Chapter 9

Design of experiments

9.1 Observational studies and designed experiments

Investigations are often conducted in which measurements are taken on some variables of interest, without any intervention. An example of this would be if the diets of 100 people were recorded over a year, together with the number of colds they got during that time. In such studies, we can examine the data collected, and seek possible reasons for the outcomes and patterns we observe. This is often a useful thing to do, but it can only *suggest* causal relationships, not prove them.

> An **observational study** is one in which we observe and record what has already happened.

Individuals and variables are measured, but the study does not attempt to influence the response. Observational studies often show an association between two variables, but they cannot in themselves prove cause and effect.

> A **designed experiment** or **trial** is where the investigator intervenes and determines which experimental units receive which treatments.

An experiment is designed to actively seek out which factors (or **treatments**) affect the response variables. This is in contrast to an observational study, where we have to take whatever comes.

Limitations of designed experiments

1. Sometimes they are not feasible, for practical or ethical reasons. For example, it is impractical to use a designed experiment to answer the question: "Does controlling inflation reduce unemployment in Australia?", essentially because replication under the same conditions is impossible.

   As another example, we can't use a designed experiment to investigate whether smoking causes lung cancer, because it would be unethical.

   In cases such as these there is no superior alternative; we have no choice but to rely on observational data.

2. Designed experiments are limited to scientifically testable theories and hypotheses. For example, it would be difficult to carry out a designed experiment to test whether eating organic food makes you a nicer person or not.

3. The criticism is sometimes made that designed experiments suffer from an artificiality: the control (of conditions, etc.) that they entail does not reflect reality, and so what we see in the experiment may not be borne out in practice. It is true that experiments are artificial; but reality is often too complex and messy to enable the sorting out of cause and effect, or the understanding of mechanisms—we need to control conditions for these to happen.

9.2   Experimental units

The first consideration in carrying out any experiment is a clear understanding of what constitutes an **experimental unit**.

> Experimental units are the objects or items to which the treatments are allocated and applied.

This is a simple definition—but it is not always easy to apply correctly!

Plant and animal experiments

In most agricultural field trials, the experimental unit is a plot of land, which usually contains several plants. But an experimental unit need not be a plot of land—it can be a single plant, or even a part of a plant (such as a stem or leaf) if different treatments are applied to the individual parts. With animal trials, an experimental unit can be a paddock of animals, a single animal, or even part of an animal (such as a trial in which different hooves of a sheep or horse are given different treatments).

Food experiments

In experiments involving food, the experimental unit is usually a portion of food. Examples are: a 100 ml sample of carrot juice; a block of cheese; a sealed package of pear slices; a box of lettuces; a container of yoghurt; or a vat of milk from which cheese is made. Sometimes the experimental unit arises from an entire run of a manufacturing process.

Experiments involving people

In most human experiments, where the treatment affects the whole individual, the individual is the experimental unit. If the effect is local in nature, an individual may sometimes be used as a pair of units (for example, right and left arms). At the other end of the scale, sometimes a *group* of people may constitute an experimental unit. For example, in an educational trial involving different teaching techniques, the class of students may be the unit.

Importantly, each experiment is different—and the experimental unit needs to be correctly identified each time.

9.3   Principles of experimental design

The two main principles of experimental design — what we aim for — are **validity** and **precision**. They are best explained by the following diagram:

Be aware of sources   Reduce bias   Reduce scatter
of error   (validity)   (precision)
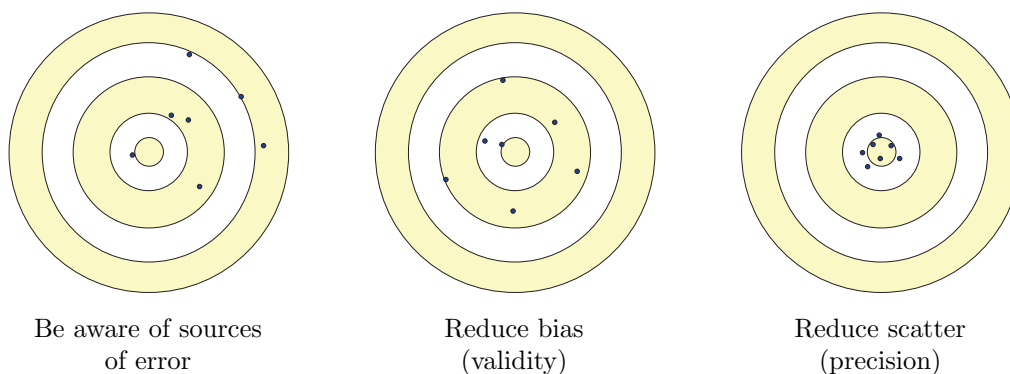
These two principles are best achieved by the following practices:

1. Randomisation (for validity);

2. Replication (for precision);

In most experiments there is a third practice which increases precision:

3. Blocking.

We will consider each of these practices in turn.

