1 Introduction

1.1 Why do life scientists need to know about experimental design and statistics?

If you work on living things, it is usually impossible to get data from every individual of the group or species in question. Imagine trying to measure the length of every anchovy in the Pacific Ocean, the haemoglobin count of every adult in the USA, the diameter of every pine tree in a plantation of 200 000 or the individual protein content of 10 000 prawns in a large aquaculture pond.

The total number of individuals of a particular species present in a defined area is often called the **population**. But because a researcher usually cannot measure every individual in the population (unless they are studying the few remaining members of an endangered species), they have to work with a very carefully selected **subset** containing several individuals (often called **sampling units** or **experimental units**) that they hope is a **representative sample** from which they can infer the characteristics of the population. You can also think of a population as the total number of artificial sampling units possible (e.g. the total number of $1m^2$ plots that would cover a whole coral reef) and your sample being the subset (e.g. 20 plots) you have to work upon.

The best way to get a representative sample is usually to choose a number of individuals from the population at **random** – without bias, with every possible individual (or sampling unit) within the population having an equal chance of being selected.

The unavoidable problem with this approach is that there are often great differences among sampling units from the same population. Think of the people you have seen today – unless you have met some identical twins (or triplets etc.), no two would have been the same. This

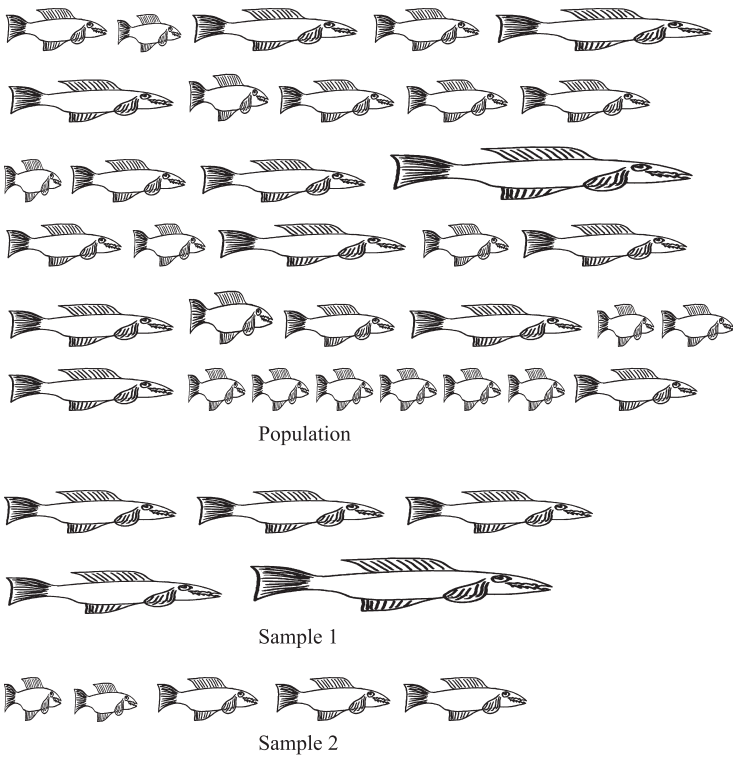

Figure 1.1 Even a random sample may not necessarily be a good representative of the population from which it has been taken. Two samples, each of five individuals, have been taken at random from the same population. By chance sample 1 contains a group of relatively large fish, while those in sample 2 are relatively small.

can even apply to species made up of similar looking individuals (like flies or cockroaches or snails) and causes problems when you work with samples.

First, even a random sample may not be a good representative of the population from which it has been taken (Figure 1.1). For example, you may choose students for an exercise experiment who are, by chance, far less (or far more) physically fit than the student population of the college they represent. A batch of seed chosen at random may not represent the variability present in all seed of that species, and a sample of mosquitoes from a particular place may have very different insecticide resistance than the same species occurring elsewhere.

