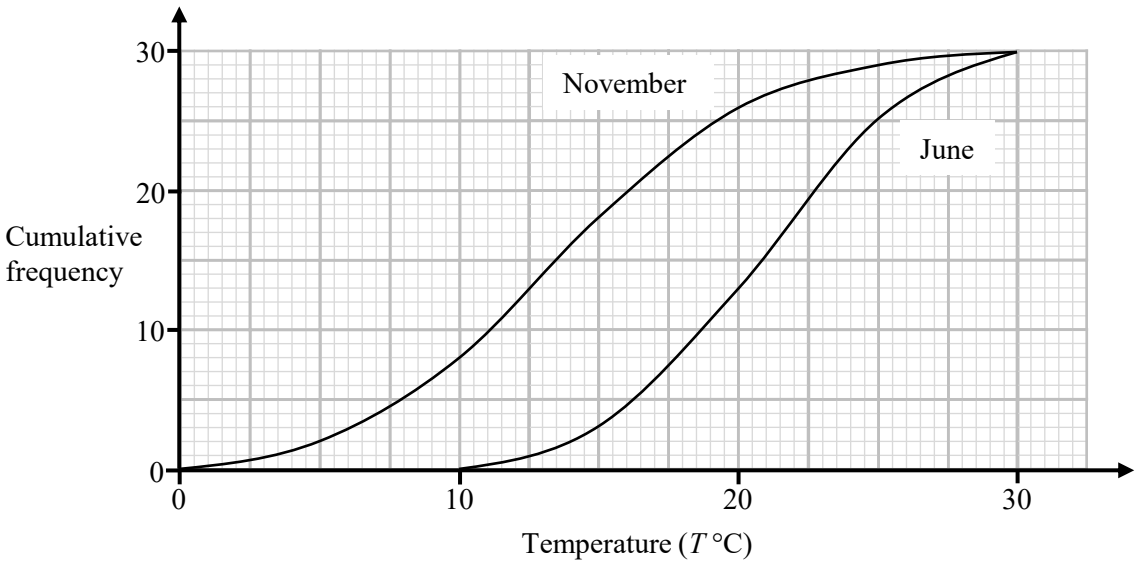
(a) On the grid, draw the cumulative frequency graph for this information. **(2)**



(b) Use your cumulative frequency diagram to estimate
(i) the median
(ii) the upper quartile
(iii) the 80th percentile
(iv) the 15th percentile **(4)**

(Total for Question 8 is 6 marks)

9 The cumulative frequency diagram below shows the maximum temperatures (T °C) for a city for each of the days during June and November.



(a) Work out estimates for the median and interquartile range of the temperatures in June. **(2)**

(b) Work out estimates for the median and interquartile range of the temperatures in November. **(2)**

(c) Compare the maximum temperatures in June and November. **(2)**

(Total for Question 9 is 6 marks)