9.4   Randomisation

> **Randomisation** is the use of chance to allocate treatments to experimental units. It forms the basis of any valid statistical experiment.

We would not compare the effectiveness of two medications by giving one medication to healthy individuals, and the other to sick individuals. Nor would we give one kind of food to large guinea pigs and another to small guinea pigs in a nutrition experiment. The results of such experiments would be limited in value, because we would not know how much of the observed difference is due to the treatment, and how much is due to the differences in health or size of the subjects. In statistical terminology, the two effects would be **confounded**, i.e. they would not be able to be distinguished from one another. Observational studies have the inherent problem that the observed associations may be due to confounding variables which have not been measured.

A (uniformly) random allocation of treatments to experimental units is defined as one is which all units have an equal probability of receiving any of the treatments. The details of exactly how the randomisation is performed will depend on the particular experimental design.

"Random" does not mean selecting units in a subjective or haphazard way — it must be based on probability to ensure validity. "Manual" methods, such as rolling dice or tossing a coin, are acceptable, but it is usually more convenient to use computer software to perform the randomisation, or in simple cases, random digits.

9.4.1   Completely randomised designs

> In a **completely randomised design**, treatments are allocated to experimental units uniformly at random, subject only to the number of times each treatment is to be used.

Drug trials are quite often conducted in this way: if there is a pool of eligible subjects, the subjects may be allocated at random to one of the "arms", or treatment groups, of the trial. There are a variety of ways that this is done in practice, usually using computer programs, but they all have the feature that each subject has the same probability of being allocated to each group.

Example

Suppose there are 18 experimental units, to which 3 treatments are to be randomised — 6 units per treatment. The units should be labelled (e.g. from 1 to 18) before the randomisation.

Using R

```
> treatment <- c(rep("A", 6), rep("B", 6), rep("C", 6))
> unit <- sample(18, 18)
> data.frame(treatment, unit)
   treatment unit
1          A   10
2          A   13
3          A    7
4          A   17
5          A    4
6          A    8
7          B    1
8          B    5
9          B   18
10         B   12
11         B    9
12         B   16
13         C    2
14         C   15
15         C    3
16         C    6
17         C   14
18         C   11
```

The `sample(x,n)` function selects n numbers uniformly at random, without replacement, from the numbers 1 to x. Therefore `sample(n,n)` randomly arranges n numbers. The `sample(x,n)` function can also draw a sample with replacement if required.

Using Excel

1. Put the numbers 1 to 18 in column A.

2. Enter the formula RAND() in cell B1 and fill down to B18.

3. Copy column B onto itself using Paste Special → Values. If you don't do this step, a new set of random numbers is generated every time a cell in the worksheet is changed, and it can get confusing.

4. Select a cell in either column A or B, and sort the worksheet by Data → Sort → Column B.

We now have (in column A) the numbers 1 to 18 in random order, so we can give the first six to treatment A, and so on.

9.5   Replication

> **Replication** is having a treatment applied to more than one experimental unit.

A single observation or a unique event seldom allows us to draw a conclusion. On repeating any observation, we nearly always find that the results of the second round are not identical with those obtained in the first. This variation has to be taken into account in the analysis of experiments.

There are actually many sources of variation. Successive measurements on the same object will usually yield different readings owing to either human variation in the act of taking observations, instrument variation, or variation in the object itself. When measurements are taken not on the same object, but on a group of seemingly uniform objects, an additional source of variation is introduced: no two objects are really identical, and the variation between objects is generally larger than the variation within objects.

The nature of the random variation between measurements can be observed and described only by repeated observations under given conditions in which all the systematic variation is controlled. The main purpose of experimentation is to separate the systematic and the random variation by repeated observations, i.e. replication. Adequate replication ensures that the random variation averages out sufficiently so that the systematic effects of the treatments can be seen.

The larger the number of replicates, the greater the **precision** will be in comparing treatment effects, i.e. standard errors will be smaller, confidence intervals narrower, and P-values smaller when testing the null hypothesis of no treatment differences.

