

D1 DATA HANDLING 1

Statistics is about handling information, or **data**.
You use statistics to collect facts and opinions.



This unit will show you how to:

- Use statistical terms
- Devise a hypothesis
- Conduct a survey
- Use a database

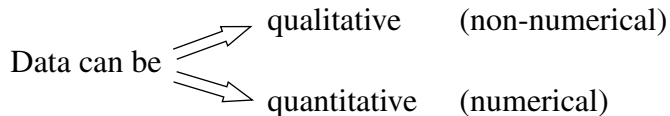
1.1 Statistical terms

First you need to know some of the words most commonly used in statistics.

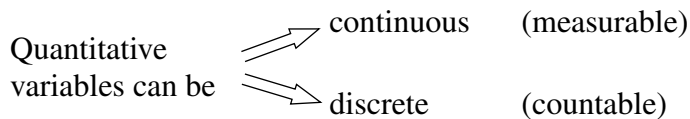
Data is information that you collect.

You collect data on a particular **variable**, for example height in cm.

Variables can change from one value to another.



- **Qualitative** data is information that can be described in words.
Examples are hair colour or pupils' names.
- **Quantitative** data is information that can be measured or counted.
It is presented as a number, for example weights of babies or the number of people attending a hockey match.



- **Continuous** variables can take any value within a given range.
Examples are height, area or speed.
- **Discrete** variables can only take certain values.
Examples are the number of pupils in a class, or clothes sizes.

Exercise 1A

In each question below, state whether the question will give qualitative or quantitative information. If the information is quantitative, state whether it will be discrete or continuous. You do not need to answer the questions.

1. What is your favourite song?
2. How many branches are on the tree?
3. What word would you use to best describe the last film you saw?

You will learn more about variables in Unit 11 when you study algebra.

Qualitative comes from the word quality; quantitative comes from the word quantity.

The height of a person is continuous, but the height of a person measured to the nearest centimetre is discrete.

4. What is the colour of your front door?
5. In which month were you born?
6. How many days are there in this month?
7. Which party will you vote for in the next election?
8. What is your trouser size?
9. How much money have you in your pocket at this moment?
10. What is the area of the carpet?
11. What is your date of birth (in numbers)?
12. What is the height of the sunflower?

1.2 Hypotheses and types of data

Before you start collecting data you should have a clear idea what you are going to investigate. It often helps if you devise a **hypothesis**.

- A hypothesis is a general statement about an area of inquiry.

Examples of hypotheses are ‘Smoking is linked to lung cancer’ and ‘Women live longer than men’.

Once the hypothesis has been made, data will need to be collected and analysed so that you can test whether the hypothesis is likely to be true or not.

Collecting data will need careful planning.

- You could gather evidence at first hand by conducting an experiment or survey, perhaps involving a questionnaire. These methods will give you **primary data**.
- If your hypothesis requires you to use data that has already been collected from a previous investigation or survey, then you will be obtaining **secondary data**.

Example

Suggest an appropriate method which could be used to test the following hypotheses.

- (a) Women drivers have fewer accidents than men drivers.
- (b) The most popular soap on terrestrial television is EastEnders.

- (a) For ‘the women drivers and fewer accidents’ hypothesis you would have to use secondary data from insurance companies.
- (b) A method for investigating the most popular soap would be to devise a suitable questionnaire and survey your local village or town.

When you have made your hypothesis and decided on the type of data you will collect, the following questions may help you to decide if your hypothesis is acceptable:

- (i) Will I be able to test my hypothesis?
- (ii) Will I be able to collect enough data?
- (iii) What techniques will I use for displaying and analysing my data?
- (iv) How will I know if I have proved or disproved my hypothesis?

If these questions raise problems, you may need to revise your hypothesis.

