

improve the comparability of the two groups by presenting the controls with a **placebo**, that is, a neutral medium ideally identical with that in which the drug was presented to the experimental group.)

The IV is controlled by the investigator, and its values are determined **before the experiment is carried out**. This is achieved either by **random assignment** of the participants to the pre-set conditions or by deciding to test each participant under all conditions, if that is feasible. The DV, on the other hand, is **measured during the course of the investigation**.

In a true experiment, the IV, unlike gender, blood group, or nationality, is not an intrinsic property of the participants: the participants are assigned at random to the experimental or the control group. In other circumstances, however, the hypothesis might be that some aspect of a person's gender, nationality, religion or age group affects, say, their views on some issue. For example, do men and women have different views about violence on television? This hypothesis might be tested by interviewing some men and women and comparing their views. Here, the IV is gender, and the DV is attitude to violence. In this case, however, the participants have not been assigned to the categories making up the IV: the latter is an intrinsic property of the participant. Such **quasi-experiments** often produce results that are difficult to interpret unequivocally. The problem with quasi-experimental research is that the IV, being a **participant (or subject) variable**, is itself a whole package of other variables, some of which may not be under the control of the investigator. Some of the difficulties with quasi-experimental research can be overcome by following an appropriate sampling strategy. For example, in the gender-and-attitudes experiment, the men and women would be sampled from comparable social strata and similar age bands. But there is always the possibility that the effects of gender will be entangled, or **confounded**, with those of uncontrolled, **extraneous** variables.

#### *Experimental versus correlational research*

In true experiments and quasi-experiments, the IV is manipulated by the researcher, either directly, as in the former, or by sampling, as in the latter. In **correlational research**, the variables are recorded as they occur in the participants. For example, it might be hypothesised that cheerfulness (the DV) is promoted by exercise (the IV). One might test this hypothesis by asking participants in a study to rate themselves on cheerfulness and to say how often they take exercise. Here the problems of interpretation are multiplied. It may well turn out that the cheeriest people take the most exercise, the glummiest take the least and moderately cheerful people lie somewhere in between. But both the putative IV (amount of exercise) and the putative DV (cheerfulness) may arise from a myriad other possible variables, including socioeconomic status, personality and physical type.

#### *Implications for data analysis*

Typically, the data from experiments and quasi-experiments are processed using statistical methods that are rather different from those used with data from correlational research. In the next section, we offer the reader some advice (together with some cautions and caveats) on choosing statistical tests.

## 1.2 CHOOSING A STATISTICAL TEST: SOME GUIDELINES

It is common for authors of statistical texts to offer advice on choosing statistical tests in the form of a flow chart, decision tree or similar diagram. The numerous schemes that have been proposed vary considerably, and sometimes contradict one another. The reason for the inconsistency is partly that most classifications tend to break down in certain circumstances. Moreover, in some situations, the correct choice has been hotly disputed.

On one matter at least, however, there is general agreement. **There is no such thing as a decision tree that will automatically lead the investigator to the correct choice of a statistical test in all circumstances.** Some of the later chapters contain illustrations of the penalties that a scheme-reliant approach can incur. At best, a decision tree can serve only as a rough guideline. Ultimately, a safe decision requires careful reflection upon one's own research aims and a **thorough preliminary exploration of the data**.

#### *Considerations in choosing a statistical test*

The choice of a statistical test depends upon:

1. The **research question**
2. The **plan**, or **design**, of the research
3. The **type of data** that you wish to analyse

This list is by no means comprehensive; nor do we intend to imply that any fixed ordering of these three considerations is appropriate in all situations. Moreover, a decision about one often has implications for the others.

#### *The research question*

We shall identify five basic research situations in which formal statistical tests can be applied (Figure 1). For each situation, the appropriate techniques will be discussed in the sections indicated in the figure.

The questions are as follows:

1. Is a difference (between averages) significant? For example, is resting heart rate the same before and after a fitness course? (Sections 1.3 and 1.4)
2. How strongly are variables associated? For example, do tall parents tend to have tall children? (Section 1.5)
3. Can scores on a target variable (or category membership, if the variable is qualitative) be predicted from data on other variables? For example, can university performance be predicted by scores on aptitude tests? (Section 1.6)
4. From a single sample of data, what can be said about the population? For example, if we know the vocabulary test scores of 100 children, what can we infer about the scores of the entire population of children in the same age group? (Section 1.7)
5. The user has a multivariate data set, perhaps people's scores on a battery of ability tests. Can these scores be accounted for (or classified) in terms

of a smaller number of hypothetical latent variables or **factors**? For example, can performance in a variety of intellectual pursuits be accounted for in terms of general intelligence? (Section 1.8)

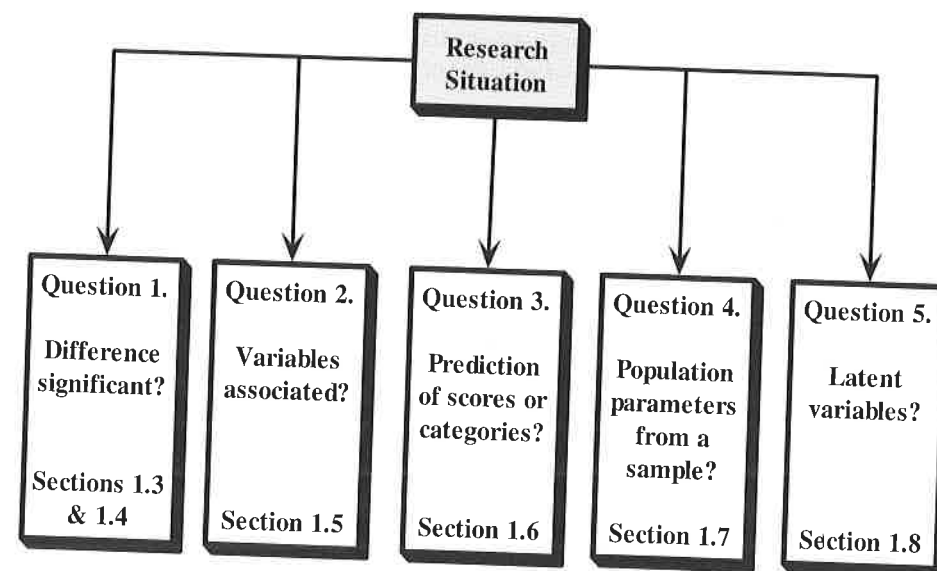


Figure 1. Five types of research situation

### 1.3 SIGNIFICANCE OF DIFFERENCES

The question of whether two means are significantly different is one that arises naturally in the context of experimental or quasi-experimental research, where the performance of the participants under different conditions is being compared.

