Skew

Key Vocab

Skew - A *measurement of the distortion of symmetrical distribution or asymmetry in a data set.*

Box Plot - A statistical diagram used to represent the distribution of the data using quartiles

Frequency Polygon - a statistical diagram for *continuous* data which plots the midpoint again the frequency

Histogram - a diagram consisting of rectangles whose area is proportional to the frequency of a variable and whose width is equal to the class interval.

Distribution - The way in which data is shared out between the variables

Normal Distribution - a function that represents the distribution of many random variables as a symmetrical bell-shaped graph.

Key Knowledge

Skew is a way of describing (or calculating) the distribution of a set of data

It is visible particularly in *box plots, frequency polygons, histograms* and most noticeably in the *normal distribution* (symmetrical)

Negative Skew

The median is closer to the upper quartile that the lower quartile

This means that although the range of the data is wide, the average of the data is towards the lower end of the scale

Positive Skew

The median is closer to the lower quartile that the upper quartile

This means that although the range of the data is wide, the average of the data is towards the upper end of the scale

Symmetrical Skew

Should the median be equidistant from both the lower and upper quartiles, then the skew is said to be 'symmetrical'

Skew can be identified from a statistical diagram such as a box plot but it can also be calculated with a numerical value

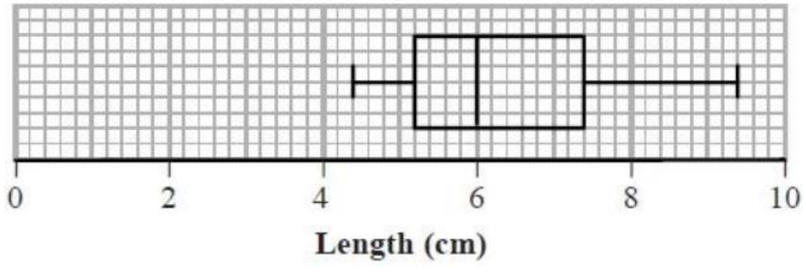
The formula for calculating skew is given as

$$\frac{3(mean - median)}{standard\ deviaiton}$$

*This formula is given on the formula page in the front of your exam paper

Example - I DO

Some students were asked to draw a 5 cm long line without using a ruler.
The actual length of each line was then measured.
The box plot shows information about the distribution of the actual lengths of the lines that were drawn by the students.



(a) Describe the skew of the distribution.

.....

(1)

(b) Describe what information is given about this distribution by its skewness.

.....

.....

(1)

Example - WE DO

The table shows information about the amount of time that each member of a group of 46 teenagers spent on social media during one day.

Time spent on social media (t minutes)	Frequency
$0 \leqslant t < 50$	1
$50 \leqslant t < 100$	4
$100 \leqslant t < 150$	8
$150 \leqslant t < 200$	17
$200 \leqslant t < 300$	16

The table below gives the mean, the standard deviation and the median for the times spent on social media during one day by a sample of sixty year olds.

Mean (minutes)	Standard deviation (minutes)	Median (minutes)
125	25	130

(a) Calculate the skew of the times spent on social media by the sample of sixty year olds.

.....

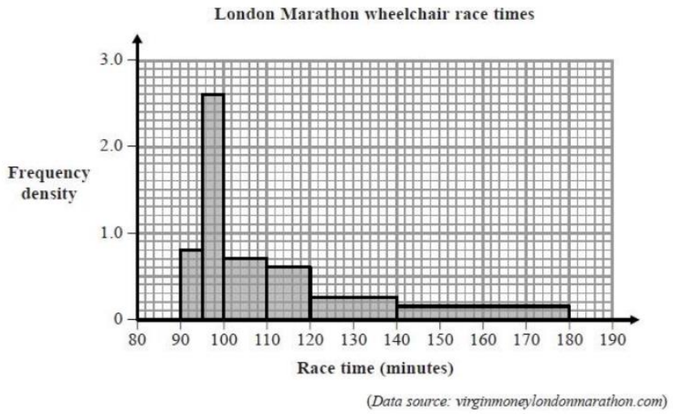
(2)

(b) Interpret your answer to part (b).

.....

YOU DO a

The histogram shows information about the race times for the 41 wheelchair competitors in the 2014 London Marathon.



