

Chapter 3 - Representation of Data



3.1 - Outliers - Pg. 2

3.2 - Box Plots - Pg. 3 - 4

3.3 - Cumulative Frequency - Pg. 5

3.4 - Histograms - Pg. 6 - 7

3.5 - Comparing Data - Pg. 8

Personal notes:



3.1 - Outliers

Notes

- An outlier is
- Conventional statistical definition of an outlier:
- You will be told in exam which one to use, including the value of k .

Example

The weight of 11 different chipmunks are measured in kilograms:

2.2 2.5 2.7 2.7 2.8 3.0 3.1 3.1 3.2 4.0 4.7

Determine the quartiles and hence any outliers.



Example

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

- Calculate the mean and standard deviation, given that $\Sigma x = 252$ and $\Sigma x^2 = 5468$.
- An outlier is an observation which lies ± 2 standard deviations from the mean. Identify any outliers for this data.

Practice

The IQ of 15 cows are given:

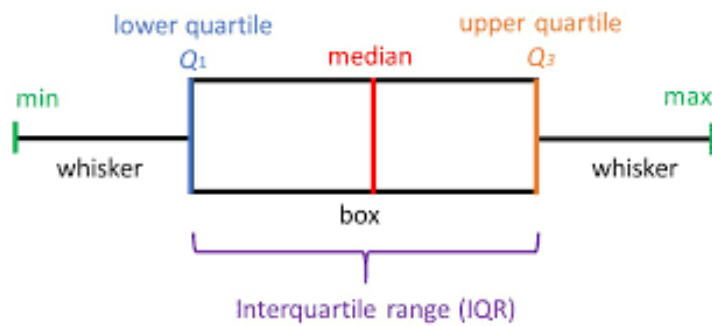
11 18 20 27 30 31 32 32 35 36 37 58 63 78 105

- If an outlier is considered to be 1.5 interquartile ranges below the lower quartile or above the upper quartile, determine any outliers.
- If instead an outlier is considered to be outside 2 standard deviations within the mean, determine any outliers. Note that $\Sigma x = 613$ and $\Sigma x^2 = 33815$



3.2 - Box Plots

Typical Box Plot (aka Box-and-whisker)



Example

Find out if there is any outlier in the data below and draw a box plot

Smallest values	Largest values	Lower Quartile	Median	Upper Quartile
0, 3	21, 27	8	10	14

Practice

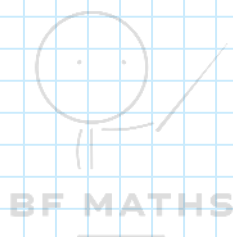
[Jan 2011 Q3] Over a long period of time a small company recorded the amount it received in sales per month. The results are summarised below.

	Amount received in sales (£1000s)
Two lowest values	3, 4
Lower quartile	7
Median	12
Upper quartile	14
Two highest values	20, 25

An outlier is an observation that falls either $1.5 \times$ interquartile range above the upper quartile or $1.5 \times$ interquartile range below the lower quartile.

(a) On the graph paper below, draw a box plot to represent these data, indicating clearly any outliers. (5)

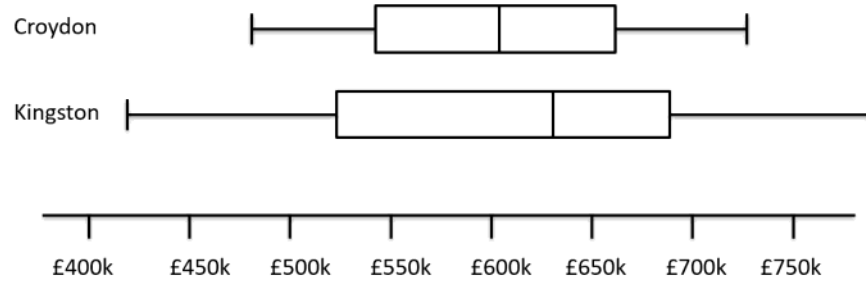
(c) The company claims that for 75% of the months, the amount received per month is greater than £10 000. Comment on this claim, giving a reason for your answer. (2)



3.2 - Box Plots

Example (Comparing Box Plot)

Box Plot comparing house prices of Croydon and Kingston-upon-Thames:



Compare the house prices in Croydon with those in Kingston. (2 marks)

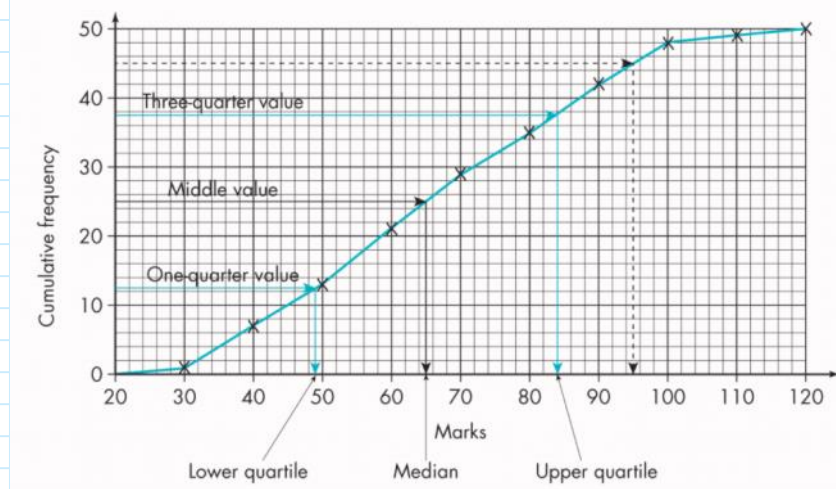


3.3 - Cumulative Frequency

Notes

- Cumulative frequency graphs are useful for estimating median and quartiles.