Exercise 1B

1. Suggest an appropriate method which could be used to test the following hypotheses.
 - (a) Boys are better than girls at estimating distances.
 - (b) Bournemouth has its least rainfall in the month of June.
 - (c) Our town needs a hypermarket to offer shoppers a greater variety of choice.
 - (d) The general public under-uses buses as a means of transport.
2. Devise a method to test the hypothesis:
‘There is more sport shown on BBC1 than on any other terrestrial TV channel.’
3. Devise a questionnaire to test the hypothesis:
‘School uniform should be abolished.’

You will learn more about questionnaires on page 000.

1.3 Surveys and questionnaires

Surveys

Surveys are used for obtaining and recording information about any chosen topic. A survey may be carried out either by **census** or by taking a **sample** from the population.

- In a census the entire population is used for the survey.
- In a sample survey only a proportion of the population is used.

The data that you collect in your survey can be either primary or secondary. The three main survey methods that you use to obtain primary data are:

- Asking questions face to face, for example giving a questionnaire to people in the street, consulting fellow pupils at school or members of your family.
- Observing situations and recording results, for example the colour of cars passing a particular road junction at a given time.
- Postal or telephone survey, where the responses to a questionnaire are recorded on a form.

These guidelines should be followed when conducting a survey:

- The survey should yield sufficient data. Usually between 20 and 50 responses would be the minimum acceptable for a conclusion to be made.
- There should be sufficient variety in the respondents to ensure a true picture of the population.
- Care should be taken to avoid **bias** when conducting a survey. Bias could occur, for example, if there is any form of deliberate selection of the people chosen to respond. Bias can distort results, and can make any conclusions drawn from the survey have little value.

Questionnaires

Questionnaires are commonly used to collect data, particularly on opinions.

When writing questions for a questionnaire, the questions must be carefully constructed and the following points must be remembered.

- The number of questions asked should be sufficient to gather all the information needed. However, the questionnaire should not be too lengthy.
- Make the questions easy to understand by using simple language.

In statistics, the word 'population' is used to mean all the items in a particular set. Examples of populations include 'the entire set of people living in Birmingham' or 'the entire set of books in Tooting library'.

Frequently businesses need to conduct **market research** to decide how to improve their products and services.

- (c) The type of answer required should be clear. For example, the respondent may need to know whether the response is a 'yes/no' type, or a selection from a choice of possible answers.
- (d) Avoid open-ended questions where possible. It is usually best to provide a choice of tick-box answers for people to select from.
- (e) Do not ask leading questions, as the results from these responses will be invalid. An example of a leading question is 'Do you agree with the fact that spinach tastes nicer than broccoli?'
- (f) Ask questions in a logical order.
- (g) Avoid personal questions.

It is sensible to try a questionnaire out on a few people at first, in a **pilot survey**. This should enable you to find out if any of the questions need alteration.

Example

The following questions (a) and (b) were suggested as alternatives for a questionnaire. Criticise them, saying which you think is best, and suggest improvements.

- (a) What is your age?
- (b) Which is your age group? (Please tick the box)

Under 16	<input type="checkbox"/>
16–20	<input type="checkbox"/>
20–35	<input type="checkbox"/>
36–50	<input type="checkbox"/>
50–60	<input type="checkbox"/>

- (a) This is a personal question, and older people may be reluctant to answer it truthfully.
- (b) This is better than (a) as it is much less personal. However, a 20-year-old will have two possible responses, and anyone aged over 60 will have no responses.

An improvement would be to have a question like (b), but alter the categories to read:

Which is your age group? (Please tick the box)

Under 16	<input type="checkbox"/>
16–20	<input type="checkbox"/>
21–35	<input type="checkbox"/>
36–50	<input type="checkbox"/>
Over 50	<input type="checkbox"/>

