The simplest type of experimental design: completely randomized, single-factor

7

- We begin by introducing one-factor completely randomized designs with two levels of the factor as a simple but powerful means of organizing an experiment (Section 7.1).
- Key to such designs, and to many experiments, is randomization to deal with confounding factors, and we discuss how to do this effectively (Section 7.2).
- We then explore how the number of levels of the factor can often usefully be increased beyond two (Section 7.3).
- We finish this chapter with an overview of the types of situation that would encourage you to adopt this type of design, and the types of situation that might cause you to instead adopt one of the designs described in later chapters (Section 7.4).

So far in this book we have focused on the basic ingredients of good experiments, such as proper replication and effective controls. We now turn to specific experimental designs to see how these ingredients can be combined to answer biological questions. One way to think about different experimental designs is as different ways to deal with all of the potential sources of confounding variation and noise. As we will see (in Section 7.4), the design that you choose will be determined by the question that you are asking, and the biology of your study system (in particular the potential sources of variation); and will determine the kind of statistical analysis that you are able to do. At first sight, the experimental design literature can seem daunting, and full of complex terms and jargon. However, there is really nothing difficult about the ideas behind these terms. We will introduce and define the commonest terms as we go along. In this chapter we will go through the process of setting up the simplest experimental design, bringing together ideas from previous chapters. In subsequent chapters we will build on this simple design, at each stage focusing on how the additional complexity of the design deals with variation and affects the questions that we can ask.

7.1 Completely randomized single-factor designs

Let's start with a simple biological question:

Does plant feed affect the growth of tomato plants?

An ***n*-factor** or ***n*-way design** varies *n* different independent factors and measures responses to these manipulations. For example, an experiment looking at the effects of both diet and exercise regime on dogs' health would be a two-factor design. If we consider three different diets, then there are three **levels** of the experimental factor 'diet'. If *n* is greater than one (i.e. multiple factors are involved), then the experiment can be referred to as a *factorial* experiment.

To answer this question we are obviously going to need to apply some plant feed to some tomato plants and measure some aspect of their growth, perhaps their dry weight after a few weeks of growing. We are also going to need some form of control so that we know what growth we might expect in the absence of the plant feed. In this case our control would be tomato plants that are not given the plant feed but are otherwise handled in an identical way.

Now we need to introduce some jargon. Our experimental manipulation affects a single factor, the presence or absence of plant feed, and so this design is known as a **one-factor design** (you will also see it referred to as a **one-way design**). Our factor has two **levels**, since our experiment has two treatments ('plant feed' and 'control').

Of course for all the reasons outlined in Chapter 4, we can't just compare the growth of two plants, one with and one without plant feed. We need several plants in both our 'plant feed' and 'control' treatment groups, or to use more jargon, our experiment needs to be replicated. In this case we have 80 tomato plant seedlings all ready to be used in our experiment. All we need to do is allocate them to our two treatment groups, apply the plant feed to those in the plant-feed treatment group, and our experiment is ready to go. However, an important question remains, how should we decide which plants go into which treatment group? In this simple experimental design, the answer is equally simple: **randomization**.

A simple one-way design allows us to compare the effects of different levels of a single factor.

7.2 Randomization

7.2.1 Randomizing study subjects

Randomizing the allocation of experimental subjects to experimental groups simply means that each subject is allocated at random to a group, with any subject being equally likely to end up being allocated to any of the groups. This process is also called *random allocation*, *simple random allocation*, *random assignment*, or *random placement*.

The experimental subjects of any study will vary for many reasons that are nothing to do with the factors that we are interested in. The chickens in Chapter 4 produced eggs with different shell thickness for many reasons that were nothing to do with the kind of feed they had been given. Similarly, in our current study, plants will vary in their growth for all sorts of reasons that have nothing to do with whether they are treated with plant feed or not. Our challenge is to separate the effects we are interested in (in this case any effect of plant feed) from all of this background noise. We want to ensure that there are no systematic differences between the plants in our two treatment groups from any of these other sources of variation in growth, otherwise any differences caused by our experimental treatment could be confounded with these other factors and cause us to draw erroneous conclusions. Imagine that there are small

Q 7.1 We have assumed that the experiment should have a concurrent control (in the jargon we introduced in Section 2.3.1). Would you be tempted to halve the size of the experiment and use a historical control instead?

