

3A Outliers

1. The blood glucose level of 30 females is recorded. The results, in mmol/litre, are shown below:

1.7	2.2	2.3	2.3	2.5	2.7
3.1	3.2	3.6	3.7	3.7	3.7
3.8	3.8	3.8	3.8	3.9	3.9
3.9	4.0	4.0	4.0	4.0	4.4
4.5	4.6	4.7	4.8	5.0	5.1

An outlier is an observation that falls either $1.5 \times IQR$ above Q_3 , or $1.5 \times IQR$ below Q_1 . Find any outliers.

2. The lengths, in cm, of 12 giant African land snails are given below:

17 18 18 19 20 20

20 20 21 23 24 32

a) Calculate the mean and standard deviation, given that $\sum x = 252$ and $\sum x^2 = 5468$.

b) An outlier is an observation which lies ± 2 standard deviations from the mean. Identify any outliers for this data.

3B Box Plots

1. The blood glucose levels of 30 males is recorded. The results, in mmol/litre, are summarized below:

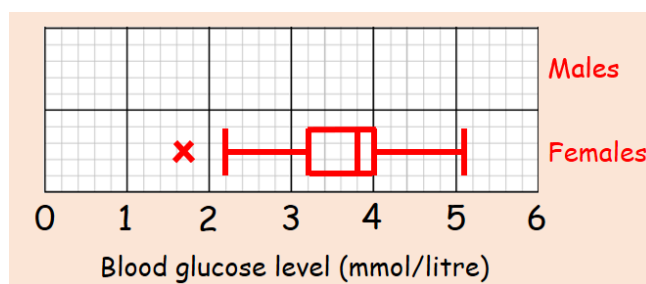
$$Q_1 = 3.6 \quad Q_2 = 4.0 \quad Q_3 = 4.7$$

Lowest value = 1.4

Highest value = 5.2

An outlier falls either $1.5 \times IQR$ above Q_3 , or $1.5 \times IQR$ below Q_1 .

- a) Given that there is only 1 outlier for males, plot this information on the same diagram as the females.



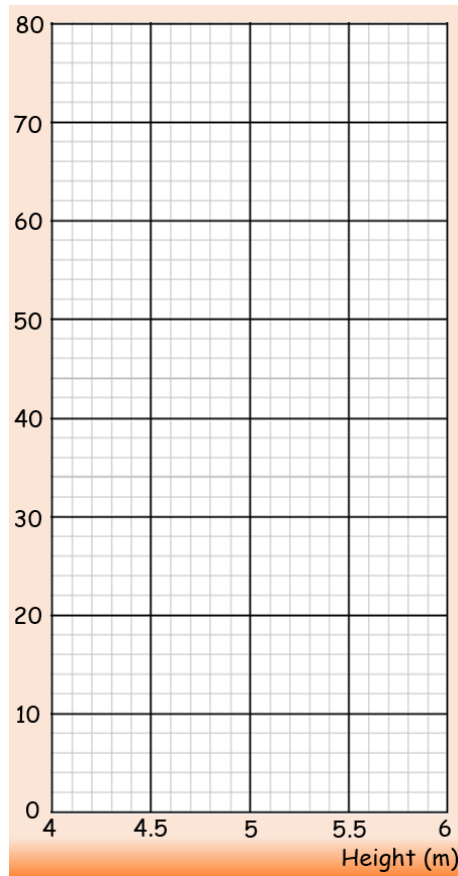
- b) Compare the blood glucose levels for males and females.

3C Cumulative Frequency

1. The data in the table shows the heights in metres, of 80 giraffes.

a) Draw a cumulative frequency diagram for the data

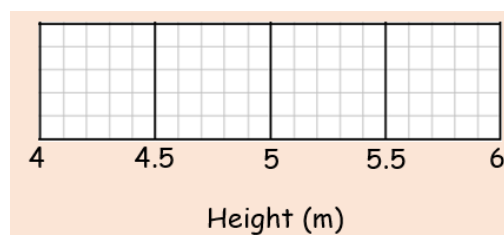
Height, h (m)	Frequency
$4.6 \leq h < 4.8$	4
$4.8 \leq h < 5.0$	7
$5.0 \leq h < 5.2$	15
$5.2 \leq h < 5.4$	33
$5.4 \leq h < 5.6$	17
$5.6 \leq h \leq 5.8$	4



b) Using the cumulative frequency diagram, estimate the median and quartiles

c) Estimate the 90th percentile

d) Draw a box plot to represent the diagram

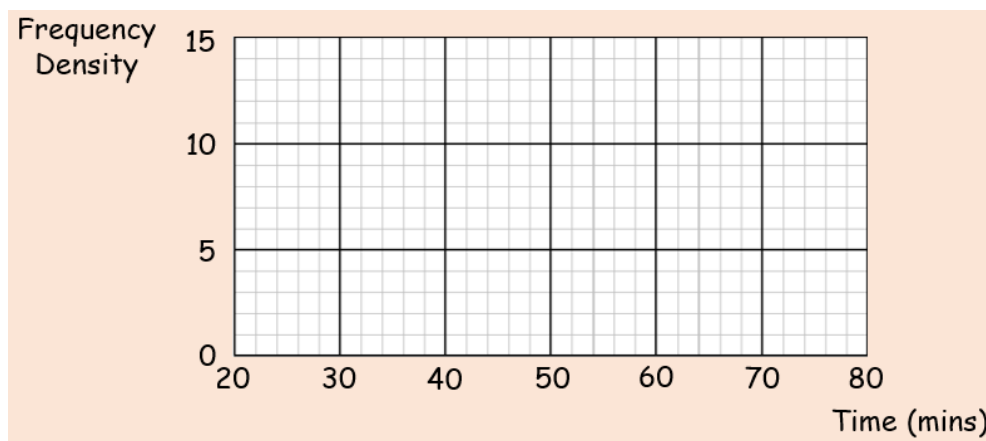


3D Histograms

1. A random sample of 200 students was asked how long it took them to complete their homework the previous night. The time was recorded and summarised in the table to the right.

- a) Draw a Histogram and frequency polygon for this data

Time, t (mins)	Frequency
$25 \leq t < 30$	55
$30 \leq t < 35$	39
$35 \leq t < 40$	68
$40 \leq t < 50$	32
$50 \leq t < 80$	6



- b) Estimate how many students took between 36 and 45 minutes to complete their homework

2. A random sample of daily mean temperatures (T , °C) was taken from the large data set for Hurn in 2015. The temperatures were summarised in a grouped frequency and represented by a Histogram.

a) Give a reason to support the use of a Histogram to represent this data

b) Write down the underlying feature associated with each of the bars in a Histogram

On the Histogram, the rectangle representing the $16 \leq T < 18$ class was 3.2cm high and 2cm wide. The frequency for this class was 8.

c) Show that each day is represented by an area of 0.8

d) Given that the total area of the Histogram was 48cm^2 , find the total number of days in the sample

3E Comparing Data

1. From the large data set, the daily mean temperature during August 2015 is recorded at Heathrow and Leeming.

For Heathrow, $\sum x = 562.0$ and $\sum x^2 = 10301.2$.

- a) Calculate the mean and standard deviation for Heathrow

For Leeming, the mean temperature was 15.6°C with a standard deviation of 2.01°C .

- b) Compare the data for the two locations using the information given