Exercise 1C

1. The following questions were suggested for a questionnaire. Consider each question and give a reason for it being either a suitable or poor form of question to be asked.
 - (a) What do you think of the improved library facilities?
 - (b) Do you like chocolate?
 - (c) How much money do you earn?
 - (d) Meat is obtained through killing animals. Do you eat meat?
 - (e) What is your date of birth?
 - (f) Don't you agree that bus lanes are good for public transport?
 - (g) Don't you think it's ridiculous to spend money on a project like the Millennium Dome or the new Wembley Stadium when people in Africa are starving?
2. Your school is planning to upgrade the library and learning resource centre. The following questions were put forward to be asked in a survey. Explain why each question is unsuitable in its present form and rewrite the question.
 - (a) What is your age?
 - (b) How often have you been to the library?
 - (c) What did you read in the library?
 - (d) How would you improve the library?
3. A survey of the school tuck shop is to be conducted. Suggest five questions which could be included with the choice of responses for each question.
4. The school council wanted to investigate the attitude of the pupils to the current school uniform. Suggest three questions which could be asked, with each question having up to five responses.
5. Ken and Jane are thinking of opening a newsagents shop in a shopping parade. Suggest four questions which could be asked to people living in the area which could enable Ken and Jane to decide whether the newsagent is likely to be successful.

1.4 Databases

Government departments routinely collect data on population, housing, transport and health. This data provides information for local councils to plan for people's needs. The vast amounts of data collected need to be entered into a computer database to enable analysis to be made quickly and efficiently.

Many different databases are used on a daily basis to check population data:

- A telephone directory is an example of a database that can be searched manually for information.
- Police have access to the national computer population database where names, addresses, dates of birth and other personal records for all UK citizens can be obtained.
- Many organisations, for example banks, gas, electricity and credit card companies, have access to a national population database which will give the names of all UK citizens who are registered as resident at a particular address.

Exercise 1D

1. Here is a database for a small second-hand car sales company.

Year	Registration letter	Make	Colour	Model	Mileage	Cost (£)
97	R	Ford	Red	Fiesta 1.25 LX	33 500	4695
97	P	Ford	Silver	Fiesta 1.25 Fusion	29 500	4695
97	P	Ford	Red	Fiesta 1.25 LX Auto	10 000	4995
97	R	Ford	Black	Fiesta 1.25 Flight	23 500	5896
97	R	Ford	Blue	Fiesta 1.8 Diesel	29 500	4995
97	P	Ford	White	Escort 1.4 Encore	30 500	3995
97	P	Ford	Red	Escort 1.8 Encore	28 500	4995
95	N	Ford	White	Escort 1.4 Encore	24 500	3485
97	R	Ford	Green	Escort 1.4 Encore	30 150	4995
96	N	Ford	Blue	Mondeo 2.0 LX	46 716	4495
96	N	Ford	Blue	Mondeo 2.5	67 129	4495
98	S	Ford	Blue	Mondeo 1.8 LX	39 500	6995
97	R	Ford	Green	Mondeo 1.8 Verona	29 500	7495
97	R	Ford	Blue	Mondeo 1.8 Verona	14 500	7495
00	W	Ford	Thistle	Fiesta 1.25 Ghia	2000	7995
00	X	Ford	Silver	Fiesta 1.25 Ghia	Delivery mileage	8495
00	W	Ford	Pepper red	Focus Zetec	2000	9999
00	X	Ford	Pacific blue	Puma 1.7 Coupe	Delivery mileage	12 495

- (a) Jean likes white cars.
Which cars are available and how much do they cost?
- (b) Bill can afford £5000 and he wants to buy a car which has recorded less than 25 000 miles.
From which cars can he choose?
- (c) Brian can afford to spend up to £10 000 on a car but does not want to buy a car manufactured before 1999.
Which cars can he choose from?
- (d) Judith likes only Mondeo cars.
From what colours of car can she choose?
- (e) How many cars are available costing between £4500 and £5000?
- (f) Tony wants to buy a green car with the lowest mileage possible! Which car would he choose and what is the recorded mileage and price of this car?

