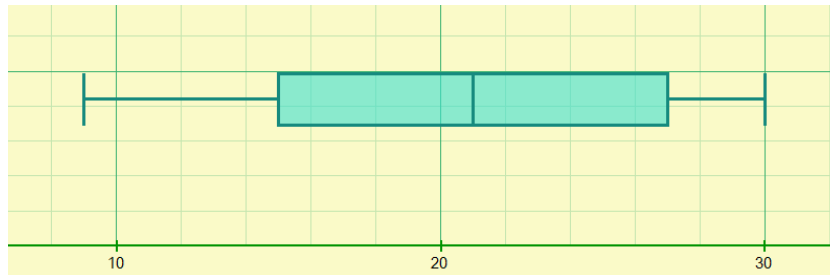## Section 2: Data presentation and interpretation

Exercise level 2 solutions

1. (i) Minimum = 9, lower quartile = 15, median = 21, upper quartile = 27, maximum = 30

   (ii) From Petr's box-and-whisker diagram, his median score is 24. This is larger than Poly's median score of 21. This supports Petr's statement.
   [You might have calculated the midrange: Poly has a midrange score of 19.5 and Petr has a midrange score of 17.5. This does not support Petr's statement.
   You cannot use the mode or mean scores for comparison as you cannot calculate these for Petr.]

   (iii) From Petr's box-and-whisker diagram, the range of his scores is 21. Poly's score also has a range of 21. This doesn't provide evidence for Petr's statement.
   From Petr's box-and-whisker diagram, the interquartile range of his scores is 5. Poly's score has an interquartile range of 12. This supports Petr's statement.

   (iv) If the maximum score in the game is 28 then it means that Poly could not have scored 30 points in any of her games. Since there is no evidence to suggest how this 30 has been recorded, it is advisable to ignore this score (rather than assume, for example, that a score of 3 has been misrecorded, or to assume that it is a maximum score of 28).
   The effect on the calculations for Poly are as follows:
   Min = 9, LQ between 14 and 15, Median = 20, UQ between 25 and 27, Max = 28
   Petr still has a higher median score and a lower midrange so the answer to (ii) is not affected.
   Petr still has a larger range and a lower IQR so the answer to (iii) is not affected.

2. (i) This is the group with the largest area: the 175 cm to 185 cm group.

   (ii) The modal group is the group with the largest frequency density: the 185 cm to 190 cm group.

(iii) We need to make an assumption about how the data is spread within each group. If we assume that the soldiers are spread equally across each group then we can use the mid-interval value as an estimate for the height of all the soldiers in that group.
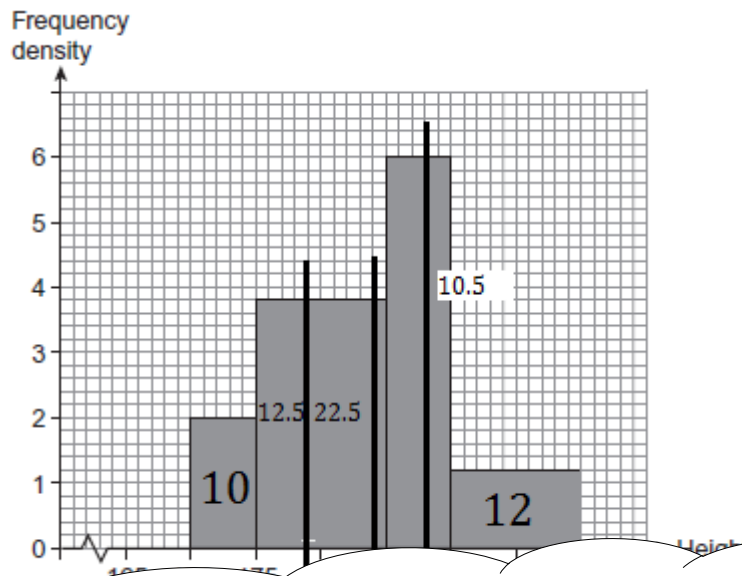
The mid-interval values are 172.5, 180, 187.5 and 195.