Suppose that in a drug experiment, performance under two different conditions (experimental and control) has been measured and that the means have somewhat different values. This may seem to support the experimenter's hypothesis; but would a similar difference be found if the experiment were to be repeated? Here the researcher wishes to test the **statistical significance** of the difference, that is, to establish that the difference is too large to have been merely a chance occurrence.

#### 1.3.1 The design of the experiment: Independent versus related samples

Of crucial importance in the choice of an appropriate statistical test is the question of whether the experiment would have resulted in **independent samples** or **related samples** of scores.

##### *Independent samples*

Suppose we select, say, 100 participants for an experiment and randomly assign half of them to an experimental condition and the rest to a control condition. With this procedure, the assignment of one person to a particular group has no effect upon the group to which another is

assigned. The two **independent samples** of participants thus selected will produce two independent samples of scores. A useful criterion for deciding whether you have independent samples of data is that there must be **no basis for pairing the scores in one sample with those in the other**. An experiment in which independent samples of participants are tested under different conditions is known as a **between subjects experiment**.

##### *Related samples*

Suppose that each of fifty participants shoots ten times at a triangular target and ten times at a square target of the same area. For each target, each participant will have a score ranging from 0 (ten misses) to 10 (ten hits). As in the previous example, there will be two samples of 50 scores. This time, however, each score in either sample can be paired with the same participant's score with the other target. We have two **related samples** of scores, or a set of **paired data**. The scores in two related samples are likely to be substantially correlated, because the better shots will tend to have higher scores with either target than will the poorer shots. An experiment like this, in which each participant is tested under both (or all) conditions, is known as a **within subjects experiment**. Within subjects experiments are also said to have **repeated measures** on the IV (the shape of the target).

There are other ways of obtaining paired data. Suppose that in the current example, the participants were pairs of identical or fraternal twins: each participant shoots at only one target and their twin shoots at the other. This experiment will also result in two related samples of scores, because, as in the repeated measures experiment, there is a basis for pairing the data. Different statistical tests are appropriate for use with independent and related samples of data.

#### 1.3.2 Flow chart for selecting a suitable test for differences between averages

Figure 2 outlines **some** of the considerations leading to a choice of a statistical test of the significance of differences between means (or frequencies, if one has nominal data). If there are more than two conditions or groups, an analysis of variance (ANOVA) may be applicable (see Section 1.4). In this section, we shall consider only the comparison between two groups or conditions, such as male versus female, or experimental group versus control group.

To use the chart, begin at the START box and consider how many conditions there are in the experiment. If there are two conditions, proceed down the chart to the next stage. The next questions are whether the samples are independent or related and whether the data are measurements, as opposed to nominal assignments. The appropriate test is then shown in the bottom box. If there are more than two conditions, an ANOVA should be considered and the flow chart in Figure 3 (see page 13) can be used to ascertain which ANOVA test might be appropriate.

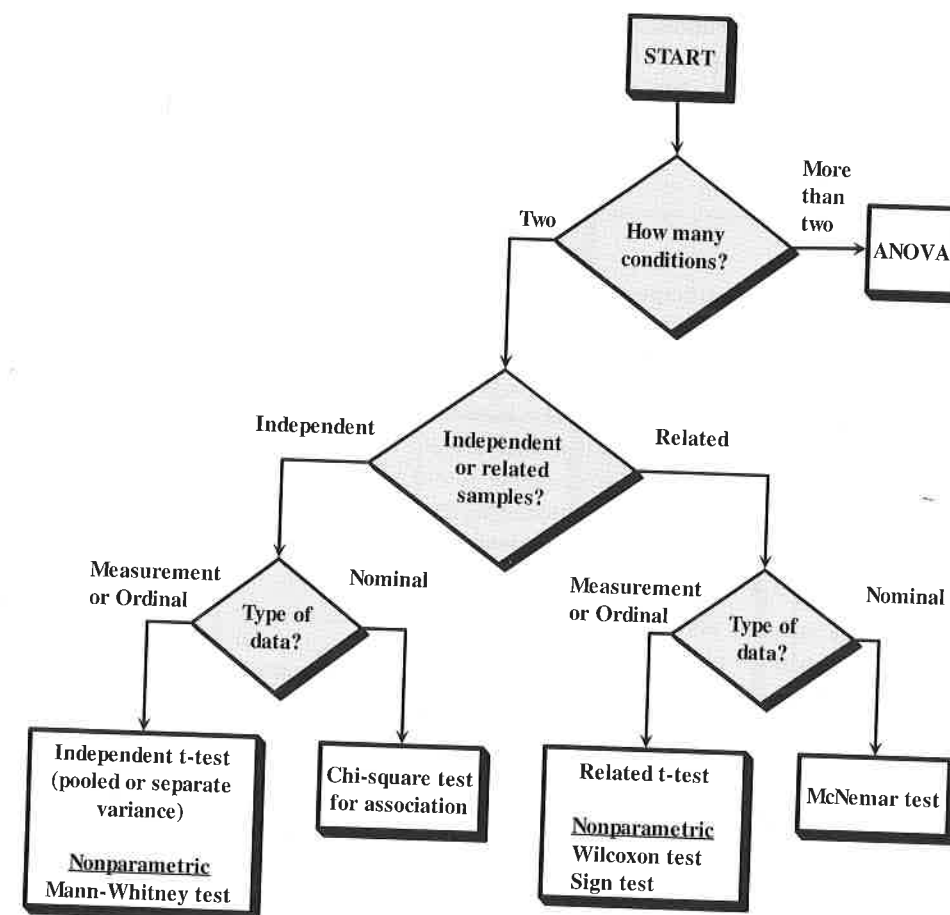


Figure 2. Flow chart showing the selection of a suitable test for differences between averages

### 1.3.3 Two conditions: The *t* tests

To compare the levels, or averages, of two *independent samples* of data in the form of measurements, the **independent-samples *t* test** is often appropriate (see Figure 2). This test and other *t* tests are fully described in Chapter 6.