(*Data source: virginmoneylondonmarathon.com*)

Describe the shape of the distribution.

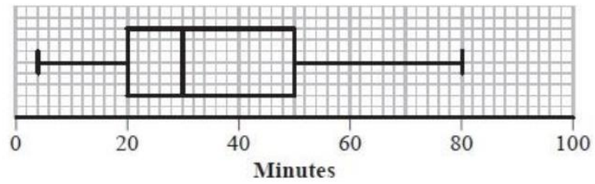
.....

(1)

YOU DO b

Malcolm asked each student in his class how many minutes of music they listen to each day.

He drew this box plot for the information.



(a) Write down the percentage of these students who listen to more than 30 minutes of music each day.

..... %

(1)

(b) Find the range.

..... minutes

(2)

(c) Describe the skew of the distribution.

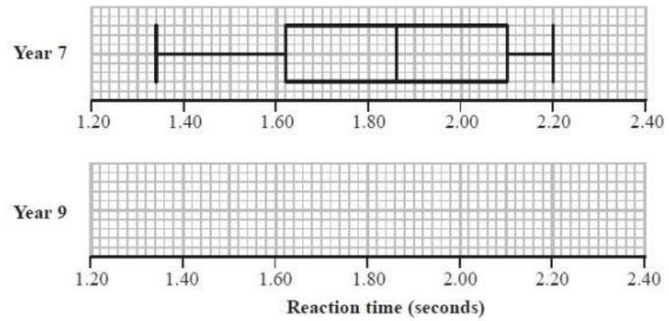
.....

(1)

YOU DO c

Alexander measured the reaction times of Year 7 and Year 9 students.

The box plot shows information about the reaction times of the **Year 7** students.



The incomplete table shows some information about the reaction times of the **Year 9** students.

Minimum	Lower quartile	Median	Upper quartile	Maximum
1.40	1.68			2.26

The Year 7 and Year 9 distributions have the same median.

The Year 7 and Year 9 distributions have the same interquartile range.

(a) On the grid above, use this information to draw the box plot for the reaction times of the Year 9 students.

(3)

(b) Describe the skew of each distribution.

.....

YOU DO d

A farmer recorded the birth weights, in kg, of a sample of 50 piglets born on his farm.

The grouped frequency table gives information about his results.

Weight (w kg)	Frequency (f)
$0.5 \leqslant w < 1.5$	4
$1.5 \leqslant w < 2.0$	12
$2.0 \leqslant w < 2.5$	17
$2.5 \leqslant w < 3.0$	13
$3.0 \leqslant w < 4.0$	3
$4.0 \leqslant w < 6.0$	1

Using the information in the table and the values of w as the class midpoints, the farmer finds that an estimate for the mean birth weight of these piglets is 2.29 kg.

He also finds that $\sum fw^2 = 286.875$

(a) Show that an estimate of the standard deviation of the birth weights of these 50 piglets is 0.7 kg, correct to 1 decimal place.

(2)

(b) Calculate an estimate for the skew of the birth weights of the piglets.

.....

(4)

(c) Interpret your answer to part (b).

.....

YOU DO e

Bella collected the finish times, x minutes, of the men's elite wheelchair race at the 2018 London Marathon.

She uses statistical software to calculate the following summary statistics.

Number of athletes = 39 Median = 97 minutes

$\sum x = 4171$ $\sum x^2 = 469\,657$

(Source: *http://results-2018.virginmoneylondonmarathon.com*)

Calculate the skew for the distribution of the finish times and interpret this value in the context of Bella's data.

YOU DO f

Sanjit writes this hypothesis.

At my school, younger boys are faster at running 400 metres than older girls.

He decides to collect some data from the students at his school to see if his hypothesis is correct.

He plans to ask each student their name, their gender, their best time to run 400 m and their age.

The table gives the mean, the standard deviation and the median of the best times taken by all the children in Sanjit's school to run 400 m.

Mean (seconds)	Standard deviation (seconds)	Median (seconds)
120	25	124

Sanjit carries out this calculation using the information in the table.

$$\frac{3 \times (120 - 124)}{25} = -0.48$$

Describe what conclusion you can make about the spread of values less than the median compared with the spread of the values greater than the median.

Give a reason for your answer.

.....

.....

.....

.....

(2)