differences in the quality of the compost in each pot that our plants are growing in, and that this has an effect on the growth of the plant. If, for whatever reason, the plants with slightly better compost end up in the plant-feed group, whilst those with slightly worse compost end up in the control group; then we are likely to see a difference in growth rate between our groups, even if the plant feed has no effect at all. This is true for any of the other factors that might affect plant growth. So our task is to ensure that the only systematic difference between the plants in each treatment group is the factor of interest (presence or absence of plant feed). The simplest solution to this problem is to let chance decide: to randomly allocate subjects to each treatment group (say by tossing a coin for each plant). By randomly allocating plants to treatment groups we randomly spread the variation due to other potentially confounding factors across treatments. This will minimize the chance of any systematic differences in other factors affecting our conclusions. The major advantage of randomization is that it not only deals with sources of variation that you might be suspicious about, such as compost quality, but also deals with all of the many other potentially confounding factors that you don't know about: everything gets randomized. It will probably not surprise you to learn that this use of random allocation to treatments is what leads to this type of study being referred to as a *completely randomized design*, since subjects are assigned to treatment groups completely at random.

Randomization is the key to avoiding confounding factors in planned experiments.

7.2.2 Randomizing other aspects of your study

Randomizing the allocation of experimental subjects to treatments is a simple and powerful way of avoiding systematic differences amongst treatment groups. However, it is only the start. Our tomato plant study is going to take place in a greenhouse. Each plant will be placed in an individual pot in the greenhouse. The size of the experiment means that our plants will need to be spread over two benches within the greenhouse. After 2 months, we will harvest the plants and measure their dry mass. There are many aspects of this experimental set up that may affect plant growth. Maybe the greenhouse is slightly warmer or lighter at one end than the other, or maybe the glass above one of the benches is slightly older and so lets through a different colour of light than the glass above the other bench. The list could obviously go on. If we put all of our plant-food-treated plants on one bench, and the control plants on the other bench, then our replicates have effectively become pseudoreplicates, since treatment effects are confounded with systematic differences between benches. The easiest way to avoid this is to carry out further randomization. We can randomize the allocation of plants to benches, and the position of plants on a single bench, such that any plant has an equal probability of being in any part of the greenhouse.

However, the need to randomize does not just apply to the setting up of the experiment. It can equally be applied to taking the measurements at the end of a study. There are numerous reasons that can lead to the accuracy of measurements

differing through time. Maybe the spectrophotometer that you are using is old, and gets less accurate as time goes on. Or maybe you have spent 10 hours looking down a microscope counting parasites, and the inevitable tiredness means that the later counts are less accurate. Or maybe after watching 50 hours of great tit courtship behaviour on video you become better at observing than you were at the beginning. In the current study, perhaps the balance that we are using to weigh our dry tomato-plant samples becomes less accurate as time goes on, or the dried samples take in moisture through time, slowly increasing in mass. Whatever the reason, this means that if you take all the measurements on subjects from one treatment group first and then all those from another treatment group, you risk introducing systematic differences between the groups because of the changes in the accuracy of your measurement methods. It is far better to organize your sampling procedure so that subjects are measured in a random order (see Section 11.1 for more about this sort of problem). In short, proper randomization at all stages of study is one of the most simple and powerful tools available to us to avoid problems of systematic bias and confounding variables. It ensures that sources of non-independence between subjects do not creep into our study and lead us to unwittingly pseudoreplicate. We emphasize the word 'proper' here, because inadequate randomization is probably one of the most common flaws in experimental design, in experiments by everyone from undergraduates to the most eminent professors.

Randomization isn't just the way to avoid confounding variables when allocating experimental subjects to different treatments; it is the key to avoiding confounding variables creeping in at all stages of your experiment.

7.2.3 Haphazard allocation

The major problem that arises in randomization is that, for many people, when they say that they randomly allocated subjects to treatments what they actually mean is that they employed *haphazard* allocation. So what is the difference? Let's briefly consider a different study. Imagine we have a tank full of 40 hermit crabs that we want to use in a behavioural experiment. The experiment requires that we allocate them to one of four different treatment groups. Random allocation would involve something like the following:

- Each crab would be given a number from 1 to 40.
- Pieces of paper with numbers 1 to 40 are then placed in a hat.
- Ten numbers are drawn blindly, and the crabs with these numbers allocated to treatment *A*.
- Ten more are drawn and allocated to treatment *B*, and so on until all crabs have been allocated.

Of course, we could equally have drawn the first number and put that crab in treatment *A*, the second in *B*, third in *C*, fourth in *D*, and then repeated until all treatments were

filled. However, what is important is that each crab has the same chance as any other of ending up in any treatment group and so all of the random variation between crabs is spread across treatments. Another way to think about a random sample is that the treatment group selected for one subject has no effect whatsoever on the treatment group selected for the next subject.