In its original form, the independent-samples *t* test uses an average or **pooled** estimate of the supposedly constant population variance. For this reason, it is sometimes known as the **pooled *t* test**. Sometimes, however, the data may not conform to the requirements of the independent-samples *t* test. If the sample variances and/or sample sizes are markedly different, the **separate-variance *t* test** may be a safer choice or possibly resort to a nonparametric test as described in the next Section.

When one wishes to compare the means of two **related samples** of data, as when each participant in an experiment has been tested under both an experimental and a control condition, the **related samples *t* test** should be considered, or possibly (depending on the data distribution) a nonparametric test as described in the next Section.

See  
Chap.  
6

### 1.3.4 Two conditions: Nonparametric tests

The *t* test is an example of a **parametric test**: that is, it is assumed that the data are samples from a population with a specified (in this case normal) distribution. Other tests, known as **nonparametric tests**, do not make specific assumptions about population distributions and are therefore also referred to as **distribution-free tests**.

There are circumstances in which a *t* test can give misleading results. This is especially likely to occur when the data set is small and there are some highly deviant scores, or **outliers**, which can inflate the values of the denominators of the *t* statistics.

Figure 2 identifies the nonparametric equivalents of the independent and related samples *t* tests. A nonparametric alternative to the independent-samples *t* test is the **Mann-Whitney U test**. Two nonparametric equivalents of the related-samples *t* test are the **Wilcoxon test** and the **Sign test**.

See  
Chap.  
6

There has been much controversy about the use of nonparametric tests instead of *t* tests with some kinds of data. While some authors (e.g. Siegel & Castellan, 1988) strongly recommend the use of nonparametric tests, others, such as Howell (2002) emphasise the robustness of the parametric *t* tests to violations of their assumptions and the loss of power incurred by the use of the equivalent nonparametric tests. We suggest that, provided the data show no obvious contraindications, such as the presence of outliers, marked skewness or great disparity of variances (especially if the last is coupled with a large difference in sample size), a *t* test should generally be used. There are also circumstances in which the removal of outliers from the data set followed by another *t* test is a defensible procedure (e.g. very long reaction times that could be due to participants not being ready). Otherwise, a nonparametric equivalent should be considered. Data in the form of ratings are a grey area, and there has been considerable debate over whether they should be analysed with parametric or nonparametric tests. If, however, the data are measurements at the ordinal level in the first place, as with sets of ranks, or nominal data, a nonparametric test is the only possibility.

In a planned experiment, the data usually take the form of measurements on an independent scale with units. Occasionally, however, one might have a situation in which each participant attempts a task and either a pass or a fail is recorded. If so, a two-group experiment will yield two independent samples of **nominal data**. Here the research question is still one of the significance of differences, albeit of differences between frequencies, rather than differences between means. With independent samples, a **chi-square test of association** will answer the question of whether the success rates in the two groups are significantly different.

See  
Chap.  
11

#### *Two correlated samples of dichotomous nominal data: The McNemar test*

Suppose that, before they hear a debate on an issue, ten people are asked whether they are for or against the proposal. Afterwards, the same people are asked the same question again. This is a **within subjects experiment**, in which each participant is observed under two conditions: *Before* (an event) and *After*.

On each occasion of testing, a person's response is coded either as 0 (*Against*) or 1 (*For*).

See  
Chap.  
6

Since the same person produces two responses, we have a paired nominal data set, for which the **McNemar test** is appropriate if one wishes to claim that the debate has resulted in a change of opinion.

#### 1.4 ANALYSIS OF VARIANCE EXPERIMENTS

The **analysis of variance (ANOVA)** is actually a whole set of techniques, each of which is based upon a model of how the data were generated and culminates in tests that are appropriate for that particular model only. It is therefore important to identify ANOVA experiments correctly, in order to choose the right tests. In this section, only some of the most common ANOVA experiments will be described. There are many others, which can be found in standard statistics textbooks such as Winer, Brown & Michels (1991).

In the ANOVA, a **factor** is a set of related conditions or categories. The conditions or categories making up a factor are known as its **levels**, even though, as in the qualitative factors of gender or blood group, there is no sense in which one category can be said to be 'higher' or 'lower' than the other. The terms **factor** and **level** are the equivalents, in the context of ANOVA, of the terms **independent variable** and **value**, respectively. A factor can be either a true independent variable or a participant characteristic, such as gender, that we 'manipulate' **statistically**, by sampling people from each of some set of categories. In ANOVA, participant variables are treated in exactly the same way as IVs that are directly manipulated by the experimenter.

##### 1.4.1 Between subjects and within subjects factors

Some factors are **between subjects**: that is, the participant is tested under only one condition (i.e. at one level) of the factor. Gender is an obvious example. Other factors are **within subjects**, that is, the participant is tested under all the different conditions (levels) making up the factor. An experiment with a within subjects factor is also said to have **repeated measures** on that factor.

Returning to the simplest two-group drug experiment, in which the performance of an experimental (drug) group is compared with that of a control (placebo) group, we have one treatment factor (*Drug Condition*), comprising two conditions or levels: *Drug Present* and *Drug Absent*. Since different samples of participants perform under the different conditions, this is a **between subjects experiment**. Should the investigator wish to study the effectiveness of more than one drug, the performance of two or more groups of participants on different drugs could be compared with that of the controls. In that case, there would still be one treatment factor, but with three or more levels (see Table 1a).