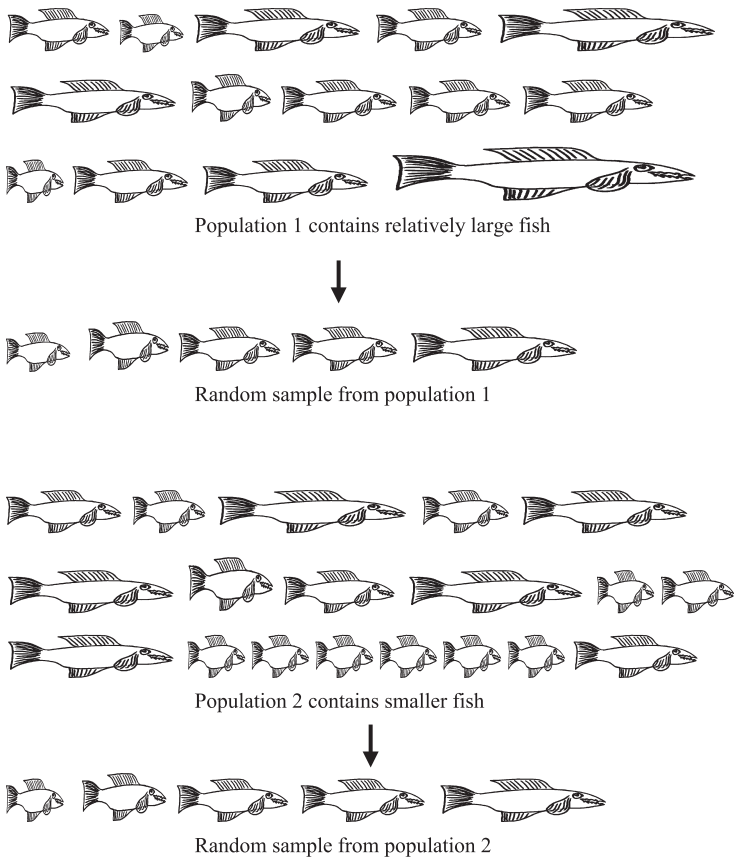

Figure 1.2 Samples selected at random from very different populations may not necessarily be different. Simply by chance the samples from populations 1 and 2 are similar, so you might mistakenly conclude the two populations are also similar.

Therefore, if you take a random sample from each of two similar populations, the samples may be different to each other **simply by chance**. On the basis of your samples, you might mistakenly conclude that the two populations are very different. You need some way of knowing if a difference between samples is one you would **expect by chance** or whether the populations they have been taken from really do seem to be different.

Second, even if two populations are very different, randomly chosen samples from each may be similar and give the misleading impression the populations are also similar (Figure 1.2).

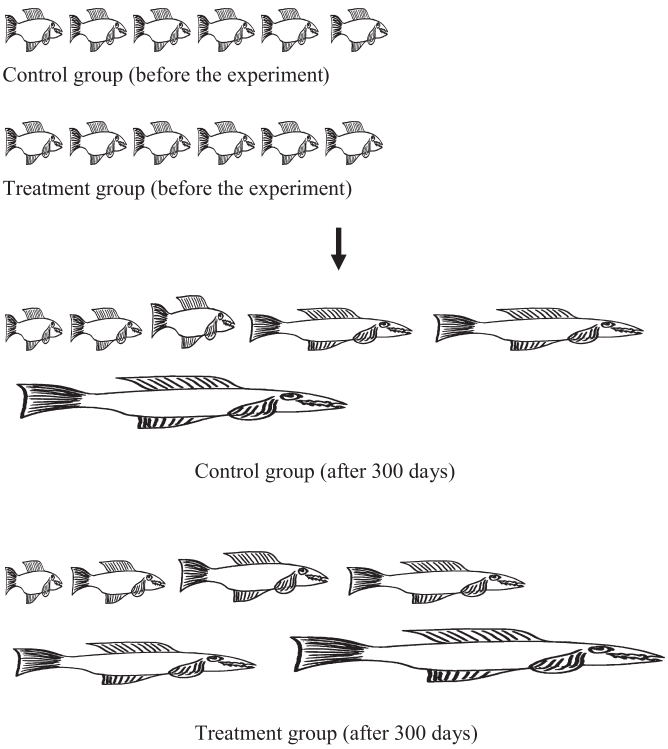

Figure 1.3 Two samples were taken from the same population and deliberately matched so that six equal-sized individuals were initially present in each group. Those in the treatment group were fed a vitamin supplement for 300 days and those in the untreated control group were not. This caused each fish in the treatment group to grow about 10% longer than it would have without the supplement, but this difference is small compared to the variation in growth among individuals, which may obscure any effect of treatment.

Finally, natural variation among individuals within a sample may obscure any effect of an experimental treatment (Figure 1.3). There is often so much variation within a sample (and a population) that an effect of treatment may be difficult or impossible to detect. For example, what would you conclude if you found that a sample of 50 people given a newly synthesised drug showed an average decrease in blood pressure, but when you looked more closely at the group you found that blood pressure remained unchanged for 25, decreased markedly for 15 and increased slightly for the remaining ten? Has the drug really had an effect? What if

tomato plants treated with a new fertiliser yielded from 1.5 kg to 9 kg of fruit per plant compared to 1.5 kg to 7.5 kg per plant in an untreated group? Could you confidently conclude there was a meaningful difference between these two samples?

