8 Interrelationship of design and analysis

In Chapter 2 we went through an example which illustrated the steps which have to be taken in designing an experiment. To recap, these were as follows:

Given that you have a problem, a relatively fuzzy idea of the area in which you want to carry out the experiment, the first step is to develop one or more precise **research questions**. These have then to be turned into a form which is capable of being tested experimentally. This means deciding on the **independent variable** and **dependent variable**, and, conventionally, setting up a **null hypothesis** and **alternative hypothesis** in terms of the specific IV and DV. Each of the variables has to be **operationalized**, i.e. you have to state what precisely you do in order to manipulate or measure it. You have also to decide how many levels of the independent variable you are going to use (i.e. how many experimental conditions or treatments). Remember that the techniques presented in this book deal directly only with the comparison of two treatments.

The next decision concerns how participants fit into the experiment. This gives us our three basic experimental designs – **independent samples**, **matched pairs** and **repeated measures**. Then you must decide on how many participants, which will obviously be influenced by many things: how difficult they are to get, how long you can spend with each, and so on. The kind of problem that you start with, coupled with the experimental design, and the type of data that this generates (scores, frequencies or counts, or ranks) effectively decides for you the statistical test (or tests – often more than one would be feasible) which you use. Relying solely on the techniques that are described in this book, there is a good range of experimental designs whose results you can analyse statistically. Of course, there are many designs for which you have not been given the appropriate statistical techniques.

A basic principle is that you should never conduct an experiment without having thought through the ways in which it can be analysed. *The decisions about the statistics to be used must be made as a part of the design process.* If you don't do this you run the grave risk of having data that is unanalysable.

This is always true, no matter how sophisticated an arsenal of statistical techniques is at your disposal. It is only in cases where experimenters have supreme confidence that they can demonstrate the effects of the independent variable completely unequivocally (as happens for instance in some cases with the application of Skinnerian techniques) that statistical analysis may be unnecessary.

How to increase the sensitivity of an experiment

Many new experimenters get very discouraged by a string of non-significant results. Somebody new to the game is likely to be unskilled in selecting problems amenable to attack by experimentation. An experiment is a precision tool where we are effectively making what might be a very risky bet that a particular independent variable, operationalized in a specific way, has an effect on a particular dependent variable, also operationalized in a specific way. Confidence that you are doing something sensible comes from your building on the work of others (which generally means that you have a good knowledge of the research literature in a particular field), or where you have built up your own knowledge and experience through working in the area.

However, there are a number of general areas to which you can give attention which will help to make your experiments more sensitive at detecting experimental effects.

1 Reduce the 'noise' level

By this, I do not mean just the physical sound level. 'Noise' is used figuratively here as a general term to cover the effects of uncontrolled variables. These effects appear in many ways, for instance in the instructions given to participants. If instructions vary slightly from participant to participant, then the scores obtained by participants might vary according to these instructions. If there is poor experimental control over the general conditions in the laboratory, or wherever the experiment takes place (e.g. people talking or laughing, to which the participants may sometimes pay attention and sometimes not), if they are bored with the experiment or more interested in interacting with the experimenter than with the experiment – these can all have effects on the data which have nothing to do with the experimental variable being manipulated.

Standardization and control can be taken too far. There is a good case for our being concerned with experimental effects which are sufficiently strong to show themselves in relatively naturalistic situations. However, if we have decided to follow the experimental approach then the line must be drawn at a point where the 'noise' is at a level where the experimental effects have a chance to come through.

So lesson one is that the sensitivity of an experiment can be increased by increasing the degree of experimental control over the conditions under which the experiment takes place. Given that we have refined our procedures as much as possible, what else can be done?

2 More participants

A second possibility is to increase the *sample size* – that is, the number of participants taking part in the experiment. This tends to make the experiment more sensitive because the effect of the experimental variable (assuming that there is an effect) will add together over participants, whereas the random error effects (which we will never be able to get rid of completely) will tend to cancel each other out as some will be in one direction, some in the other.

Another way of saying the same thing is that we are more likely to get a statistically significant result if we increase the sample size. In fact, if statistical significance were all that one was interested in you could virtually guarantee it in *any* experiment by choosing a sufficiently large sample size! This is because there will almost always be some (even though very minor) effect of the IV on the DV. So, somewhat paradoxically, there is a case for paying more attention to statistically significant results obtained from relatively small samples than from those with large samples. The former are more likely to be 'significant' in the sense of 'important' or 'strong' as they have, as it were, emerged successfully from a considerable amount of 'noise' from random effects.

However, in the small experiments that you are likely to be carrying out as novice experimenters, a good rule is to work with as many participants as you can get hold of and deal with properly in the time available.