Now suppose that another investigator wishes to test the hypothesis that words presented in the right visual hemifield are recognised more quickly than those presented in the left hemifield. Fifty participants are each presented with twenty words in the left and right hemifields and their median response times are recorded. In this experiment (as in the drug experiment), there is **one treatment factor** (in this example, *Hemifield of Presentation*) comprising two levels (conditions), *Left Field* and *Right Field*. But this time, the same participants perform under both conditions. This experiment is said to be of **within subjects** design, or to have **repeated measures on the Hemifield factor**.

In the general case of a one-factor, within subjects experiment, the factor would comprise three or more treatment conditions, or levels, and each participant would perform under all conditions. For example, to test the hypothesis that the accuracy with which a participant shoots at a target depends upon the shape of the target's perimeter, each participant might be tested with circular, square, triangular and diamond-shaped targets. Here the treatment factor *Shape of Target* would have four levels: *Circle*, *Square*, *Triangle* and *Diamond* (Table 1b).

| Table 1. Between subjects and within subjects experiments in which there is one treatment factor with four levels |                                                    |         |          |         |
|-------------------------------------------------------------------------------------------------------------------|----------------------------------------------------|---------|----------|---------|
| (a) A one-factor between subjects experiment                                                                      |                                                    |         |          |         |
|                                                                                                                   | Levels of the Drug factor                          |         |          |         |
|                                                                                                                   | Control                                            | Drug A  | Drug B   | Drug C  |
| Participants                                                                                                      | Group 1                                            | Group 2 | Group 3  | Group 4 |
| (b) A one-factor within subjects experiment                                                                       |                                                    |         |          |         |
|                                                                                                                   | Levels of the Shape factor                         |         |          |         |
|                                                                                                                   | Circle                                             | Square  | Triangle | Diamond |
| Participants                                                                                                      | The same participants perform with all four shapes |         |          |         |

##### 1.4.2 Factorial designs: Between subjects and within subjects experiments

In addition to the effects of different drugs upon fresh participants, a researcher might also wish to investigate their effects upon those who have voluntarily gone without sleep for twenty-four hours. Accordingly, an experiment is designed as shown in Table 2, which shows two factors: *State (Fresh, Tired)*; *Drug (Drug A, Drug B)*. An experimental design with two or more factors is known as a **factorial** design.

| Table 2. A two-factor factorial between subjects experiment |       |                           |         |
|-------------------------------------------------------------|-------|---------------------------|---------|
| Levels of the State factor                                  |       | Levels of the Drug factor |         |
|                                                             |       | Drug A                    | Drug B  |
| Fresh                                                       | Tired | Group 1                   | Group 2 |
|                                                             |       | Group 3                   | Group 4 |

##### Between subjects factorial experiments

Notice that, in Table 2, different samples of participants perform in the four combinations of the two treatment factors. Since each participant is tested only once, this type of experiment is said to be of **between subjects factorial** design. It can also be described as having **two factors with no repeated measures**. Factorial experiments are discussed in Chapter 8.

See  
Chap.  
8

### Within subjects factorial experiments

Table 3 shows another kind of factorial experiment.

| Table 3. A two-factor factorial within subjects experiment |                                                                               |       |        |       |
|------------------------------------------------------------|-------------------------------------------------------------------------------|-------|--------|-------|
| Levels of the Shape factor:                                | Triangle                                                                      |       | Square |       |
| Levels of the Colour factor:                               | Blue                                                                          | Green | Blue   | Green |
| Participants                                               | The same participants perform under all four combinations of shape and colour |       |        |       |

As before, there are two factors, which in this case are:

1. Shape of Target (Triangle, Square);
2. Colour of Target (Blue, Green).

This time, however, every participant is tested at both levels of each factor. This type of factorial design is known as a **within subjects factorial** experiment. Alternatively, the experiment in Table 3 can be said to have **repeated measures on both factors**. Within subjects experiments are discussed in Chapter 9.

See  
Chap.  
9

### 1.4.3 Mixed or split-plot factorial experiments

Suppose that, in addition to the effects of target colour, the experimenter wishes to test the hypothesis that men and women differ in the colour of target with which they are more accurate (Table 4).

| Table 4. A two-factor mixed factorial experiment with one between subjects factor (Gender) and one within subjects factor (Colour of Target) |                                                      |      |
|----------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------|------|
| Levels of the Gender factor                                                                                                                  | Levels of the Colour of Target factor                |      |
|                                                                                                                                              | Red                                                  | Blue |
| Male                                                                                                                                         | Each participant is tested with red and blue targets |      |
| Female                                                                                                                                       | Each participant is tested with red and blue targets |      |

Once again, there are two factors:

1. Gender (Male, Female);
2. Colour of Target (Red, Blue).

The first factor is between subjects; but the second factor is within subjects, because each participant shoots at both the red and the blue target. Factorial experiments with a mixture of between subjects and within subjects factors are known as **mixed** (or **split-plot**) factorial experiments. These are discussed in Chapter 10.

See  
Chap.  
10

### 1.4.4 Flow chart for ANOVA

The between subjects, within subjects and mixed (or split-plot) factorial ANOVA are based upon different models of the data, and require different tests. The presence of a within subjects (or repeated measures) factor in an experiment results in the gathering of related (rather than independent) samples of data, which has important implications for the computation of the F ratios. Figure 3 offers some guidelines for the choice of the appropriate ANOVA or its nonparametric equivalent (if there is one).

