

6 EXPERIMENTAL DESIGN

Objectives

After studying this chapter you should

- understand what is meant by experimental error, bias, replication and blocking;
- appreciate why experimental and control groups are used;
- understand what is meant by blind and double blind trials.

6.0 Introduction

A greengrocer normally obtained her fruit and vegetables from a market in Manchester. She wished to find out whether obtaining her supplies from a different market in Preston would increase her takings. As an experiment she used the Preston market on eight days and recorded her daily takings. The results together with her takings on ten days when she used the Manchester market are shown below.

	Takings, £
Preston market	323 274 269 552 435 391 208 529
Manchester market	286 517 492 264 367 399 198 581 362 303

The first point to note is that the takings for both markets vary. If every time she bought from Preston her takings were £323 and every time she bought from Manchester her takings were £286 it would be obvious that buying from Preston increased her takings and there would be no need for any statistical analysis.

However the takings do vary and this is said to be caused by **experimental error**. This does not mean that a mistake has been made. It simply means that factors other than the market she is buying from will affect the takings. In this case, it is probable that the weather, the traffic conditions in the area and the types of fruit and vegetables available will affect the takings. Almost certainly the shop will be busier at the weekend than in the middle of the week and this will have a substantial effect on the takings.

What other factors might affect the takings?

The problem with experimental error is that if there is a difference in the average takings from the two markets, it may be difficult or impossible to tell whether this is due to the effect of the factor(s) being investigated - in this case the market - or due to experimental error.

Experimental error should be minimised by keeping factors which are not being investigated as constant as possible and by experimental design. In this case it may be impossible to standardise the weather or the traffic conditions but the opening hours of the shop and the number of assistants working there should be kept constant. In a laboratory experiment temperature, humidity and various other factors could be held constant.

What factors could be kept constant in an experiment to compare the petrol consumption of two different makes of car?

If repeated observations under apparently identical conditions are made then the magnitude of the experimental error may be estimated. The repeated observations are known as **replicates** and the magnitude of the experimental error is usually estimated by calculating the standard deviation of the replicates. In the experiment above there were 8 replicates of the takings from Preston and 10 replicates of the takings from Manchester. As far as the factor under investigation was concerned (i.e. the market), the conditions were identical. The data could be analysed using an unpaired *t*-test (see Section 4.3) to compare the means. This involves making a pooled estimate of the standard deviation, i.e. estimating the size of the experimental error.

6.1 Experimental design

Rather than carrying out the analysis on the data above, it would first be better to use **experimental design** to try to reduce the size of the experimental error. The simplest experimental design is the use of **paired comparisons**. Here two treatments being compared are each applied to similar raw material. For example, if yield of two types of wheat were to be compared, a field might be split into small plots and the two types of wheat planted in adjacent plots. This is to minimise differences in the conditions in which the wheat grows and reduce experimental error due to the two types of wheat growing under different conditions.

Similarly, to compare the weight loss due to two different slimming diets, an ideal design would be to secure the cooperation of several pairs of identical twins. One twin of each pair would follow one diet and the other twin the other diet. Thus experimental error due to physiological differences in the people undertaking the diets would be minimised.

To return to the greengrocer, an obvious source of experimental error is the day of the week. Takings are likely to be much higher at weekends. Therefore if we examine the differences between the takings from each market on the same day of the week, one major source of experimental error will have been eliminated.

The experiment might be carried out as follows

Week 1	Manchester market					
	Mon	Tue	Wed	Thur	Fri	Sat
Takings	272	295	318	307	532	599
Week 2	Preston market					
	Mon	Tue	Wed	Thur	Fri	Sat
Takings	268	272	324	352	511	604

The data can now be analysed using a paired t -test (see Section 4.4). This would be perfectly satisfactory for analysing data of weight loss by identical twins who had followed different diets. However in this case, the design can be improved further. Suppose that Week 1 was fine and dry but Week 2 was wet and windy. Alternatively, suppose new potatoes were available from the market in Week 2 but not in Week 1. If a difference was found between the takings in the two weeks we would not know whether this was due to a difference between the markets or whether it was due to the different weather conditions (or the availability of new potatoes). The two effects are said to be confounded.

A better arrangement would be as follows

Week 1	Mon	Tue	Wed	Thur	Fri	Sat
Market	A	B	A	B	A	B
Takings £	284	296	333	376	494	517
Week 2	Mon	Tue	Wed	Thur	Fri	Sat
Market	B	A	B	A	B	A
Takings £	276	308	307	400	482	512

Why is this a better arrangement?

A and B represent the two markets. The final decision to be taken is whether A represents Manchester and B represents Preston or the other way round. Where there is no obvious reason for choosing one way in preference to the other the choice should be made by a random process such as tossing a coin. This is known as **randomisation** and is to prevent unconscious or unsuspected **bias** from affecting the result. There are no obvious further improvements which can be made to this design. Suppose A was chosen to be Manchester, the analysis would be as follows:

	Mon	Tue	Wed	Thur	Fri	Sat
Manchester	284	308	333	400	494	512
Preston	276	296	307	376	482	517
Difference	12	12	26	24	12	-5

H_0 : mean difference = 0 H_1 : mean difference \neq 0

This can be tested using the paired t -test. This assumes that the differences are normally distributed. The sample mean will be approximately normally distributed even if the individual differences do not exactly follow a normal distribution. It is therefore safe to apply this test unless it is quite clear that the distribution is extremely skew.