2. Here is part of a database about students.

Student	Gender	Month of birth	Numeric day in the month	Day of birth
Aaron	M	February	17	Monday
Brenda	F	December	14	Wednesday
Colin	M	September	22	Friday
Debbie	F	June	1	Tuesday
Eric	M	March	3	Tuesday
Fiona	F	August	31	Friday
Gita	F	February	14	Wednesday
Helen	F	June	12	Friday
Ian	M	August	7	Monday
Jeremy	M	June	5	Wednesday
Kevin	M	December	27	Thursday

- (a) How many students were born later than May in the year?
- (b) How many male students were born in December?
- (c) How many female students were born before the 16th of their month of birth?

3. The following database gives information about houses available for sale from an estate agent.

Town	Type of property	Number of bedrooms	Central heating	Cost
Dudley	Terrace	3	No	£34 950
Gornal	Flat	1	No	£15 000
Dudley	Detached	4	Yes	£107 950
Tipton	Terrace	3	Yes	£38 500
Dudley	Semi-detached	3	Yes	£48 000
Tipton	Detached	4	Yes	£76 500
Gornal	Semi-detached	3	No	£41 950
Dudley	Flat	2	Yes	£33 650
Gornal	Detached	5	Yes	£95 000
Dudley	Semi-detached	3	No	£45 000
Tipton	Detached	2	Yes	£32 000

- (a) How many properties have 3 bedrooms?
 (b) How many semi-detached houses cost more than £40 000?
 (c) How many Dudley properties have 3 or more bedrooms, central heating and cost less than £50 000?

4. The following database gives information about 'Last Minute' holidays for 7 nights available in June 2001.

Country	Resort	Hotel rating (*'s)	Departure airport	Cost per adult (£)
Cyprus	Paphos	4	Birmingham	440
Rhodes	Rhodes Town	5	Luton	385
France	Cannes	4	Manchester	495
Crete	Rethymnon	3	Gatwick	219
Spain	Benidorm	3	Heathrow	199
Cyprus	Limassol	4	Manchester	289
Malta	Valetta	3	Manchester	430
Malta	Sliema	3	Gatwick	275
Spain	Nerja	4	Luton	195
Crete	Aghios Nikolaos	5	Luton	395
France	Nice	5	Luton	550
Crete	Rethymnon	2	Birmingham	175

- (a) How many holidays at 4* hotels cost less than £350 per adult?
- (b) I want to travel to France or Spain on holiday. What is the lowest cost of holiday available and at which resort?
- (c) How many holidays costing between £300 and £400 are available at 5* hotels?

5. The database below contains information about a group of Year 8 pupils.

Pupil	Gender	Height (cm)	Hand span (cm)	Shoe size
Anne-Marie	F	165	18.5	38
Julian	M	183	29.5	48
Alex	M	147	16.8	36
Judith	F	158	17.5	38
Clive	M	182	21.0	48
George	M	177	20.5	48
Brian	M	195	24.8	52
Peter	M	163	18.0	44
Ann	F	131	18.3	40
Robin	M	175	20.0	46
Inder	F	168	19.5	42

- (a) Which gender of student has the larger hand span?
- (b) Which students have the same shoe size?
- (c) How many students are shorter in height than Robin?
- (d) What is the difference in height between the tallest and shortest pupils?

Summary

1. You know whether data is quantitative or qualitative.
2. You know the difference between discrete and continuous variables.
3. You know the meaning of a database.
4. You know the difference between a census and a sample.
5. You know how to carry out a survey.
6. You know how to construct a questionnaire.
7. You know what is meant by pilot survey, bias and hypothesis.