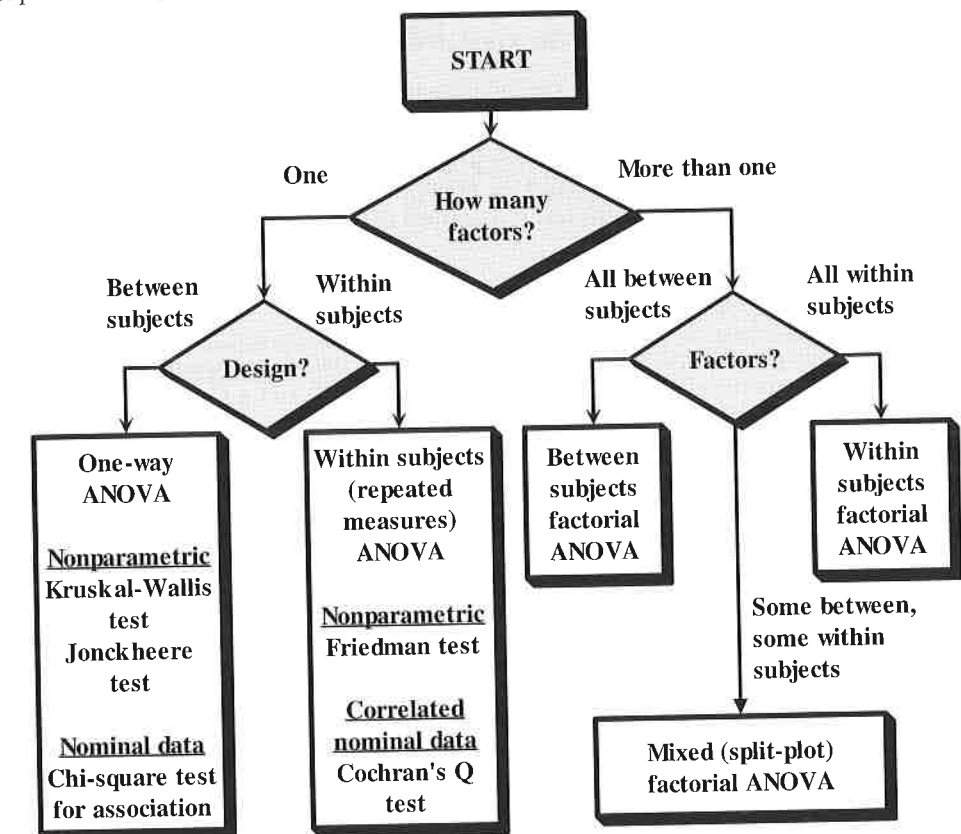


Figure 3. Flow chart for ANOVA experiments

To use the chart (Figure 3), begin at the START box and consider how many factors there are in the experiment. If there are two or more factors, work down the right-hand side of the chart. If there is just one factor, work down the left-hand side.

### 1.4.5 Analysing the results of one-factor experiments

Figure 3 outlines the decisions leading to the correct choice of ANOVA model for an experiment with one factor. The **one-factor between subjects ANOVA** is often called the **one-way ANOVA** and is more fully described in Chapter 7. The **one-factor within subjects ANOVA** (sometimes called the **subjects by treatments** or **randomised blocks ANOVA**) is described in Chapter 9.

See  
Chaps.  
7 & 9

*The case of two conditions only: Equivalence of the  $t$  tests and ANOVA*

The reader will have noticed that at the beginning of the section on ANOVA, the one-factor ANOVA was illustrated with the same simple two-condition experiments that were used to illustrate the use of the  $t$  tests. A natural question, therefore, is whether, in this special case, the independent and related  $t$  tests would lead to the same decision about the null hypothesis as would the one-way and within subjects ANOVA, respectively. In fact, they do.

*An alternative to the one-way ANOVA: The Kruskal-Wallis  $k$ -sample test*

So far, we have been considering the use of nonparametric alternatives to  $t$  tests. For data sets comprising more than two sets of measurements, however, there are also equivalent nonparametric tests. The **Kruskal-Wallis  $k$ -sample test** is equivalent to the one-way ANOVA: it is appropriate for independent samples of scores. The Kruskal-Wallis test is described in Chapter 7.

See  
Chap.  
7

*Independent samples of dichotomous nominal data*

Suppose that a sample of participants is divided randomly into three equally-sized groups: two experimental groups (*Group A* and *Group B*) and a *Control* group. Each participant is tested with a criterion problem, a 1 being recorded if they pass, and a 0 if they fail. With the resulting nominal data set, a **chi-square test** for association can be used to test the null hypothesis that, in the population, there is no tendency for the problem to be solved more often in some conditions than in others (see Chapter 11).

See  
Chap.  
11

*Correlated ordinal data: The Friedman test*

Suppose that twenty people rank ten paintings in order of preference, each person assigning the ranks of 1 and 10 to their most preferred and least preferred painting, respectively. This operation will yield ten related samples of data, each sample comprising the ranks assigned to one painting by the twenty raters. Is there a tendency for some paintings to receive higher ranks than others? An appropriate test for use with this type of data is the **Friedman test**, which is considered more fully in Chapter 9.

See  
Chap.  
9

*Three or more sets of correlated dichotomous nominal data: Cochran's  $Q$  test*

Suppose that each person in an experiment is tested under all conditions, and that a 1 is recorded for a successful performance under a given condition and a 0 otherwise. This is a one-factor, within subjects experiment, which has produced related samples of nominal data. Figure 3 suggests that an appropriate test of the null hypothesis of no difference in performance among the conditions is **Cochran's  $Q$  test**, which is described in Chapter 9.

See  
Chap.  
9

### 1.4.6 Analysing the results of factorial experiments

Figure 3 shows that for each of the three types of factorial experiment described here, there is a special ANOVA, which is appropriate for that kind of experiment only. The between subjects, within subjects and mixed ANOVA are described in Chapters 8, 9 and 10, respectively.

See  
Chaps.  
8, 9 &  
10

## 1.5 MEASURING STRENGTH OF ASSOCIATION BETWEEN VARIABLES

Do tall fathers tend to have tall sons, short fathers to have short sons and fathers of medium height to have sons of medium height? This question is one of a **statistical association** between the two variables *Father's Height* and *Son's Height*. To answer the question, you would need a data set comprising the heights of a substantial sample of fathers and those of their (first) sons.

### 1.5.1 Flow chart for selecting a suitable test for association

Figure 4 outlines the questions one needs to answer in order to make a decision about an appropriate measure of association.