The mean difference $\bar{d} = 13.5$ and the standard deviation of the differences $\hat{\sigma} = 11.095$. So

$$t = \frac{13.5}{\left(\frac{11.095}{\sqrt{6}}\right)} = 2.98$$

Critical values of t_5 for a 5% 2-sided risk of **Type 1 error** are ± 2.571 .

H_0 is rejected and, since the mean difference for Manchester - Preston is positive, we conclude that the takings are higher when the Manchester market is used.

It is impossible to rule out completely the possibility that the difference is due to some factor other than the markets. Chance is always a possibility. However, a well-designed experiment eliminates all likely alternative explanations.

Activity 1

A road haulage firm moves lorryloads of slate from a quarry to a depot 70 miles away. You are asked to advise the firm how to investigate which of two alternative routes is quicker. Design an experiment to compare the routes efficiently. Explain the role of replication and randomisation in your design.

Activity 2

You are asked to compare the effect of two different fertiliser treatments on the yield of a particular variety of carrots. Describe how you might design this experiment and the precautions you might take to ensure any conclusions you came to were valid. Explain the role of replication and randomisation in your design.

Blocking

If the greengrocer wished to compare the effect on takings of several possible markets she could visit each market on randomly chosen days and record her takings.

	Takings £					
Kirkham	344	479	503	290	207	
Manchester	459	234	602	222	598	479
Oldham	322	600	308	344	506	
Preston	292	588	347	399	544	406

This is known as a **completely randomised design** and would be analysed using **one factor analysis of variance** (see Section 7.3).

However, the ideas of the paired comparison can be extended to this case. The experiment could be carried out as follows.

	Takings £					
	Mon	Tue	Wed	Thur	Fri	Sat
Kirkham	294	306	343	386	494	527
Manchester	277	318	399	360	524	566
Oldham	299	265	302	410	488	530
Preston	260	289	299	391	460	488

This is known as a **randomised block design**. The markets are blocked by days of the week. This design is analysed using **two factor analysis of variance** (see Section 7.4). One of the factors - in this case the markets - is the object of the investigation. The other factor - in this case the day of the week - is introduced because it is thought that it might have a substantial effect on the results. As with the paired comparison, the introduction of the days of the week as a factor will, if the design is successful, reduce the experimental error. It will therefore make it easier to detect a difference between markets, if such a difference exists.

Control groups

In the examples above two or more different markets were compared. Sometimes there are not two or more treatments, but only one. For example, we may wish to observe the effect of a particular medical treatment on arthritis or the effect of a coaching course on a student's tennis playing skills. The effect of these treatments cannot be judged in isolation. An arthritis sufferer may improve (or deteriorate) with no treatment. Similarly, a tennis player may improve without attending a coaching course.

It is necessary to have a **control group** and an **experimental group**. These two groups should be matched as closely as possible. That is, the people in one group should be as similar as possible to the people in the other group as far as characteristics relevant to the investigation are concerned. This does not mean that all the people in a particular group must be similar to each other, but that the group as a whole must be similar to the other group.

For example, in the case of the arthritis sufferers the two groups should contain people of similar age, sex, general health and severity of arthritis. In the case of the tennis players, the groups should contain students of similar age, sex, fitness and tennis playing ability.

The groups should be selected and then one group should be chosen at random to receive the experimental treatment (or tennis coaching) and the other group will be the control group. The control group will receive no treatment (or coaching) or will continue with the standard treatment. The effect on the two groups can then be compared.

Blind and double blind trials

In the case of medical treatment, it is sometimes thought that patients will improve or recover without treatment and that in some cases this improvement will be greater or quicker if they are told they are having treatment, even if they are not. Thus it is standard practise in drugs tests to give the control group a **placebo**. This is a harmless substance which looks like the real medication but, in fact, does not contain any drug. Many patients will improve after taking placebos. To show a drug to be effective, significantly more patients who took the drug must show improvement than those who took the placebo. (There are, of course, other issues such as possible side effects to consider as well).

If the patients who took the placebos knew that they were taking placebos the effect would of course be lost. It is essential that the patients should not know whether or not they are taking placebos and this is known as a **blind trial**.

Even more subtle effects can be at work. It has been found that, even if the patients do not know whether or not they are taking placebos, the doctor may expect those patients taking the drug to fare better than those taking placebos. This expectation may somehow transmit itself to the patient whose condition may improve as a result. It is therefore, necessary that the doctor does not know which patients are receiving placebos and which are receiving the drugs. Of course, someone must know who is receiving the drugs otherwise it would be impossible to analyse the results. However, it should be someone who has no direct contact with the patient. Trials where neither the patient nor the doctor know who is receiving the drugs are known as **double blind trials**.

It has been suggested that the person carrying out the statistical analysis should also not know which patients took the drug to prevent this influencing the analysis. This would be described as a triple blind trial.

Drug trials are greatly affected by the ethical problems involved. Firstly, it is of course essential that patients taking part should be fully informed of the nature and possible risks of the experiment. Secondly, once a treatment has been established as beneficial, it would be wrong not to let all the patients taking part in the trial benefit from it even if this interferes with strictly statistical considerations. Thirdly, it is clearly wrong to carry out an experiment, with all the inconvenience it may cause and false hopes it may raise, if it has not been well designed. Trials have been carried out which have been too small to establish the effectiveness of a treatment, whatever the results obtained. In other cases, lack of effective design has made the experimental error much larger than it need be thus making a real effect impossible to detect. The statistician has an important role to play in this fascinating area.