Example



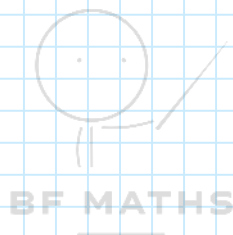
Find the median, IQR and number of students who score 80% or above.

Practice

The table shows the marks of 100 students in an English examination.

Mark	Number of students	Cumulative frequency
$21 \leq x \leq 30$	3	3
$31 \leq x \leq 40$	9	12
$41 \leq x \leq 50$	12	24
$51 \leq x \leq 60$	15	39
$61 \leq x \leq 70$	22	61
$71 \leq x \leq 80$	16	77
$81 \leq x \leq 90$	10	87
$91 \leq x \leq 100$	8	95
$101 \leq x \leq 110$	3	98
$111 \leq x \leq 120$	2	100

- Draw a cumulative frequency graph.
- Use your graph to find the median and the interquartile range.
- Students who score less than 44 have to resit the examination. Estimate the number of students who will have to resit the examination.



3.4 - Histograms

Notes

- A histogram looks similar to a bar chart, but there are 4 fundamental differences:

- $Frequency\ Density = \frac{Frequency}{Class\ width}$

Example / Practice

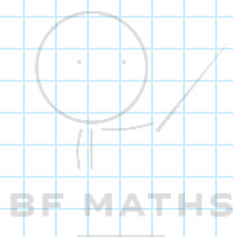
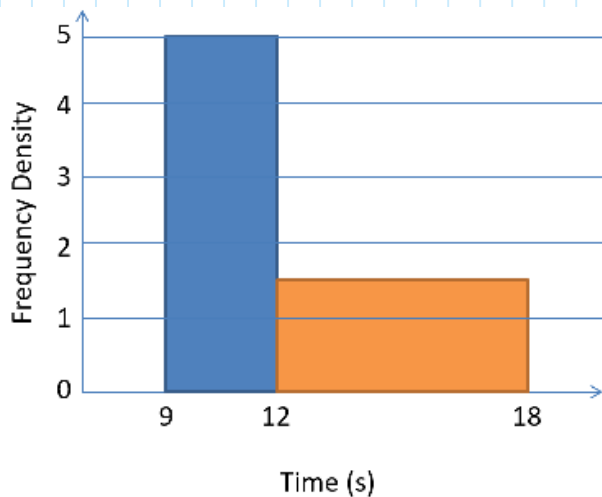
Complete the following table

Weight (w kg)	Frequency	Frequency Density
$0 < w \leq 10$	40	
$10 < w \leq 15$	6	
$15 < w \leq 35$		2.6
$35 < w \leq 45$		1

Example (Scaling Area)

There were 60 runners in a 100m race. The following histogram represents their times.

Determine the number of runners with times above 14s.



3.4 - Histograms

Example

A random sample of daily mean temperatures ($T, ^\circ C$) was taken from the large data set for Hurn in 2015. The temperatures were summarised in a grouped frequency table and represented by a histogram.

- Give a reason to support the use of a histogram to represent this data.
- 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.

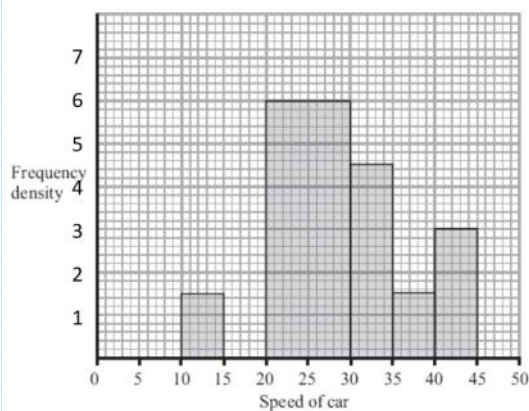
- Show that each day is represented by an area of 0.8cm^2 .
- Given that the total area under the histogram was 48cm^2 , find the total number of days in the sample.

Practice

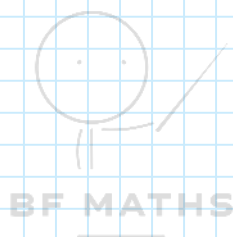
Edexcel S1 May 2012 Q5

A policeman records the speed of the traffic on a busy road with a 30 mph speed limit. He records the speeds of a sample of 450 cars. The histogram in Figure 2 represents the results.

(a) Calculate the number of cars that were exceeding the speed limit by at least 5 mph in the sample. (4 marks)



(b) Estimate the value of the mean speed of the cars in the sample. (3 marks)



3.5 - Comparing Data

- When comparing data sets, you should comment on

-

-

Example

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

For Heathrow, $\Sigma x = 562.0$ and $\Sigma x^2 = 10\,301.2$.

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.