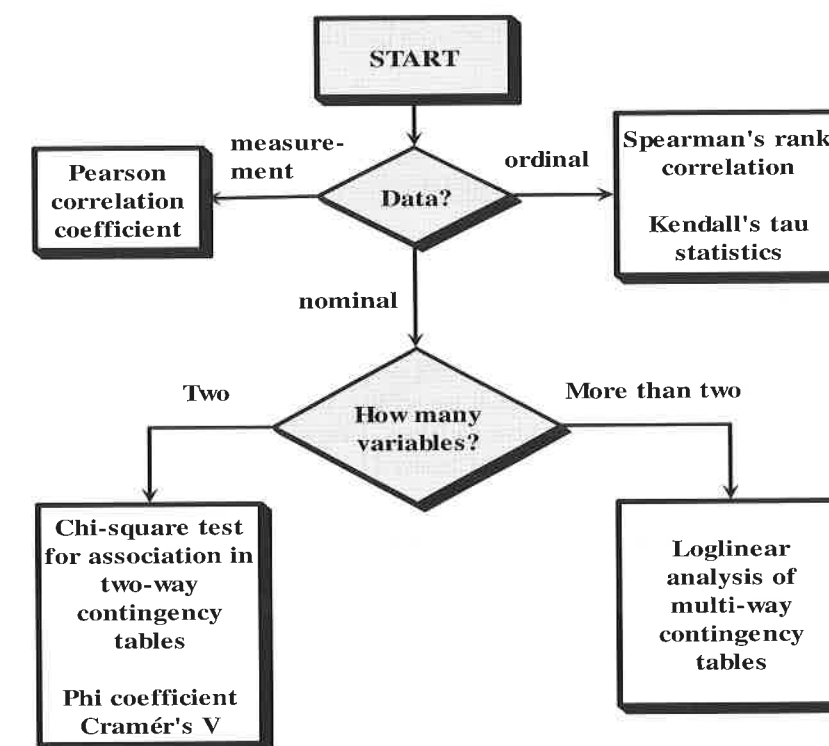


Figure 4. Flow chart showing measures of association

Begin at the START box and consider whether the data are measurements or ordinal. If the two variables are in the form of measurements, a **Pearson correlation** should be considered. However, as we shall see in Chapter 11, there are circumstances in which the Pearson correlation can be highly misleading. **It is essential to examine the data first before proceeding to obtain the correlation coefficient.**

See  
Chap.  
11

#### Measuring association in ordinal data

Now suppose we ask two judges to rank twenty paintings in order of preference. We shall have a data set consisting of twenty pairs of ranks. Do the judges agree? Again, our question is one of a statistical association. However, since the data are ordinal, a **rank correlation** is an appropriate statistic to use. There are two kinds of rank correlation: 1. **Spearman's rank correlation**; 2. the **Kendall tau** statistics. Both are considered more fully in Chapter 11.

See  
Chap.  
11

#### 1.5.2 Measuring association in nominal data: Contingency tables

Suppose that Fred claims to possess telepathic powers. Accordingly, in an experiment designed to test his claim, an experimenter tosses a coin 100 times, and Fred, seated behind a screen, states, on each occasion, whether the coin has turned up heads or tails. Table 5 presents the results of the experiment in what is known as a **contingency table**, an array designed to display the association (if any) between qualitative variables.

| Table 5. A contingency table |                     |          |
|------------------------------|---------------------|----------|
| Fred's guess                 | Experimenter's toss |          |
|                              | Head (H)            | Tail (T) |
| H                            | 45                  | 9        |
| T                            | 8                   | 38       |

The presence of an association can be confirmed by using a **chi-square test** (see Chapter 11). Since the value of the chi-square statistic depends partly upon the sample size, it is unsuitable as a measure of the strength of association between two qualitative variables. Figure 4 names two statistics that measure strength of association between qualitative variables: **Cramér's V** and the **phi coefficient**. Both measures are discussed in Chapter 11.

See  
Chap.  
11

#### 1.5.3 Multi-way contingency tables

In the past three decades, there have been dramatic developments in the analysis of nominal data in the form of multi-way contingency tables. Previously, tables with three or more attributes were often 'collapsed' to produce two-way tables. The usual chi-square test could then be applied. Such 'collapsing', however, is fraught with risk, and the tests may give highly misleading results. The advent of modern **loglinear analysis** has made it possible to tease out the relationships among the attributes in a way that was not possible before (see Chapter 13).

See  
Chap.  
13

## 1.6 PREDICTING SCORES OR CATEGORY MEMBERSHIP

If there is an association between variables, it is natural to ask whether this can be exploited to predict scores on one variable from knowledge of those on another. Such prediction is indeed possible, and the methods by which this is achieved will be briefly reviewed in this section.

There are also circumstances in which one would wish to predict not scores on a target or criterion variable, but membership of a category of a qualitative variable. For example, it is of medical and actuarial interest to be able to assign individuals to an 'at risk' category on the basis of their smoking and drinking habits. Statistical techniques have been specially devised for this purpose also.

The purpose of the methods reviewed here is to predict a target, or **criterion** variable (the term **dependent variable** is also used in this context) from scores on other variables, known variously (depending on the context) as **regressors**, **predictors**, **independent variables**, and **covariates**. The predictors need not always be quantitative variables: qualitative variables, such as gender and blood group, are often included among the predictor variables in research of this kind.

### 1.6.1 Flow chart for selecting the appropriate procedure for predicting a score or a category

To use the flow chart (Figure 5) for selecting the appropriate prediction procedure, begin at the START box and consider whether the target variable is qualitative (e.g. a set of categories such as *Pass* and *Fail*) or quantitative (e.g. examination scores).

