- The **interquartile range** (IQR) is the upper quartile, Q_3, minus the lower quartile, Q_1.
- When the data is arranged in order, the **lower quartile** is the data point at the 25th percentile and the **upper quartile** is the data point at the 75th percentile.
- The mean of a set of numbers is $\bar{x}$ and the standard deviation is σ_x. If you add k to or subtract k from each of the numbers, then the mean becomes $\bar{x} \pm k$ and the standard deviation remains σ_x. If you multiply each number by k then the mean becomes $k \times \bar{x}$ and the standard deviation becomes $|k| \times \sigma_x$.
- The **population** is the whole group from which you can collect data.
- A **sample** is a small group chosen from the population.
- **Simple random sampling** is selecting a sample completely at random, for example by using a random number generator or picking numbers from a hat.
- **Systematic sampling** is, for example, taking every fifth entry starting at a random place.
- **Convenience sampling** is getting data from people who are easy to reach, for example the members of a school, club, etc. It does not select a random sample of participants and so the results could be biased.
- A **biased** sample is one that is not random, for example researching spending habits on cars and only interviewing people exiting a garage.
- **Quota sampling** is setting certain quotas for your sample, for example selecting a sample of eight boys and eight girls.
- **Stratified sampling** is selecting a sample where the numbers in certain categories are proportional to their numbers in the population. For example, if 20% of students in a school were in Grade 7, then you would choose 20% of your sample from Grade 7.
- To draw a box-and-whisker plot you need five pieces of information: the smallest value, the lower quartile (LQ), the median, the upper quartile (UQ) and the largest value.
- An outlier is a point less than the LQ $- 1.5 \times$ IQR or greater than the UQ $+ 1.5 \times$ IQR.
- Interpreting a box-and-whisker plot:
  - 25% of the values are between the smallest value and the LQ.
  - 25% of the values are between the LQ and the median.
  - 25% of the values are between the median and the UQ.
  - 25% of the values are between the UQ and the largest value.
- The **cumulative frequency** is the sum of all the frequencies up to a particular value. To draw a cumulative frequency curve, you need to construct a cumulative frequency table, with the upper boundary of each class interval in one column and the corresponding cumulative frequency in another. Then plot the upper boundary on the x-axis and the cumulative frequency on the y-axis.
- To find any **percentile**, p%, you read the value on the curve corresponding to p% of the total frequency.
- **Bivariate** data has two variables; univariate data has only one variable.
- With bivariate data you have paired data on two variables that you want to compare to see whether there is any **correlation** between the two variables.
- Correlation can **positive** or **negative**, or there may be **no correlation**, and correlation can also be described as **strong**, **moderate** or **weak**.