3 Floor and ceiling effects

Avoid possible 'floor' and 'ceiling' effects on the measures that you are taking. The level of difficulty of the experimental situation should be adjusted so that scores lie in the mid-range of any scale that is used. If, in a memory experiment, all the participants score between 90 and 100 per cent correct, with a lot of 100s, then any difference between experimental conditions would be reduced simply because the results of some participants are bumping on the ceiling and hence not going as high as they would otherwise. The converse occurs if the material is too difficult and floor effects result.

The solution lies in careful pilot work to establish the kind of scores participants are likely to obtain *under the precise conditions of the experiment.*

4 Increasing reliability of the measure

Another way in which the sensitivity can be increased is by *increasing the reliability of the measures to be analysed.* One way in which this can be done in psychological experimentation is by basing the measure not on a single observation but on a series of observations and then using the mean or some other measure of central tendency in analysis. Any random effect unconnected with the experimental effect, say due to a loud noise just before an observation is made, might influence a single observation a great deal, but would have much less effect on the median score. Consider carefully, of course, whether the experiment is such that taking a series of observations from a single participant is possible. In doing that, we are assuming that the observations are independent of each other, that they are all measures of the same thing. The reasonableness of this assumption varies very much from one experimental situation to another.

5 Which design?

Finally, one should consider the relative sensitivity of the independent samples, matched pairs and repeated measures designs. Generally, the sensitivity increases as one goes from independent samples to matched pairs to repeated measures. This is due to the increasing degree of control over any variables associated with participants. In a repeated measures design it is the same participant who appears under both conditions. In so far as scores under the experimental conditions may be affected by such things as age, sex, intelligence, personality characteristics, etc., we obviously have perfect matching, and hence direct control over these variables, when it is the same participant under both conditions.

With the matched pairs design we retain some matching, but this is usually done on just a single variable. Hence the degree of matching is less than with repeated measures and the efficiency of the design will depend on how close a correlation exists between the matching variable and the dependent variable. If there is a high correlation, then the matching will be very effective. Your problem is to find variables with this high degree of correlation. It is not easy.

With independent samples designs there is no attempt at all to match participants on a one-to-one basis, and therefore no participant variables are controlled, and this design is the least sensitive.

This analysis should not be taken as an indication that we always aim for repeated measures designs and avoid independent samples designs. The great weakness of repeated measures designs lies simply in the fact that they have repeated measures! Because participants have to perform under both experimental conditions, there are all kinds of nasty effects which might occur.

One of the special features of humans is the extent to which they are learning animals: the extent to which their present behaviour is modified by their past experiences. If we test the same person under two conditions there is likely to be an order effect. The result of the second test may well be modified by their experience on the first one. Counterbalancing or randomization of the order of presentation of the two conditions can and should be used, but it will only completely neutralize a simple order effect which adds (or subtracts) a constant amount to, or from, whatever is done second. There is no guarantee that the effects will be as simple as this.

Because of this the repeated measures design is best used in situations where the order effect is known to be small or negligible – for example in simple motor tasks where no knowledge of results is provided to participants. Alternatively, if the random variability between participants on a dependent variable was very high (so that an independent samples design would be unlikely to yield any results), one might be tempted to use a repeated measures design.

It is possible to convert an independent samples design into a matched pairs design, providing that some meaningful way of matching can be devised (i.e. a matching variable which is known to correlate reasonably highly with performance on the dependent variable). The only disadvantage is the labour involved in getting the scores on the matching variable to make up the pairs of participants.

The strategy suggested then is to use a matched pairs design if there is a matching variable which correlates highly with the dependent variable. If this is not available, then a choice between independent samples and repeated measures designs depends upon the likelihood that repeated measures would be independent. If this appears unlikely, the independent samples design should be used.

More complex designs

Some indications of the kinds of ways in which the basic experimental designs can be complicated will now be considered.

More than two experimental conditions

Whilst it is possible to use our two-condition experimental design in order to test for the effect of the independent variable, there are several defects to this simple design. It may happen, for instance, that the two values of the independent variable which we have chosen happen not to show any effects, whereas the choice of two other values might have shown an effect. The obvious way of getting over this difficulty is to include a larger number of experimental conditions and to look for differences between these conditions.

You can use the t-test (or its non-parametric equivalents where appropriate) to look at conditions in pairs, but there are difficulties connected with significance level if this is done. Do not forget that a 5 per cent significance level means that there is a 5 per cent chance of mistakenly deciding that the IV is affecting the DV when only random effects are present. This means that if you made 20 comparisons between pairs of conditions, you would expect 1 of these 20 (i.e. 5 out of 100, or 5 per cent) to come up with a significant effect *even when there is no actual effect of IV on DV.* Similarly, if you just look at the results after they have been obtained, and pick out, say, the conditions with lowest and highest means, you are effectively going through all the other tests implicitly and the difficulties with significance level remain even though one might only compute a single t-test.