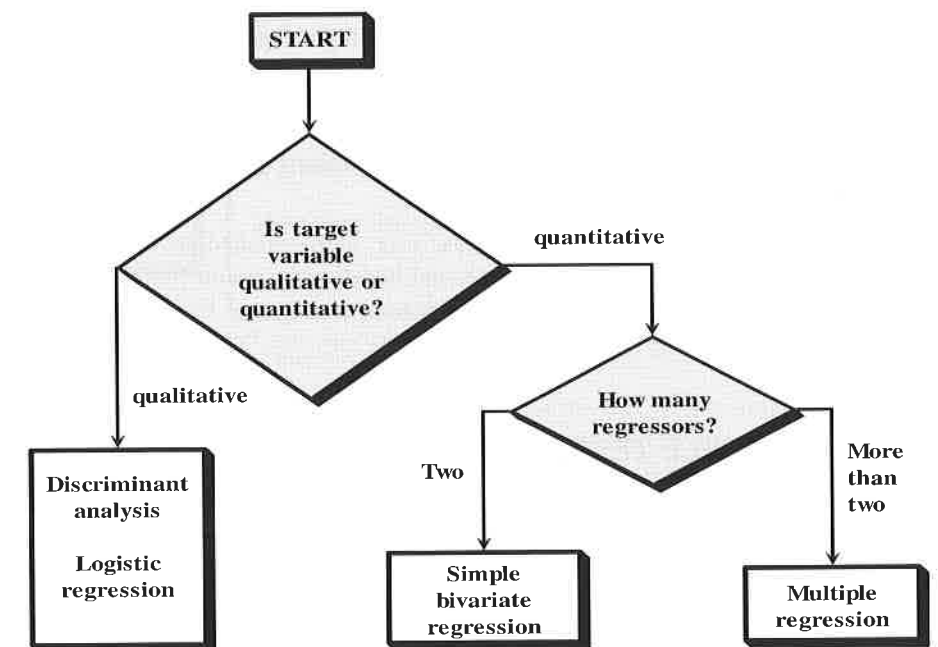


Figure 5. Flow chart showing procedures for prediction

Begin at the START box and consider the purpose of the test. If it is to test for goodness-of-fit, move down the left-hand side of the chart. If it is to estimate the population mean or its probable range, move down the right-hand side. The next consideration is the nature of the data: different types of data require different tests. If the target variable is quantitative, a **regression** method should be considered. In **simple regression**, there is one predictor; in **multiple regression**, there are two or more. Regression is the subject of Chapter 12. If the criterion variable is qualitative, the techniques of **discriminant analysis** and **logistic regression** should be considered. They are discussed in Chapter 14.

See  
Chap.  
14

### 1.6.2 Simple regression

In some US universities, the authorities are interested in predicting students' grade point averages after a year of study from the scores they achieved on aptitude and intelligence tests when they matriculated.

Suppose that, given a student's verbal aptitude score at matriculation, we want to predict the same student's grade point average a year later. This is a problem in **simple regression**, and its solution is described in Chapter 12.

See  
Chap.  
12

### 1.6.3 Multiple regression

A student's grade point average is associated not only with verbal aptitude, but also with numerical ability. Can grade point average be predicted even more accurately when both verbal ability and numerical ability are taken into account? This is a problem in **multiple regression**. If grade point average is correlated with both verbal and numerical aptitude, multiple regression will produce a more accurate prediction of a student's grade point average than will a simple regression upon either of the two regressors considered separately.

See  
Chap.  
12

### 1.6.4 Predicting category membership: Discriminant analysis and logistic regression

Two statistical techniques designed to help the user make predictions of category membership are **discriminant analysis** and **logistic regression** (both are discussed in Chapter 14). In recent years, logistic regression, being a somewhat more robust technique than discriminant analysis, has become the preferred method.

See  
Chap.  
14

## 1.7 ONE-SAMPLE TESTS

Much psychological research involves the collection of two or more samples of data. This is by no means always true, however: sometimes the researcher draws a **single** sample of observations in order to study just **one** population.

The situations in which one might use a one-sample test are of two main kinds:

1. One may wish to compare a sample distribution with a hypothetical distribution, such as the normal. This is a question of **goodness-of-fit**.
2. One may wish to make inferences about the parameters of a single population from the statistics of a sample, either for the purpose of

ascertaining whether the sample is from a known population or estimating the parameters of an unknown population.

### 1.7.1 Flow chart for selecting the appropriate one-sample test

Figure 6 summarises the circumstances in which a researcher might make various kinds of one-sample tests. The tests reviewed in this section are more fully considered in Chapter 6.