This randomization procedure contrasts with a haphazard sampling procedure. A typical haphazard sampling procedure would involve placing one's hand in the tank and grabbing a crab without consciously aiming for a particular individual. *This will not give you a random sample.* Even if you shut your eyes and think about your bank balance, this is still not going to give you a random sample. The reason for this is that there are a large number of reasons that could cause the first crabs to be picked out to be systematically different from the last crabs. Perhaps smaller crabs are better at avoiding your grasp than larger ones. 'Hold on,' you say, 'what about if I pick the crabs out in this way and then allocate them to a group without thinking, surely this will give me random groups?'. Well, maybe it will, but probably it won't, depending on how good you are at not thinking. It is very tempting to subconsciously think, 'I've just allocated a crab to treatment *A*, so I guess the next one should be given a different treatment'. This is not random. So if you really want random groups, the only way to get them is to randomize properly, by pulling numbers from a hat, or generating random sequences on a computer. It may seem like a nuisance, and it may take you an extra half an hour, but an extra half hour is a small price to pay for being confident that the results that you obtain after weeks of work actually mean something.

Always randomize properly, don't be tempted into using easier procedures which are open to bias.

7.2.4 Balanced and unbalanced allocation

Now that we understand the importance of randomization, we allocate our tomato plants at random to our two different treatment groups. One way to do this would be to take each plant in turn and flip a coin. We then allocate the plant to the plant-feed group if we see heads, and the control is we see tails. This procedure would certainly achieve a random allocation of plants, but because of the vagaries of random sampling we are very likely to end up with slightly different numbers of plants in each treatment group: perhaps 36 in one and 44 in the other. In statistical jargon, our experiment is **unbalanced**. This is not a fatal flaw in our experiment, but it does have drawbacks. Generally the statistical methods that we will ultimately use to analyse the data are most powerful, and least influenced by violations of assumptions, when there are equal numbers in each group. Only in unusual circumstances should we deliberately seek to use unequal numbers. Thus aiming for equal numbers, or as a statistician would say for a **balanced design**, should be a general tenet of your experimental designs. A better way to assign the plants would be to add a constraint that the final number in each group must be 40. Practically, this could be done by numbering all of the plants, then putting these numbers on identical pieces of card in a bag, mixing them thoroughly

A **balanced** experimental design has equal numbers of experimental units in each treatment group; an **unbalanced** design does not.


You can find one example of when you might use unbalanced groups for ethical reasons in Section 7.2.4 of the supplementary material. Go to: **www.oxfordtextbooks.co.uk/orc/ruxton4e/**.

then drawing them out. The first 40 numbers to be drawn would be assigned to the first treatment group, and so on.

We have now designed our study, so let's revisit some of the jargon. We have used *a completely randomized one-factor design, with a factor that has two levels. Our design is also fully replicated and balanced.* You can see how the terminology can become complicated very quickly, but the underlying logic is straightforward. This design allows us to look for differences between two groups. As we emphasized in Statistics Box 2.1, the final component we need to think about as we plan our study is how we will analyse our data. A major advantage in using simple, well understood designs is that they lead very naturally to statistical tests that we can use to answer the question once we have our data. The most obvious choices to analyse data from a single-factor design with two levels of the factor would be to use a *t*-test, or its non-parametric alternative, the Mann-Whitney U test.


Always aim to balance your experiments, unless you have a very good reason not to.

7.3 Factors with more than one level

Now that we understand this simple design we can start to expand it to address other questions. One obvious way we might expand our study is to add additional levels for our factor. Thus, suppose that we wanted to explore the effects of supplying the plant feed at different rates (let's call the rates *low*, *medium*, and *high*). In principle we could repeat the study above for each rate in turn, but a much better alternative would be to examine the feeding rates in the same study, by including four groups: (i) control (no plant feed), (ii) feed provided at low rate, (iii) feed provided at medium rate, and (iv) feed provided at high rate. This is still a one-factor design because we are still only manipulating a single factor (rate of plant feed delivery), but now our factor has four levels instead of two. There are multiple advantages to this; most obviously it allows us to compare each feeding rate to the others, as well as to the control. That is, we can answer several questions with a single study:


Q 7.2 If we carried out three separate experiments, one immediately after the other, each comparing a control group to either a low, medium, or high rate of plant-feed application, couldn't we still compare the effects of different feeding rates?

Does providing plant feed at a low rate affect the growth of tomato plants?

Does providing plant feed at a medium rate affect the growth of tomato plants?