The frequencies are 10, 38, 30 and 12 (the areas of the bars of the histogram)

Thus the total height of all the soldiers is 16530cm, and there are 90 of them.

The mean height of a soldier is is 184cm (3sf).

(iv) Break the area of the graph into quarters. The total area is 90 so break it into sections of area 22.5 as shown:

This means that the shortest soldier has first $\frac{1}{10}$ of the area of the first bar, meaning his height, h, satisfies $170 < h < 170.5$, the next shortest satisfies $170.5 < h < 171$ and so on (each soldier shares the 5 cm width of the 170 cm to 175 cm equally between them).

Again if we assume that the soldiers heights are distributed evenly throughout each interval:

Min=170

$$LQ = 175 + \frac{12.5}{38} \times 10 = 178.3 \ (1 \text{ d.p.})$$

$$Median = 175 + \frac{35}{38} \times 10 = 184.2 \ (1 \text{ d.p.})$$

$$UQ = 190 - \frac{10.5}{30} \times 5 = 188.3$$

Max=200

IQR = 10

(v)   10 new recruits are all in the group 185 cm – 190 cm. The height of this bar will increase by 2 units and the total frequency increases to 100. The lower quartile will therefore increase because an area of 25 in front of it will move it to the right. The upper quartile will also increase (the number of people in this group has increased by a third, so the number of people above the old upper quartile will increase by a third from 10.5 to 14) by the width attributed to one person in this group, but the increase will not be as much as the increase in the lower quartile. So the interquartile range will decrease.

To check: new $LQ = 175 + \dfrac{15}{38} \times 10 = 178.9$ (1 d.p.)

new $UQ = 190 - \dfrac{13}{40} \times 5 = 188.4$ (1 d.p.)

New IQR = 9.4 (1 d.p.) so IQR has decreased.

3. (i)   Billy has calculated the midrange. The median age is the age which 400 of the people attending will be below, and 400 will be above. From the cumulative frequency diagram this is somewhere between 12 and 13 years old.

(ii)   Using the cumulative frequency diagram we can see that there were around 680 people who were below 18 (we actually find this by starting at 18 years on the x-axis as the maximum age of such a person is 18 years). So there were around 120 people paying full price. So the total revenue was £11,500.

(iii) Assume that children under 10 are dangerous to be left unaccompanied, and that parents and guardians are over 20. Then the cumulative frequency diagram suggests that there were around 260 to 280 children accompanied by at most 80 parents. This does seem quite a high ratio of children to parents, especially with around 100 children under 5 who usually require 1-to-1 attention.

(iv)   The histogram suggested the organisers would like to get more adults coming to the concert. This year only 10% (80/800) of the attendees were 20 or over, but next time they are aiming at 2/3 of the attendees being 20 or over

4. (i)   The mode is the highest bar – 4 red cars was the most common number parked on her street.
The total of the frequencies is 35 so the median is the 18th item. The total of the frequencies for 0, 1, 2 and 3 is 18, so 3 is the median.

(ii)   There are more data items larger than 3 than smaller than 3, with significantly more items one or two items above the median than one or two below (there are slightly more items three below the median than above but this will be out-weighted by the others). So I would expect the mean number of red cars to be larger than 3.

(iii) LQ is 9th data item which is 2
UQ is 27th data item which is 5
Min=0, LQ=2, Median=3, UQ=5, Max=6

(iv) The IQR is currently 3. We can see that the UQ will be very sensitive to an additional data item as it is currently right at the bottom of the next group. If the next observation is fewer than 5 red cars then exactly a quarter of the data will be 5 or 6 red cars. So the UQ will be 4.5 and the IQR will decrease since this extra data item will not change the LQ. So n = 0,1,2,3,4 are the possible values of n.