Replication is also needed to estimate the error variance, so as to obtain a *measure* of precision; we can't estimate the variability of anything from one reading.

9.5.1   Replication—true and false

"False replication" or "fake replication" occurs when experimental units are in some sense measured more than once but treated as separate experimental units. This gives a distorted view of the natural variation present, because measurements on the same experimental unit are related (i.e. not independent), so they usually vary less than measurements on different units. The amount of extra information about the treatments is usually much less when you make more measurements on an experimental unit than when you measure more units. Some examples of false replication are:

- Measuring trees within a plot of a forest experiment and analysing the data as if each tree was an experimental unit.

- Applying milking treatments to only a few cows and then measuring the milk yield of each cow several times to increase the number of "replications".

- Testing one new Holden and one new Ford for acceleration on 5 stretches of road, and making conclusions about the acceleration of all new Holdens and Fords.

- Taking ten samples from a mature block of cheese and ten from an immature block, and performing an analysis of variance to assess the maturing process.

Note that there is nothing wrong per se with measuring experimental units more than once, or taking sub-samples. It is often a good thing to do, and will enhance the quality of your data, by giving a less variable value of the response variable for a particular experimental unit. The problem occurs when these measurements are included in the analysis as if they were separate experimental units. False replication is sometimes described in nicer terms such as "pseudo-replication" or "quasi-replication". These sound more scientific, but don't be fooled—they usually mean false replication.

9.6 Blocking

A **block** is a group of experimental units that are similar in some way that is expected to affect the response to the treatments. Blocks should consist of units which are likely to be more homogeneous than the entire collection of available units. Treatments are then randomly assigned to experimental units within blocks, and so comparisons between treatments are essentially performed within blocks. The purpose of blocking is to reduce the variance of the error, and hence increase *precision*, by accounting for some of the variation between units.

The term "blocking" was first used by R. A. Fisher in agronomic experiments, where blocks were literally blocks of land in a field, each comprising a number of experimental plots. This traditional terminology has continued in experimental design, although other terms such as "matching", "pairing" or "stratification" have become common in some disciplines. Whatever they are called, the important aspect is that such groups contain similar or homogeneous experimental units; this is shown in the following diagram, and in the one on the next page.
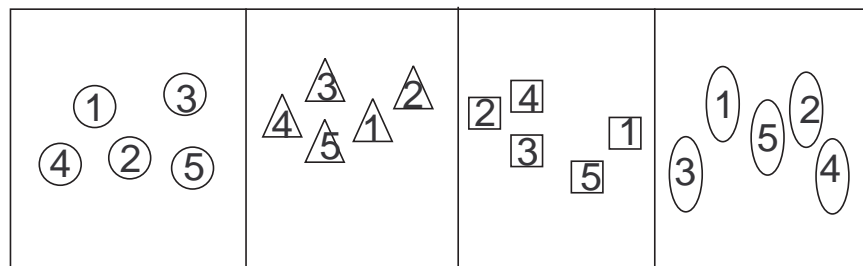
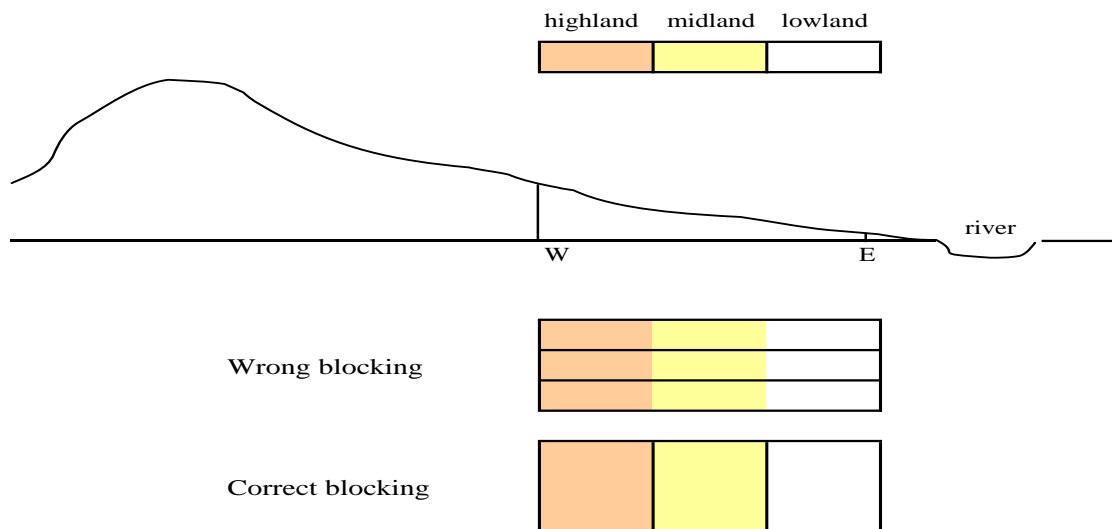Non-homogeneous
experimental units
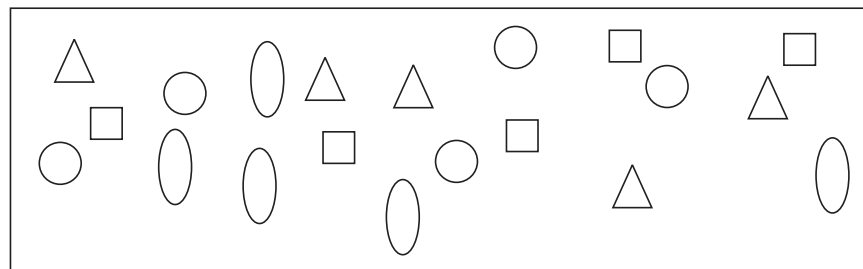
Blocking into
homogeneous groups
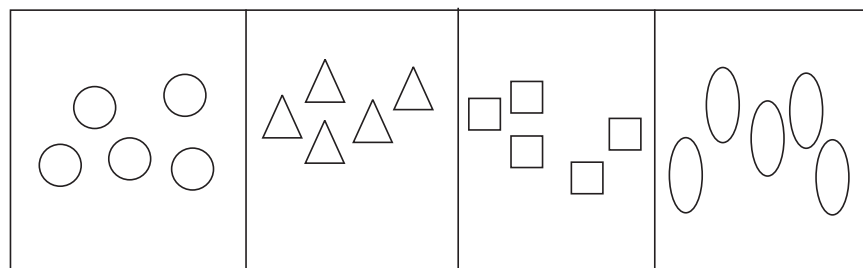
Randomisation
within blocks

