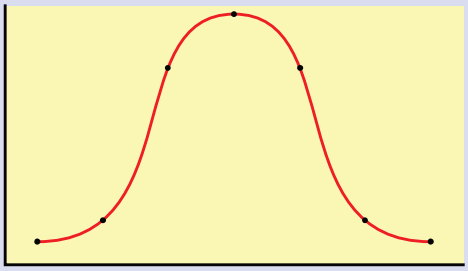
34 Statistics

A Basic statistical terms

A **normal distribution** of data means that most of the examples in a **set of data** are close to the **average**, also known as the **mean**, while relatively few examples tend to one extreme or the other. Normally distributed data shown on a chart will typically show a **bell curve**. It will often be necessary to work out the extent to which individuals **deviate**[1] **from the norm**[2] and to calculate the figure that represents **standard deviation**[3].



Six children are 7, 8, 8, 8, 11 and 12 years old. Their average or mean age is 9 years old (the **sum** of their ages divided by six). The **mode** (the most frequent value) is 8. The **median** is 9.5 (the **halfway point** between the two **extremes** of the **range**).

Statisticians are often concerned with working out **correlations**[4] – the extent to which, say, left-handedness **correlates with** intelligence. They must ensure that any data they collect is **valid**, i.e. that it is measuring what it claims to measure – all the subjects in the **sample**[5] must be appropriately and accurately assessed as left or right-handed, for example. The figures must also be **reliable**, i.e. they would be **consistent**[6] if the measurements were repeated. Usually, statisticians hope that their calculations will **show/indicate** a **tendency**, e.g. that left-handed people will be shown to be **significantly**[7] more intelligent than right-handed people.

[1] differ [2] the average [3] average difference from the norm [4] connections, often as cause and effect [5] the subjects of the experiment or group representing the total population measured [6] the same [7] noticeably

B A probability[1] problem

Sue picks a card **at random**[2] from an ordinary pack of 52 cards. If the card is a king, she stops. If not, she continues to pick cards at random, without replacing them, until either a king is picked or six cards have been picked. The **random variable**[3], C, is the total number of cards picked. Construct a diagram to illustrate the possible **outcomes**[4] of the experiment, and use it to calculate the **probability distribution**[5] of C.

[1] likelihood of something happening [2] by chance [3] number or element of a situation that can change [4] results [5] assessment of probabilities for each possible value of C

C Other useful nouns for talking about statistics

In a class of 8 women and 4 men, what **proportion**[1] are male? Answer: one third

In the same class what is the female to male **ratio**[2]? Answer: 2:1 (two to one)

The figures show a **trend**[3] towards healthier eating habits.

The study investigates the increase in the **volume**[4] of traffic on the roads.

Most of the students achieved marks between 45% and 65% but there were a couple of **outliers**[5] who got 32% and 84%

[1] number compared with another number [2] relationship between two numbers showing how much bigger one is [3] change in a particular direction [4] amount, quantity [5] figures very different from others in the set

Common Mistake

We say *10* ***per cent*** (NOT ~~the 10 per cent~~ or ~~10 percentage~~) of students got an A for the exam but the ***percentage*** of students achieving an A has increased.

Exercises

34.1 Complete the sentences with words from A opposite.

1 The six subjects who took the test scored 24, 22, 16, 16, 16 and 14 points out of 30. The ..... was 16. The ..... score was 19 and the ..... or ..... score was 18.
2 The ..... of all donations to the charity in 2003 was $3,938. The smallest donation was $10 and the largest was $130. Most were around the ..... point of $60.
3 The centre has recorded a wide ..... of temperatures, with the two ..... being 35 in the summer and –6 in the winter.

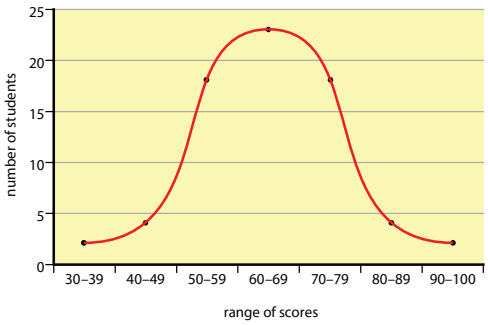
34.2 Complete the text with words from the box. There are three words you don't need.

distribution trends varieties significantly probability sample random correlate outcomes variables insignificantly

Life insurance companies base their calculations on the laws of [1] ....., that is they assess the likely [2] ....., given the different [3] ..... such as age, sex, lifestyle and medical history of their clients. The premiums are therefore not chosen at [4] ..... but are carefully calculated. The [5] ..... of ages at which death occurs and causes of death are studied to see if they [6] ..... with other factors to be taken into account in setting the premiums. Naturally, the companies also monitor social [7] ..... and react to any changes which might [8] ..... affect mortality rates.

34.3 Answer the questions.

1 There are 12 male students and 6 female students in the class. What is the ratio of males to females? And what proportion of the class is male?
2 If in a sample of 100 students, 98 evaluate a module as 5 or 6 out of 10 but 1 gives it 1 and 1 gives it 10, what are the scores of 1 and 10 called?
3 If my data show a tendency for students to choose the type of clothing their friends choose, does it mean that they always, often or rarely choose similar clothes?
4 If I repeat the same experiment three times and the results are not consistent, is my method reliable?
5 If 20 out of 200 students fail an exam, what proportion, in percentage terms, failed?
6 If the average score in a test is 56, and Barbara scores 38, by how many points has she deviated from the norm?
7 If the volume of court cases increases, what changes: the type of case, the size of each case or the total number of cases?
8 What does standard deviation tell us: (a) What the standard of something is, (b) what the norm is, or (c) what the average difference from the norm is?
9 If a general survey of teenage eating habits asks questions about what teenagers eat for breakfast and lunch, is the survey likely to be valid?
10 Here is a graph showing how many students got scores within each 10-mark band in a biology test. Do the scores show a normal distribution? What is the shape of the graph called?



Over to you

What kinds of statistical data are likely to be discussed in your discipline? Find a relevant chart, graph or table and write about it using some terms from this unit.