A second criticism of experiments with only two experimental conditions is that, whereas they can indicate whether or not an independent variable has an effect on a dependent variable (if we are lucky or cunning with the particular levels of the IV that we have chosen), they cannot tell us anything of the nature or the relationship between the two variables.

Figure 24 shows three different possibilities which would fit in

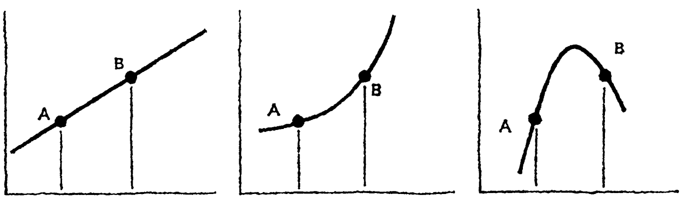

Figure 24 Three different relationships between independent and dependent variables consistent with known values at A and B

with the same results on conditions A and B. The only way in which the nature of these relationships can be made clear is by including in the same experiment several values on the independent variable, so that more points can be filled in on the graph.

A statistical technique which is useful for designs with several levels of the independent variable is the **analysis of variance**, which is covered in detail in most advanced texts of psychological statistics.

More than one independent variable

There is no reason why an experimental design need be limited to a single independent variable. The design can be extended to include as many variables as you wish but there are considerable advantages over single independent variable experiments if just two IVs are included.

One such design involves all possible combinations of levels of the different IVs, and is known as a **factorial design**. It can tell us about the effect of a particular IV, not just when all other variables are held constant (as in the single variable design), but over the different levels of the other IV.

A great advantage of factorial designs is that they bring out possible **interactions** between variables. An interaction occurs when the effect of one independent variable is not constant, but varies according to the level of another independent variable.

Suppose, for example, that a number of children were assessed on their degree of initiative on the one hand, and the extent of parental encouragement on the other. Four groups of children were formed: low parental encouragement with low initiative, low parental encouragement with high initiative, high parental encouragement with low initiative, and high parental encouragement with high initiative. Subsequent intelligence tests might have given the results shown graphically in Figure 25.

This shows an interaction between the two variables – parental encouragement and initiative – in the sense that, for children with low parental encouragement it made little or no difference whether they had low or high initiative. For children with high parental encouragement, those with high initiative scored considerably higher than those with low initiative.

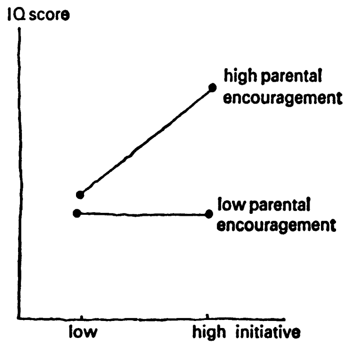

Figure 25 Interaction between the variables of parental encouragement and initiative

These (fictitious) results can also be used to repeat earlier warnings about the interpretation of the findings in a study. It is highly unlikely that a study of this type was an experiment involving random assignment of participants to the different parental encouragement and initiative conditions. So, any inferences about causative relationships are very difficult to make because of the possible existence of other factors.

Designs involving more than one independent variable cannot be analysed directly by the techniques covered in this text. The analysis of variance, referred to in the previous section, would commonly be used. It would be possible to perform separate analyses of the effect of parental encouragement on low-initiative children (e.g. a t-test in each case). Alternatively, or additionally, one could perform a similar test on the effect of initiative on the group with low parental encouragement and then the effect of initiative on the group with high parental encouragement.

Such tests can provide some kind of analysis of the data, but do not tell us anything directly about possible interactions, and they are really not an adequate substitute for a full analysis.

More than one dependent variable

Just as it is possible to make use of more than one independent variable, so you can have more than one dependent variable. There might be advantages in looking at the effect of a particular independent variable on several dependent variables. If, for example, we are investigating the effects of sleep deprivation, it might appear sensible to use a battery of different testing situations, some cognitive, some perceptual, and so on.

If it is necessary to find the effects of more than one dependent variable simultaneously, then so-called 'multivariate' procedures should be used. Most of these procedures are extremely complicated and tedious to compute and it would be foolish to adopt them without appropriate computer software which can do the drudgery for you. However, designs using a single dependent variable can be applied appropriately to many research problems. In practice it is often impossible to measure more than two or three dependent variables. These can be used one by one in separate analyses, although, as with multiple independent variables, important aspects might be lost by doing this.