Cross section of a piece of land. The section from W to E is to be used for experiments.
It is to be assumed that the fertility and mosture increases from W (highland) to E (lowland).

Apart from blocks of contiguous plots in a field, other examples of blocks in experiments are:

- pairs of identical twins;
- litters of animals;
- groups of vehicles manufactured at the same time;
- periods of time, such as a day;
- batches of material, such as a vat of milk or a truckload of timber;
- groups of people matched by age, gender and ethnicity;
- trays of apples in cold storage.

Precision

There are essentially two ways of increasing the precision of an experiment. The first is by increased replication; no matter how large the error variance may be, in principle the desired precision can be achieved by increasing the number of replicates. However, most experiments have limited resources, and so increasing the replication massively is not feasible. The second way is to use blocking; if within-block variability is substantially less than between-block variability, the gain in precision can be substantial. Improvement of precision by blocking is usually cheaper than improvement by increasing the number of experimental units.

9.6.1   Randomised block designs

In a **randomised block design**, each treatment occurs the same number of times (usually once) in each block. Treatments are randomly assigned to experimental units within each block separately. A randomised block design with exactly one replicate of each treatment in each block is sometimes called a **randomised complete block design**. The blocks are *complete* in the sense that all treatments are present in each block. This is not always possible in practice, and blocks with fewer units than there are treatments are called **incomplete blocks**.

The randomised block design revolutionised experimental design 70 years ago. It was the first valid design to estimate experimental error and test the significance of treatment effects with reasonable precision in spite of the heterogeneity of the experimental units. In a sense, it is still the most fundamental design in most forms of experimentation.

In some fields of study (especially where the experimental units are not physical areas) this type of design is sometimes called "stratified randomisation". The principle is exactly the same—similar units are grouped into strata (i.e. blocks) and a separate randomisation performed within each stratum. When there are just two treatments, it is sometimes called "matched pairs", because pairs of experimental units (often people) are matched (i.e. blocked) according to one or more factors which are expected to affect the response variable.

Example: powdery mildew on tomatoes.

