



Section 2.2 Collecting data



53

Data is collected for a variety of reasons and from a variety of sources.

Companies do market research to find out what customers like or dislike about their products and to see whether or not they would like new products. The government carries out a **census** of every person in the country every five years. Local government, education authorities and other organisations use the information obtained for further planning.

Data can be collected through direct observation such as a naturalist observing animal behaviour. In an observational study, the observer wishes to record data without interfering with the process being observed.

Apart from observational studies, data may also be collected by

- › carrying out a survey
- › doing an experiment
- › conducting interviews or completing questionnaires
- › using a data logger which records data or readings over a period of time, using a sensor.

1. Surveys

Surveys are particularly useful for collecting data that is likely to be personal.

The main survey methods are:

- › postal surveys in which people are asked questions
- › personal interviews in which people are asked questions; this type of survey is very widely-used in market research
- › telephone surveys; here the interview is conducted by phone
- › **observation**, which involves monitoring behaviour or information.

Survey method	Advantages	Disadvantages
Observation	› Systematic and mechanical	› Results are prone to chance
Personal interview and telephone survey	› Many questions can be asked › High response rate	› Expensive › Interviewer may influence responses
Postal survey	› Relatively cheap › Large amounts of data can be collected	› Limited in the type of data that can be collected › Poor response rate

2. Questionnaires

One of the most commonly-used methods of conducting a survey is by means of a questionnaire.

A **questionnaire** is a set of questions designed to obtain data from individuals.

People who answer questionnaires are called **respondents**.

There are two ways in which the questions can be asked.

- › An interviewer asks the questions and fills in the questionnaire.
- › People are given a questionnaire and asked to fill in the responses themselves.

When you are writing questions for a questionnaire,

- › be clear on what you want to find out and what data you need
- › ask short, concise questions
- › start with simple questions to encourage the person who is giving the responses
- › provide response boxes where possible: Yes No
- › avoid leading questions such as
 - ‘Don’t you agree that there is too much sport on television?’
 - or ‘Do you think that professional footballers are overpaid?’
- › avoid personal questions such as,
 - ‘Do you live in an affluent area?’
 - or ‘Are you well educated?’
 - or ‘Are you overweight?’

A choice of responses can be very useful in replying to the question, 'What age are you?'

Here is an example:

Tick your age in one of the boxes below:

 Under 18 years 18–30 31–50 Over 50

Notice that there are no gaps in the ages and that only one response applies to each person.

When you are collecting data, you need to make sure that your survey or experiment is **fair** and avoids **bias**. If bias exists, the data collected might be unrepresentative.

The boxes given below contain questions that should be avoided because they either are too **vague**, too **personal**, or may **influence** the answer.

How often do you play tennis?

Sometimes Occasionally Often

The three words *sometimes*, *occasionally* and *often* mean different things to different people.

Normal people enjoy swimming.

Do you enjoy swimming?

Yes No

This is a leading question and may cause the result to be biased.

The first sentence should not be there.

Have you ever stolen goods from a supermarket?

Yes No

Few people are likely to answer this question honestly if they have already stolen.

Whenever you undertake a survey or experiment, it is advisable to do a pilot survey. A pilot survey is one that is carried out on a very small scale to make sure the design and methods of the survey are likely to produce the information required. It should identify any problems with the wording of the questions and likely responses.

3. Control group

If we wish to investigate whether a new drug has any effect on those who take it, we select a group of patients, chosen at random, to form a sample. The sample is then divided randomly into two groups. Both groups think that they are taking the new drug, but only the first group actually take it.

The second group are given an inactive substance (or placebo) but they think they have taken the drug. This second group is called a **control group**. If more patients get better in the first group, then the drug has an effect.

4. Designed experiments

In statistics, the word 'experiment' generally refers to a situation where the experimenter carries out some controlled activity and records the results by counting or measuring or simply observing.

Thus an experiment may consist of

- › tossing three coins and recording the number of times two heads show
- › measuring the circumference of oak trees in a wood
- › throwing a dice several times to determine if it is biased
- › recording the side-effects of a new drug
- › investigating whether people are better at remembering words, numbers or pictures.

Explanatory and response variables

In a statistical experiment, one of the variables will be controlled while its effect on the other variable is observed.

The controlled variable is called the **explanatory variable**.

The effect being observed is called the **response variable**.

Example 1

A research team is investigating whether the adding of fish oil to the daily diet of school students increases their IQ. A school of 500 students is selected. Two groups, each of 50 students, are selected at random.

Group A is given a daily ration of fish oil.

Group B is given the same food as Group A, but no fish oil.

- (i) Which group is the control group?
- (ii) What is the explanatory variable in this experiment?
- (iii) What is the response variable?

(i) Control group - Group B

(ii) Explanatory variable - Fish oil

(iii) Response variable - IQ