Does providing plant feed at a high rate affect the growth of tomato plants?

Do the three feeding rates differ in their effect on the growth of tomato plants?

This single four-level experiment is also a more efficient use of plants than three two-level ones, since we can compare the three feeding rate treatments to the same control plants rather than needing three separate sets of control plants. Of course, everything that we discussed above about setting up the simple design applies equally to this design. We would randomly allocate our plants to the treatment groups, and randomize everything else that we could. We would also probably have equal numbers of plants in each treatment group,

ensuring that our design is balanced. Putting all of this together we could describe our experiment as a *completely randomized one factor design with four levels of the factor and that it is fully replicated and balanced.* As for the analysis, in this case your design naturally leads to either a one-way analysis of variance (ANOVA) or a Kruskal-Wallis test.

A larger experiment with several levels of the same factor is always more efficient than a collection of smaller experiments.

7.4 Advantages and disadvantages of complete randomization

Imagine that you have twenty plants for a growth trial, and you split them randomly into four groups, each of which is given a different feed treatment. If all goes well, you will have twenty growth rates to analyse at the end of the experiment. However, if a plant becomes diseased, or is knocked off the bench, or some other catastrophe befalls it such that the growth rate of that plant cannot be obtained or no longer provides a fair reflection of the treatment, then you will have fewer measurements to work with. The lost or omitted cases are generally called *missing values* or *drop-out cases*. Obviously, experiments should be designed to minimize the likelihood of such drop-outs occurring. However, accidents do happen, so you should seek experimental designs where a small number of drop-outs do not have a devastating effect on the usefulness of the remaining data.

Q 7.3 What could we do in our tomato plant trial to minimize drop-outs?

The attraction of complete randomization is that it is very simple to design. It also has the attraction that the statistics that will eventually be used on the data are simple and robust to differences in sample sizes, so having missing values is less of a problem. Also, in contrast to some designs that we will see in Chapter 10, each experimental unit undergoes only one manipulation; this means that the experiment can be performed quickly and that ethical drawbacks to multiple procedures or long-term confinement are minimized. It tends to be that the probability that a subject will drop out of the study increases with the length of time over which they need to be involved in the study, so the drop-out rate should be lower than in some other types of experiments.

The big drawback to full randomization is that we are comparing between subjects. As we saw in previous chapters, between-subject variation due to random factors makes it difficult for us to detect the effects of the manipulations that we have carried out, and so reduces our statistical power. If the growth rates of tomato plants are highly variable due to other factors, this may make it very hard for us to detect the effects of our different feeding regimes. With a completely randomized design, then the only real option to deal with this problem is to increase sample size, with all of the associated ethical, conservation, or financial costs this may incur (and the experiment will also use more of our time). An alternative approach is to consider using a more complex experimental design that explicitly takes some of these sources of variation into account. We will explore these in Chapters 9 and 10; but before that we will spend Chapter 8 exploring why you might often want to vary more than a single factor within an experiment, and how you might most effectively design experiments that incorporate multiple factors.

Q 7.4 A researcher is interested in whether competition affects the development time of beetle larvae that complete their larval development within mung beans. To do this she allows adult females to lay eggs on mung beans for several hours, and then examines each bean, counting the number of eggs that it carries. She randomly picks 50 of the beans carrying single eggs for her low-competition treatment, and 50 of the beans carrying two eggs for her competition treatment, and then measures the time larvae take to develop in these two treatments. Do you see any problems in this approach? How might it be improved upon?

If between-subject variation is low, then complete randomization can be quite powerful, hence it is commonly used in laboratory studies. It is also attractive if you expect high drop-out rates among subjects. Field studies and clinical trials tend to suffer more from between-subject variation, and so for these the designs discussed in Chapters 9 and 10 may be more attractive.

Summary

- A simple one-way design allows us to compare the effects of different levels of a single factor.
- Randomization is the key to avoiding confounding factors in planned experiments.
- Randomization isn't just the way to avoid confounding variables when allocating experimental subjects to different treatments; it is the key to avoiding confounding variables creeping in at all stages of your experiment.
- Always randomize properly, don't be tempted into using easier procedures which are open to bias.
- Aim to have a balanced (or near-balanced) experiment with the same (or very similar) numbers of experimental subjects in each treatment group.
- A larger experiment with several levels of the same factor is always more efficient than a collection of smaller experiments.
- If between-subject variation is low, or drop-out rates are expected to be high, then complete randomization is an attractive technique. Otherwise the methods considered in Chapters 9 and 10 may be more attractive.