Powdery mildew is a fungal disease which affects tomatoes and other crops. A randomised block experiment is set up to test eight chemical treatments to see how well they control the disease. Forty plots of tomato plants are available, in 5 blocks of 8 plots each. The trial area is on a slope, and because slope affects many outcomes of field trials (e.g. greater moisture towards the bottom may lead to more disease), it is wise to arrange blocks according to the slope, as in the following diagram:

slope ————————➤
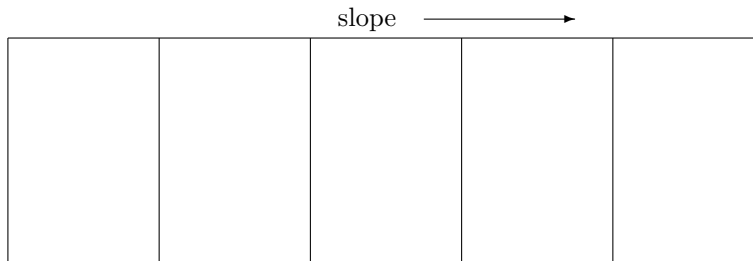
Assuming that the tomato plants can be arranged in any way, how should the blocks be divided into plots? (Remember the main principle of blocking: *the experimental units should be as similar as possible within the block*).

Randomisation in R (only the first 2 blocks are printed below)

```
> block <- rep(1:5, each = 8)
> plot <- rep(1:8, 5)
> treatment <- c(sample(8, 8), sample(8, 8), sample(8, 8), sample(8,
+     8), sample(8, 8))
> data.frame(block, plot, treatment)[1:16, ]
```

```
   block plot treatment
1      1    1         1
2      1    2         4
3      1    3         2
4      1    4         3
5      1    5         6
6      1    6         8
7      1    7         5
8      1    8         7
9      2    1         1
10     2    2         3
11     2    3         8
12     2    4         4
13     2    5         2
14     2    6         6
15     2    7         5
16     2    8         7
```

Note that in the R code above, the repetition of `sample(8,8)` is rather clumsy – if there were more than 5 blocks, we would probably use a loop.

9.6.2   Latin square designs

Example: Potato wireworm experiment

Potato wireworm, a serious pest of potato crops, can be controlled by chemicals, but such chemicals may be harmful to the environment in the long term, and alternative "cultural treatments" need to be tested. An experiment is set up with four treatments: 1. a control (no treatment applied); 2. the standard chemical; 3. and 4. two methods of "hilling" soil around the plants (cultural treatments). Each plot consists of 5 rows × 10 plants, and the response variable is the number of plants out of 50 seriously affected by wireworm at harvest. There are 16 plots, arranged in a 4 × 4 square, with a slope running from one side of the square to the other. A suitable experimental design would be a randomised block design, with blocks of 4 plots as shown in the diagram.

7

slope ⟶

3	2	3	1
1	1	2	2
4	3	1	4
2	4	4	3

Now suppose that in addition to the slope, there is a gradient in soil texture which is perpendicular to the slope, starting from clay at the top of the diagram to loam at the bottom. If soil texture influences the effect of the treatments, we then have an additional blocking factor which needs to be accounted for. If we ignore it, some treatments may be advantaged or disadvantaged; for example, in the above design, treatment 4 occurs only in the bottom half of the diagram, and so would not be used at all with clay soil. A type of experimental design which can be used in situations like this is the Latin square.

A **Latin square** design incorporates two blocking factors, which are usually represented as **rows** and **columns**. There must be as many levels of each blocking factor as there are treatments, and each treatment must appear exactly once in each row and in each column.

The following Latin square enables both blocking factors (slope and soil texture) to be accounted for in the potato wireworm experiment:

slope ⟶

clay
↓

3	1	2	4
1	4	3	2
4	2	1	3
2	3	4	1

loam

The rows and columns of a Latin square do not have to be physical rows and columns. They can be periods of time, groups of people or animals, and so on—in fact, any blocking factor is a potential row or column of a Latin square, as the following example illustrates.

Example: food supplements and milk yield of cows.

An experiment was conducted to comparing the effectiveness of three food supplements for increasing the milk yield of dairy cows. The supplements (A, B and C) were given to three cows, and repeated in three successive time periods while rotating the cows. Milk yields, in grams/day, are given below. The cow (1, 2, or 3) was one blocking factor and the time period (I, II, or III) was the other. The supplements were assigned according to a Latin square, as shown. The data are also shown.