This uncertainty is usually unavoidable when you work with samples, and means that a researcher has to take every possible precaution to ensure their samples are likely to be representative of the population as a whole. Researchers need to know how to sample. They also need a good understanding of experimental design, because a good design will take natural variation into account and also minimise additional unwanted variability introduced by the experimental procedure itself. They also need to take accurate and precise measurements to minimise other sources of error.

Finally, considering the variability among samples described above, the results of an experiment may not be clear cut. It is therefore often difficult to make a decision about a difference between samples from different populations or from different experimental treatments. **Is it the sort of difference you would expect by chance or are the populations really different? Is the experimental treatment having an effect?**

You need something to **help you decide**, and that is what statistical tests do by calculating the **probability** of a particular difference among samples. Once you know that probability, the decision is up to you. So you need to understand how statistical tests work!

1.2 What is this book designed to do?

A good understanding of experimental design and statistics is important for all life scientists (e.g. entomologists, biochemists, environmental scientists, parasitologists, physiologists, genetic engineers, medical scientists, microbiologists, nursing professionals, taxonomists and human movement scientists), so most life science students are made to take a general introductory statistics course. Many of these courses take a detailed mathematical approach that a lot of life scientists find difficult, irrelevant and uninspiring. This book is an introduction that does not assume a strong mathematical background. Instead, it develops a conceptual understanding of how statistical tests actually work by using pictorial explanations where possible and a minimum of formulae.

If you have read other texts or already done an introductory course, you may find that the way this material is presented is unusual, but I have found that non-statisticians find this approach very easy to understand and sometimes even entertaining. If you have a background in statistics, you may find some sections a little too explanatory, but at the same time they are likely to make sense. This book most certainly will not teach you everything about the subject areas, but it will help you decide what sort of statistical test to use and what the results mean. It will also help you understand and criticise the experimental designs of others. Most importantly, it will help you design and analyse your own experiments, understand more complex experimental designs and move on to more advanced statistical courses.

2 Doing science: hypotheses, experiments and disproof

2.1 Introduction

Before starting on experimental design and statistics, it is important to be familiar with how science is done. This is a summary of a very conventional view of scientific method.

2.2 Basic scientific method

These are the essential features of the 'hypothetico-deductive' view of scientific method (see Popper, 1968).

First, a person observes or samples the natural world and uses all the information available to make an intuitive, logical guess, called an **hypothesis**, about how the system functions. The person has no way of knowing if their hypothesis is correct – it may or may not apply.

Second, a **prediction** is made on the assumption the hypothesis is correct. For example, if your hypothesis were that 'Increased concentrations of carbon dioxide in the atmosphere in the future will increase the growth rate of tomato plants', you could predict that tomato plants will grow faster in an experimental treatment where the carbon dioxide concentration was higher than a second treatment set at the current atmospheric concentration of this gas.

Third, the prediction is tested by taking more samples or doing an experiment.

Fourth, if the results are **consistent with the prediction**, then the hypothesis is **retained**. If they are not, it is **rejected** and a new hypothesis will need to be formulated (Figure 2.1).

The initial hypothesis may come about as a result of observations, sampling and/or reading the scientific literature. Here is an example from ecological entomology.

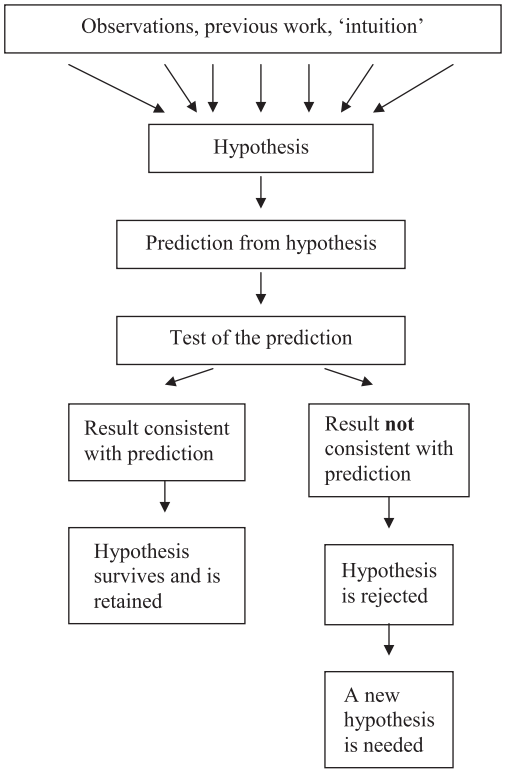

Figure 2.1 The process of hypothesis formulation and testing.

