

UNIT 18 *Measures of Variation*

Introduction to SIM

Learning objectives

This is the last of the four related units on *data handling*. The final two units in this Strand, Units 19 and 20, cover the topic of *probability*.

After completing this unit of STUDENT INSTRUCTIONAL MATERIAL (SIM) you should be able to

- construct cumulative frequency tables and graphs (curves)
- understand how to use the cumulative frequency curve to estimate the median and upper and lower quartiles of a set of data
- construct and use box and whisker plots (not in CXC syllabus)
- calculate the standard deviation of a data set (not in CXC syllabus).

Key points and principles

- Finding the median and upper and lower quartiles is an efficient way of obtaining a measure of the variation of the data from the median.
- You can use cumulative frequency graphs (curves) to obtain estimates for the median and upper and lower quartiles.
- Box and whisker plots are a very convenient way of comparing two sets of data.
- Standard deviation is another measure of the variation in the data, this time away from the median.

Facts to remember

- The cumulative frequency curve starts at the point $(a, 0)$, where a is the start of the first data group and continues to b , where b is the largest data value (see graph under *Glossary of terms*); the cumulative frequency is given as a percentage.
- The cumulative frequency graph will be an S-shaped curve (see graph).
- The median corresponds to the 50% value on the y-axis; the upper and lower quartiles correspond to the 75% and 25% values, respectively (see graph).
- Box and whisker plots show the smallest and largest data values as well as the lower quartile, median and upper quartile (see graph).

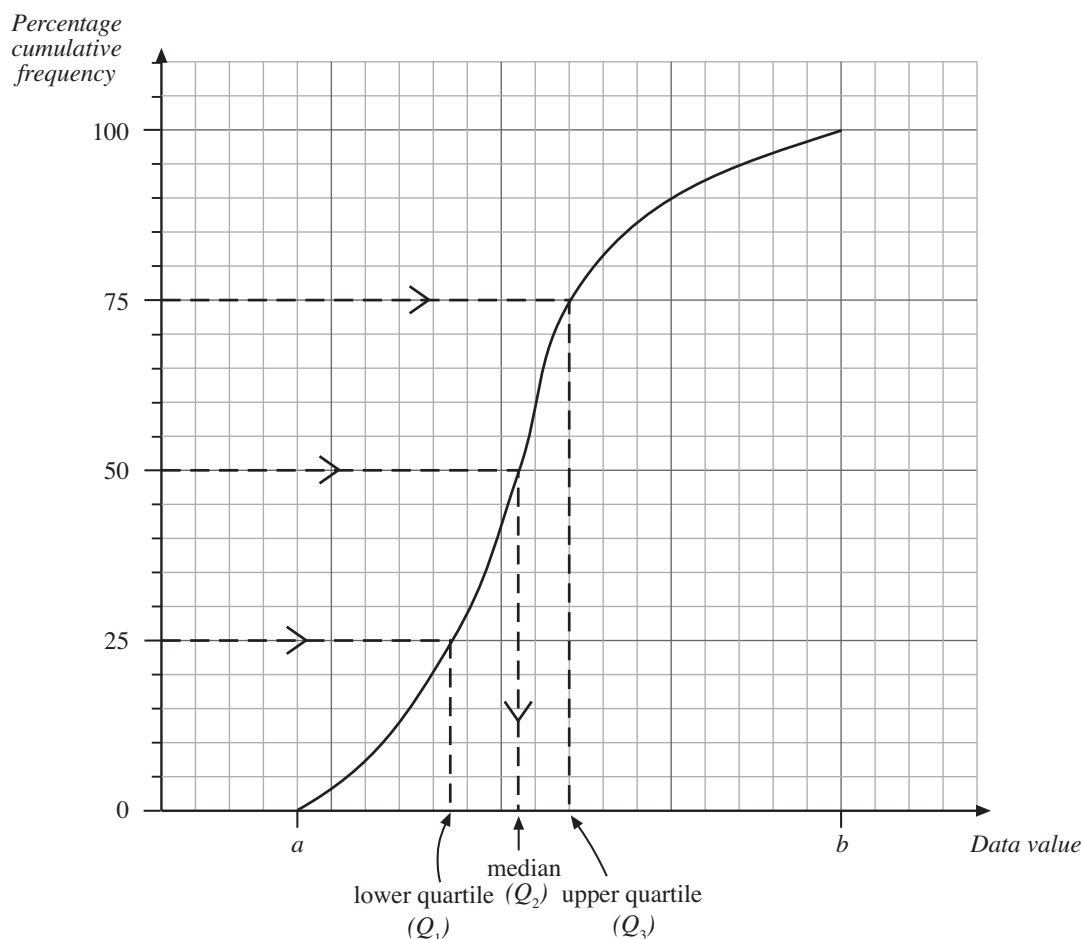
 denotes that this is not on the current CXC/CSEC Mathematics syllabus and therefore not examined, but is of relevance to the development of the topic.

UNIT 18 Measures of Variation

Introduction to SIM

Glossary of terms

- The *cumulative frequency curve (graph)* illustrates the cumulative frequency of the set of data.

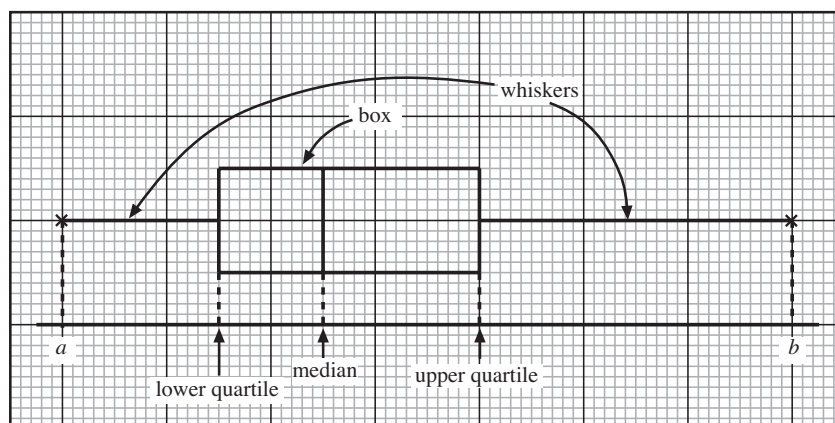


- The *lower quartile* (Q_1) corresponds to 25% on the y-axis which means that 25% of values are *below* this value.
- The *median* (Q_2) is the value corresponding to 50% on the y-axis. This means it is the *middle* value of the data.
- The *upper quartile* (Q_3) corresponds to 75% on the y-axis which means that 25% of values are *above* this value.

UNIT 18 Measures of Variation

Introduction to SIM

- A **box and whisker plot** is a way of illustrating a set of data, showing its smallest and largest values, median and upper and lower quartiles (see below).



- The **standard deviation** is a measure of the variation of the data from the mean. It is usually denoted by σ and defined by the formula

$$\sigma = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

$$\left(\text{or } \sigma = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2} \right)$$