Cow	Period					
	I		II		III	
1	A	608	B	885	C	940
2	B	715	C	1087	A	766
3	C	844	A	711	B	832

Randomisation of treatments in a Latin square

The treatments cannot be randomised separately in each row or each column, because the arrangement may not end up with complete blocks in both directions. The whole square has to be considered together, and the easiest way to do this is by starting with a standard "generating" Latin square, and randomly arranging the rows and columns. The procedure is as follows:

- Randomly choose a standard generating Latin square of the correct size from a list;

- Randomly reorder the rows;

- Randomly reorder the columns.

Some statistical packages have capabilities for design of experiments, including Latin squares.

Multiple Latin squares

Latin squares can also be used if the number of levels of one of the blocking factors is an exact multiple of the number of levels of the other blocking factor and the treatment. For example, suppose that in the food supplement experiment, six cows were available instead of three. Two Latin squares could be used, as shown below. Note that the two squares should have separate randomisations of treatments to plots.

Cow	Period		
	I	II	III
1	A	B	C
2	B	C	A
3	C	A	B
4	B	A	C
5	A	C	B
6	C	B	A

9.7   Balance

In all our discussion so far, it has been assumed that each treatment in an experiment is randomised to the same number of experimental units. In other words, the replication is the same for each treatment. This is known as *balance*.

> **Balance** means applying each treatment to the same number of experimental units.

A balanced design has two advantages:

1. It simplifies the analysis and to some extent, the interpretation. This is desirable for the experimenter, though it is not as crucial as it was in the past, when calculations were done by more primitive methods. A modern statistical package can generally cope with lack of balance.

2. It gives the most precise comparisons. For example, allocating 20 farms to each of feed types A and B results in a more precise comparison than allocating 25 farms to A and 15 farms to B (or vice-versa). However, allocating 25 farms to A and 20 farms to B gives greater precision than allocating 20 to each.

Balance is therefore desirable *within the limitations already placed on an experiment*, e.g. cost, or available space. But if more experimental units become available within those limitations, they should in general be used, even if it results in lack of balance. In other words, balance is good—but increased replication is even better.

9.8   Controls

"Gastric freezing" was a treatment for ulcers in the upper intestine. The patient swallowed a deflated balloon with tubes attached, then a refrigerated liquid was pumped through the balloon for an hour. The idea was that cooling the stomach would reduce the production of acid, and so relieve ulcers. An experiment reported in the *Journal of the American Medical Association* indicated that gastric freezing did reduce acid production and relieve ulcer pain. The treatment was safe and easy and was widely used for several years. However, the experiment was poorly designed, with no controls.

A later experiment divided ulcer patients into two groups. One group was treated by gastric freezing as before. The other group received a **placebo** treatment in which the liquid in the balloon was at room temperature rather than freezing. The results: 34% of the 82 patients in the treatment group improved, but so did 38% of the 78 patients in the placebo group. This and other properly designed experiments showed that gastric freezing was no better than a placebo, and its use was abandoned.

In the earlier experiment, the patients' responses may have been due to the **placebo effect.** A placebo is a dummy treatment, and the response to a dummy treatment is the placebo effect. Many patients respond favourably to *any* treatment, even a placebo, presumably because of trust in the doctor or the procedure, and expectation of a cure. The placebo effect is well documented, even for treatments which are quite invasive.

A placebo is a form of a **control** or control treatment. The idea behind a control is that when treatments are to be compared, all other variables should be held as constant as possible, i.e. *controlled*. The argument of causation is sustainable because the experiment compares "what happened with what would have happened without the intervention".

Experiments without controls have been shown to be biased in favour of the treatment being tested. This is also true for "historical controls", in which the effect of a control treatment is estimated from past records etc. rather than being included in the experiment. Confounding with variables that change over time is a major weakness of trials which use historical controls.

What form should the control take?

A control does not have to be no treatment at all. It can be the currently accepted treatment or the "industry standard"; it depends what you want to compare the new treatment with. For example, in assessing the efficacy of a new insecticide, an obvious control would be an established insecticide. However, the number of insects when no insecticide at all is used may also be of considerable interest. It is sometimes appropriate to include both forms of control in the experiment.

Sometimes a placebo is unethical, for example, in experiments comparing treatments for cancer. A suitable control in such trials would be the currently used treatment. Also, in experiments involving the testing of a new drug, it is often more meaningful to compare the new drug with the drug currently prescribed, rather than with a placebo. It is relatively easy to show that a new drug is better than a placebo, but not so easy to show its superiority to another drug.