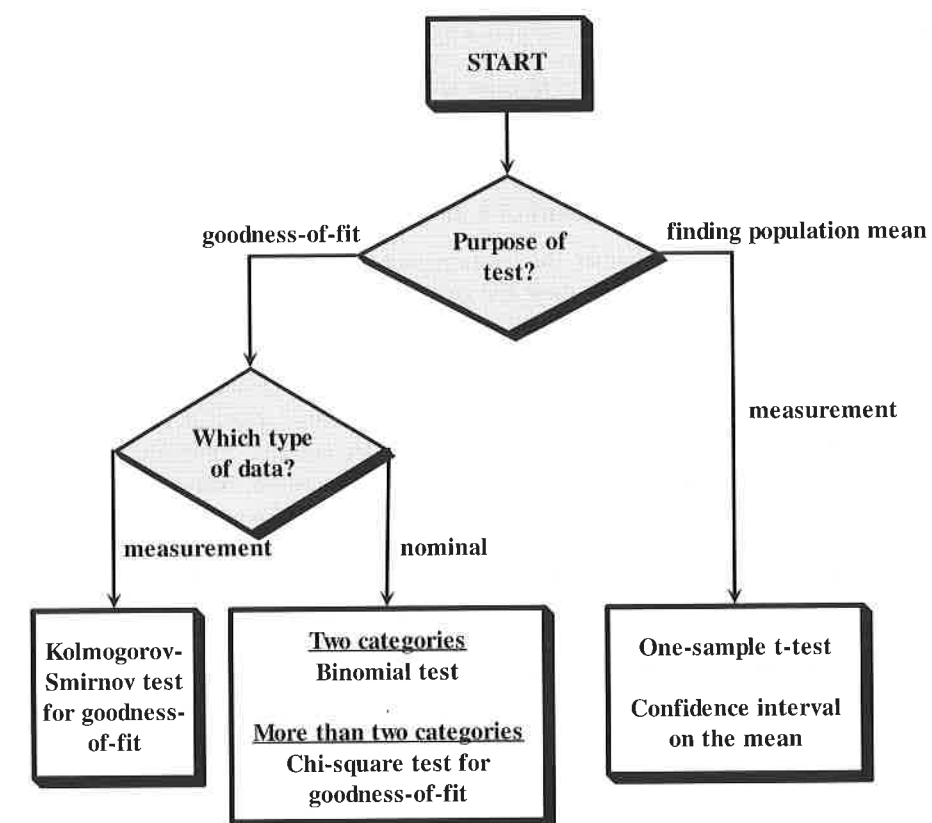


Figure 6. Flow chart of one-sample tests

Begin at the START box and consider the purpose of the test. If it is to test for goodness-of-fit, move down the left-hand side of the chart. If it is to estimate the population mean or its probable range, move down the right-hand side. The next consideration is the nature of the data: different types of data require different tests.

### 1.7.2 Goodness-of-fit: Data in the form of measurements

A question about a single population is often one of **goodness-of-fit**: has the sample been drawn from a population with a specified distribution shape? Suppose, for example, that one has a sample of measurements and wishes to ascertain whether these have been drawn from a normal population. Figure 6 shows that the **Kolmogorov-Smirnov test** is appropriate for this purpose.

See  
Chap.  
6

### 1.7.3 Goodness-of-fit: Nominal data

Suppose a researcher wants to know whether 5-year-old children of a certain age show a preference for one of two toys (A or B). The choices of one hundred 5-year-olds are noted. Here the population comprises the choices (A or B) of 5-year-olds in general. Of the hundred children in the study, 60 choose toy A and 40 toy B. The null hypothesis states that the probability of choosing A (or B) is 0.5: more formally, it states that we have sampled 100 times from a Bernoulli population with  $p = 0.5$ . Does this theoretical distribution fit our data? Figure 6 indicates that a **binomial test** can be used to test this hypothesis.

See  
Chap.  
6

If, in the foregoing example, there were three or more toys to choose from, the **chi-square test for goodness-of-fit** can be used to test the null hypothesis that the children have no preference for any particular toy.

### 1.7.4 Inferences about the mean of a single population

Suppose we want to know whether the performance of a group of schoolchildren on a standardised test is typical of those in their age group. Figure 6 shows that a **one-sample  $t$  test** can be used to test the null hypothesis that the population mean has the putative 'population' value. Often, however, as when the researcher is working with a non-standardised test, it may not be possible to specify any null hypothesis. Suppose that a lecturer wishes to ascertain the typical reaction speed of first year university students within a certain age group. The lecturer may have data on, say, two hundred first year students; but the research question, being about the reaction speeds of first year students in general, concerns the population of reaction times. The sample mean is a **point estimate** of the unknown population mean. The  $t$  distribution can also be used to build a **confidence interval** around the sample mean, so that the researcher has a range of values within which the true population mean is likely to lie.

See  
Chap.  
6

### 1.7.5 Nominal data: Testing a coin for fairness

When we toss a coin a large number of times to ascertain its fairness, we obtain a sample from the (infinite) population of such tosses. We might find that the coin turned up heads on 58 out of 100 tosses. Is the coin 'fair', that is, in the population, are the relative frequencies of heads and tails both 0.5?

The **binomial test** can be used to test the hypothesis that the population proportion is  $\frac{1}{2}$  (or, indeed, that it is any other specified proportion). A **confidence interval** can also be constructed on the sample proportion to give a range of values within which we can be confident that the true population proportion lies.

See  
Chap.  
6

## 1.8 FINDING LATENT VARIABLES: FACTOR ANALYSIS

Suppose that 500 people are measured on twenty tests of ability. It is likely that if the correlations between each test and every other test are arrayed in a **correlation matrix (R)**, there will be substantial positive correlations among the tests in the battery.

**Factor analysis** (see Chapter 15) is a set of techniques which, on the basis of the correlations in an **R matrix**, classify all the tests in a battery in terms of relatively few underlying dimensions or **factors**. The reader will have noticed that the term **factor** has more than one meaning in statistics, because it has already occurred in the context of analysis of variance (ANOVA), where it denotes an independent variable that is manipulated either experimentally or statistically by the experimenter. Actually, the two meanings are not quite so disparate as might first be supposed. Arguably, the factors that emerge from factor analysis are also independent variables, since they contribute to performance on tests in the battery, which can therefore be regarded as dependent variables. However, the factors emerging from factor analysis can never be directly measured or controlled by the researcher: they are **latent** variables.

See  
Chap.  
15

In **exploratory factor analysis**, the object is to find the minimum number of **factors** necessary to account for the correlations among the psychological tests. In **confirmatory factor analysis**, specified models are compared to see which gives the best account of the data.

## 1.9 A FINAL COMMENT

In this chapter, we have offered some advice about the circumstances in which one might consider using formal statistical tests to support the researcher's claim that what is true of a particular data set is likely to be true in the population. At this point, however, a word of warning is appropriate.

### *Formal tests, statistical models and their assumptions*

The making of a formal statistical test of significance always presupposes the applicability of a statistical **model**, that is, an interpretation (usually in the form of an equation) of the data set as having been generated in a specified manner. The model underlying the one-sample  $t$  test, for example, assumes that the data are from a normal population. To some extent, statistical tests have been shown to be **robust** to moderate violations of the assumptions of the models upon which they are based, that is, the actual error rates are not markedly different from the nominal values. But there are limits to this robustness, and there are circumstances in which a result, declared by an incautious user to be significant beyond, say, the 0.05 level, may actually have been considerably more likely than that. There is no way of avoiding this pitfall other than by thoroughly exploring the data first (see Chapters 4 and 5) to ascertain their suitability for specified formal tests.

See  
Chaps.  
4 & 5