Check out D1

1. State whether the following are quantitative or qualitative data:
 - (a) the number of cars in a traffic queue
 - (b) the colour of the first car in a traffic queue
 - (c) the number of books in a pupil's school bag
 - (d) the weight of books in a school bag.
2. For those in question 1 which are quantitative data, is the variable continuous or discrete?
3. How would a bank check that you live at your address?
4. Why does the BBC not use a census to see how many people watched 'Neighbours' last week?
5. What is the minimum number of people needed in a survey to make an appropriate conclusion?
6. (a) State three rules that should be used when making a questionnaire.
 (b) You wish to investigate whether a person's income is related to the number of cigarettes they smoke. Write a short questionnaire (five or six questions) on this topic.
7. You wish to test the hypothesis that the time of day at which people visit a supermarket relates to their age.
 - (a) Give three questions which you could ask to test this hypothesis.
 - (b) How would you use a pilot survey to help you in your questionnaire?
 - (c) How would you attempt to eliminate bias in the responses?

Revision exercise D1

1. A teacher asks all his class
‘How many children are there in your family?’
Here are their replies.

Number of children in the family	Number of replies
1	17
2	12
3	5
4	2
5	0

- (a) How many children are in the class?
- (b) What is the most common number of children in the family for this class?
- (c) Calculate the mean number of children per family in this class.
Give your answer to 1 decimal place.
2. This statement is made on a television programme about health: ‘Three in every eight pupils do not take any exercise outside school.’
- (a) A school has 584 pupils.
According to the television programme, how many of these pupils do not take any exercise outside school?
- (b) Clare says, ‘I go to the gym twice a week after school.’ She decides to do a survey to investigate what exercise other pupils do outside school.
Write down two questions that she could ask.
- (c) Matthew decides to do a survey in his school about the benefits of exercise.
He decides to ask the girls’ netball team for their opinion.
Give two reasons why this is not a suitable sample to take.

(d) This is part of Matthew's questionnaire.

Question	<i>Don't you agree that adults who were sportsmen when they were younger suffer more from injuries as they get older?</i>		
Response	<i>Tick one box</i>		
	<input type="checkbox"/> <i>Yes</i>	<input type="checkbox"/> <i>Usually</i>	<input type="checkbox"/> <i>Sometimes</i> <input type="checkbox"/> <i>Occasionally</i>

- (i) Write down one criticism of Matthew's question.
- (ii) Write down one criticism of Matthew's response section.

3. Jane does a survey about vehicles passing her school. She wants to know about the types of vehicles and their colours.

Design a suitable observation sheet to record this information.

Fill in your observation sheet as if you had carried out this survey.

You should invent suitable data for 25 vehicles.

4. Winston has designed a data collection sheet to record the number of bottles that each person puts into a bottle bank.

Number of bottles	Tally	Frequency
0 to 2		
3 to 6		
6 to 8		

(a) Give three criticisms of the class intervals that Winston has chosen.

Anna and Patrick watch people using the bottle bank.

Anna watches 60 people and calculates the mean to be 8.5 bottles per person.

Patrick watches 15 people and calculates the mean to be 9.2 bottles per person.

- (b) Which of the two means would you expect to give the more reliable estimate of the mean number of bottles per person?

Give a reason for your answer.

5. A survey on clothes shopping included the following questions.

1 What is your total annual income?				
2 How much do you spend per month on clothes?				
£0–£20	£20–£40	£40–£60	£60–£80	£100 or more
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 What are your hobbies?				

- (a) Why should question 1 not be asked?
 (b) Write down one criticism for each of question 2 and question 3.

6. The table shows information about some cars.

Make	Colour	Mileage
Vauxhall	blue	8 606
Ford	white	12 214
Vauxhall	white	5 567
Rover	red	11 984
Rover	blue	9 085
Vauxhall	red	6 984
Ford	blue	8 763
Vauxhall	white	14 675

- (a) What is the range in the mileage of these cars?
 (b) Calculate the mean of the mileages.
 (c) What is the modal make of car?

- (d) A car is chosen from the list at random.
- (i) What is the probability that it has a mileage of more than 10 000?
 - (ii) What is the probability that it is a white Ford?
- (e) A Vauxhall car is chosen at random.
The probability that it is blue is 0.25.
What is the probability that it is not blue?