Control treatments should be included as an integral part of the experimental design, not shuffled to one side in any way. They should also be included in the statistical analysis, though there are some exceptions, such as when all the control observations are zero.

Blind studies

In a medical experiment, if people know they are receiving a placebo, then the placebo effect is less likely. For this reason the subjects should not be told which treatment they are given, if at all possible. Such studies are referred to as **blind** studies. Obviously the placebo should match the treatment in as many ways as possible; for example, if it is a tablet, it should be the same size, shape and colour.

Even better are experiments in which neither the subject, nor the experimenters working with them, know which treatment they are on, until the study is completed. Such studies are referred to as **double-blind** studies. If, for example, a doctor believes in the value of a particular treatment, they may treat or evaluate a patient more favourably. It is therefore desirable for them to be "blind" if possible.

For experiments involving animals, plants, or inanimate material, blindness of the subject is not an issue. However, it is still desirable for the experimenter to be blind where possible. For example, when injecting an animal, or scoring disease on a plant, or assessing the fatigue of a metal panel, it is better if the treatment is unknown to the person performing the task.

9.9   Exercises

1. For the two studies below, consider the following questions:

   (a) Is it an observational study or a designed experiment?

   (b) Is the random variable discrete or continuous?

   (c) What parameters are we likely to be interested in estimating?

   Vitamin C in tomato juice

   From the shelf of a supermarket, all the cans of a certain brand of tomato juice on the shelf on a particular day were taken and analysed. There were 17 cans, and their vitamin C readings (mg/100 gm) were as follows:

16	22	21	20	23	22	17	15	13
22	17	18	29	17	22	16	23	

Rainfall and cloud seeding

A study was conducted to investigate the effectiveness of cloud seeding. The rainfall in hectare-metres was recorded from 52 clouds, of which 26 had been chosen randomly and seeded with silver oxide.

Rainfall from control clouds				Rainfall from seeded clouds			
0.60	101.27	4.47	10.61	24.23	36.94	29.59	202.03
11.59	3.18	2.65	45.43	3.99	33.51	3.83	0.50
3.54	42.15	5.01	0.60	33.51	334.96	24.49	4.95
3.49	39.19	146.72	19.92	119.32	0.94	14.52	40.76
1.40	2.11	3.21	29.80	52.46	15.81	31.11	85.81
0.12	18.03	2.98		207.13	14.46	2.13	
9.91	5.77	8.36		14.07	59.67	11.27	

2. A study from the Education faculty of the University of Melbourne found that children who had been in child-care from an early age experienced more educational difficulties than those who had not. It was also found that children cared for by a private nanny had fewer difficulties. The report concluded that child-care caused educational difficulties. Do you think this conclusion is justified? Can you identify any possible confounding variables?

3. How random are the following methods of allocating treatments to experimental units?

   (a) Pieces of paper drawn from a hat.

   (b) Rolling a die.

   (c) Thinking of numbers between 1 and 10 in your head.

   (d) Allocating cows to treatments according to the order in which they enter the milking shed.

   (e) A Tattslotto draw.

   (f) Walking through a field selecting plants all over the place.

   (g) Selecting digits from numbers in the telephone book.

   (h) Selecting students from a class whose surnames start with A or B.

   (i) Reaching into a closed box to select half the chicks for a bright light treatment which may increase growth (the remaining chicks get normal light).

4. For the two studies below, assess whether the claimed replication is false, and how many (true) experimental units there are.

   (a) **Bacteria in hen houses**

   Six different disinfectants are being studied for their effectiveness in reducing bacteria in hen houses. One of 6 available henhouses is randomly selected to receive the first disinfectant, another to receive the second, and so on. A week after applying the disinfectants, bacterial counts are taken on the feathers of 4 hens from each house, so that the data for analysis consist of 4 readings for each of the 6 disinfectants.

   (b) **Drought-affected lemon trees**

   An experimenter, Mary, visits 6 citrus orchards in which drought is causing problems with the lemon trees. At each orchard she selects 3 rows of lemon trees, with 10 trees in each row. Mary then randomly allocates one of 3 mulches to each row, and two months later notes the health of each tree. The final data consist of 180 readings, or 18 if they are averaged across each row.

5. The progress of a type of cancer differs in women and men. A clinical trial is to be conducted to compare three therapies for this cancer. How should the subjects be allocated to the therapies? What other variables might be important to consider in the design of the trial?