The Portuguese millipede *Ommatioulus moreleti* was accidentally introduced into southern Australia from Portugal in the 1950s. This millipede lives in leaf litter and grows to about four centimetres long. In the absence of natural enemies from its country of origin (especially European hedgehogs which eat a lot of millipedes), its numbers rapidly increased to plague proportions in South Australia. Although it causes very little damage to agricultural crops, *O. moreleti* is a serious 'nuisance' pest because it invades houses. In heavily infested areas of South Australia during the late 1980s, it used to be common to find over 1000 millipedes invading a moderate-sized house in just one night. When you disturb one of these millipedes, it ejects a smelly yellow defensive secretion. Once inside the house, the millipedes would crawl across the floor, up the walls and over the ceiling from where they even fell into food and into the open mouths of sleeping people. When accidentally crushed underfoot, they stained carpets and floors,

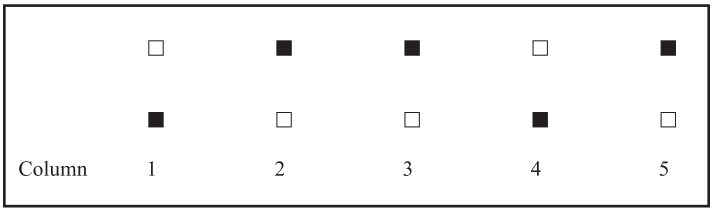

Figure 2.2 Arrangement of a 2 × 5 grid of lit and unlit tiles across a field where millipedes were abundant. Filled squares indicate unlit tiles and open squares indicate lit tiles.

and smelt. The problem was so great that almost half a million dollars (which was a lot of money in the 1980s) was spent researching how to control this pest.

While working on ways to reduce the nuisance caused by the Portuguese millipede, I noticed that householders who reported severe problems had well-lit houses with large and often uncurtained windows. In contrast, nearby neighbours whose houses were not so well lit and who closed their curtains at night reported far fewer millipedes inside. The numbers of *O. moreleti* per square metre were similar in the leaf litter around both types of houses. From these observations and very limited sampling of less than ten houses, I formulated the hypothesis, 'Portuguese millipedes are attracted to visible light at night.' I had no way of knowing whether this very simple hypothesis was the reason for home invasions by millipedes, but it could explain my observations and seemed logical because other arthropods are also attracted to light at night.

From this hypothesis it was straightforward to predict 'At night, in a field where Portuguese millipedes are abundant, more will be present in areas illuminated by visible light than in unlit areas.' This prediction was tested by doing a simple and inexpensive manipulative field experiment with two treatments – lit areas and a control treatment of unlit areas.

Because any difference in millipede numbers between only one lit and one unlit area might occur just by chance or some other unknown factor(s), the two treatments were each **replicated** five times. I set up ten identical white ceramic floor tiles in a two row × five column rectangular grid in a field where millipedes were abundant (Figure 2.2). For each column of two tiles, I tossed a coin to decide which of each pair was going to be lit. The

other tile was left unlit. This ensured that replicates of both the treatment and control were dispersed across the field instead of having all the treatment tiles clustered together and was also a precaution in case the number of millipedes per square metre varied across the field. The coin tossing also eliminated any likelihood that I might subconsciously place the lit tile of each pair in an area where millipedes were more common.

I hammered a thin two-metre long wooden stake vertically into the ground next to each tile. For every one of the lit tiles, I attached a pocket torch to its stake and made sure the light shone on the tile. I started the experiment at dusk by turning on the torches and went back three hours later to count the numbers of millipedes on all tiles.

From this experiment, there were at least four possible outcomes:

(1) No millipedes were present on the unlit tiles, but lots were present on each of the lit tiles. This result is consistent with the hypothesis, which has survived this initial test and can be retained.
(2) High and similar numbers of millipedes were present on both the lit and unlit tiles. This is not consistent with the hypothesis, which can probably be rejected since it seems light has no effect.
(3) No (or very few) millipedes were present on any tiles. It is difficult to know if this has any bearing on the hypothesis – there may be a fault with the experiment (e.g. the tiles were themselves repellent or perhaps too slippery, or millipedes may not have been active that night). The hypothesis is neither rejected nor retained.
(4) More millipedes were present on the unlit tiles than on the lit ones. This is a most unexpected outcome that is not consistent with the hypothesis, which is extremely likely to be rejected.

These are the four simplest outcomes. A more complicated and much more likely one is that you find **some** millipedes on each of the tiles in both treatments, and that is what happened – see McKillup (1988) for more details. This sort of outcome is a problem because you need to decide if light is having an effect on the millipedes or whether the difference in numbers between lit and unlit treatments is simply **happening by chance**. Here statistical testing is extremely useful and necessary because it helps you decide whether a difference between treatments is meaningful.