6. An experiment is to be conducted to examine the efficacy of three types of sprays (S1, S2 and S3) in controlling fungal rots on strawberries. Three adjacent rows of strawberries are available, each with 24 plants. Sprays can be applied to individual strawberry plants. The response variable is the proportion of strawberries with rots. For the designs 1 to 4 below, specify the experimental unit, any blocking factor, and how many replications of the treatments there are.

   (a) The sprays are randomly allocated to rows, and 8 strawberry plants randomly selected from each row for assessment.

   (b) Each row is divided into 3 plots of 8 plants each. The sprays are randomly allocated to plots within each row.

   (c) The sprays are randomly allocated to individual plants across the entire trial area.

   (d) The trial area is divided into 4 quarters, each consisting of 3 rows × 6 plants. Within each quarter, the sprays are randomly allocated to rows.

   (e) At the conclusion of the trial it is found that one end of the trial area is substantially wetter than the other, and moisture generally increases the risk of fungal rots. Which design had the most appropriate blocking to account for this?

   (f) For part (c), perform the randomisation
       (i) using R.
       (ii) using Excel.

7. A cheese manufacturer wants to test two additives for their effect on texture of cheese. There are three treatments—additive A, additive B, and no additive. A different batch of milk is delivered to the factory each morning. Because of the complexity of the machinery, only 3 cheese-making runs are possible on each day. The experiment can run for one week, from Monday to Friday.

   (a) How would you design this experiment?

   (b) Perform the randomisation in R.

   (c) Each cheese-making run requires one employee to supervise it. The manufacturer can call on any of three different employees at any time, but each employee can work a maximum of three days in total. How would you allocate them to the runs?

8. In order to assess the effects of exercise on reducing cholesterol, a researcher sampled 50 people from a local gym who exercised regularly and 50 people from the surrounding community who did not exercise regularly. They each reported to a clinic to have their cholesterol measured. The subjects were unaware of the purpose of the study, and the technician measuring the cholesterol was not aware of whether subjects exercised regularly or not. This is

   A. an observational study;
   B. an experiment, but not a double-blind experiment;
   C. a double-blind experiment.

9. A new headache remedy was given to a group of 25 patients who suffered severe headaches. Of these, 20 reported that the remedy was very helpful in treating their headaches. From this information you can conclude

   A. the remedy is effective for the treatment of headaches;
   B. very little, because the sample size was too small;
   C. very little, because there was no control group for comparison;
   D. very little, because there were no objective measurements taken.

10. To determine whether a particular hormone injection produces a change in iron level in the blood of mice, 20 mice are measured for their iron level, before injection. Three days after injection, the iron level is measured again, and for each mouse the difference (before vs after) is calculated. The results are statistically analysed, and used to conclude that the hormone injection has caused a change in iron level.

   Is the conclusion justified?

11. An apple orchard has 32 trees set aside for an experiment which aims to examine the effect of mulching on tree growth. There are 4 mulching treatments: 1. Control (no mulch); 2. Wood chips; 3. Garden compost; 4. Clippings from a local council collection. The trees are in a 4×8 rectangle, as shown in the diagram below. The ground slopes down from the left to the right of the diagram. The experimenter has resources to maintain 16 plots, each consisting of 2 trees.

(a) Construct an appropriate experimental design, giving an example of the allocation of treatments to plots.

```
o   o   o   o   o   o   o   o
o   o   o   o   o   o   o   o
o   o   o   o   o   o   o   o
o   o   o   o   o   o   o   o
```
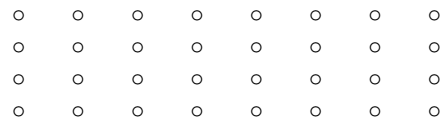
(b) If the slope of the ground was from the top to the bottom of the diagram (instead of from left to right), construct an appropriate design.

```
o   o   o   o   o   o   o   o
o   o   o   o   o   o   o   o
o   o   o   o   o   o   o   o
o   o   o   o   o   o   o   o
```
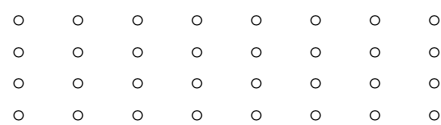
(c) If the slope of the ground was in both directions (top to bottom and left to right), construct an appropriate design.

```
o   o   o   o   o   o   o   o
o   o   o   o   o   o   o   o
o   o   o   o   o   o   o   o
o   o   o   o   o   o   o   o
```