5 The normal distribution and distributions associated with it

We have made use of the histogram on several occasions so far to give a pictorial or graphical view of how a set of scores is distributed. Let us take as a further example the distribution of heights of children of a given age. This distribution might look something like Figure 12. Each bar of the histogram represents a height range of

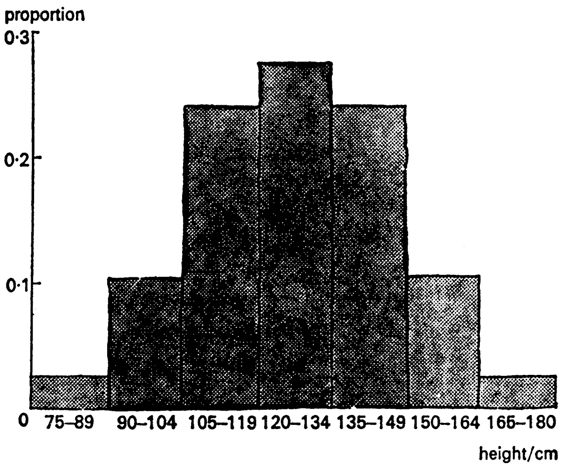

Figure 12 Histogram showing numbers of children with different heights

fifteen centimetres. There is no compelling reason why this range should be chosen. It could be smaller; indeed, the interval could be as small as one liked, providing that the measurements were sufficiently sensitive. (It only makes sense to do this when drawing a histogram if you have a large enough sample to ensure that the

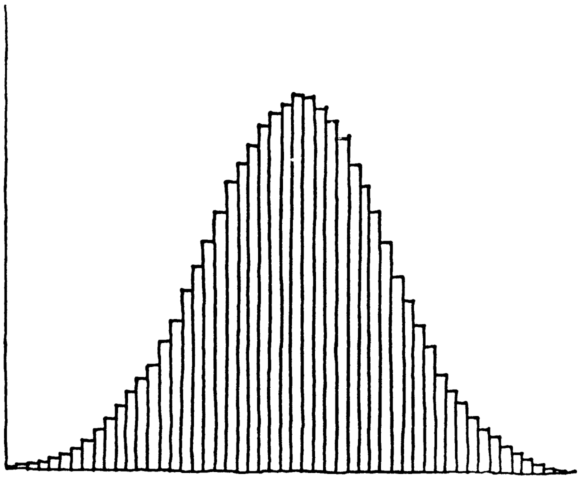

Figure 13 Approximation of a histogram to a smooth curve as the interval decreases

frequency or number of cases in different categories still remains reasonably large.)

A variable of this type, which can be continuously sub-divided, is called a **continuous** variable. There are other types of variables. Ones which can only take on particular numerical values are called **discrete** variables (e.g. size of family). There are also non-numerical **categorical** variables (e.g. gender – female or male).

By reducing the size of the interval with a continuous variable, the rectangles increase in number and become thinner so that the histogram approximates to a smooth curve when there is a sufficiently large sample (Figure 13). One such curve which you will hear referred to frequently is the **normal distribution curve**. As its properties were first investigated by Gauss, it is also commonly known as the **Gaussian distribution**. Incidentally, there is nothing abnormal or peculiar about other distribution curves – it just so happens that the so-called 'normal' curve is one which crops up many times and which has particularly useful, simple, and well-known mathematical properties. Different normal distributions vary only in their means and standard deviations and hence, if they are standardized so that they have the same mean and standard deviation, they have identical shapes. This intimate relationship between the standard deviation and the normal distribution is one of the reasons for the widespread use of the standard deviation.

The importance of the normal distribution

1 On theoretical grounds

It can be shown theoretically that, if we assume that there are many small effects all operating independently of each other to influence a particular score or other outcome, then the resulting distribution of scores is the normal distribution. As we have discussed previously, performance in an experiment is conceptualized as resulting from a large number of separate random errors arising from uncontrolled variables, in addition to the possible effect of the independent variable.

2 On practical grounds

If a sufficiently large number of observations or measurements are made so that the shape of the distribution can be assessed it will very frequently transpire that the distribution does actually approximate more or less closely to the normal distribution. For example, human height is distributed in this way (providing you control for gender and ethnic background – a distribution with both males and females would be bimodal – see p. 43). So is human intelligence as measured by IQ tests, although this tells us more about the standardization procedures used than anything else.

3 On mathematical grounds

It has already been pointed out that the normal distribution is particularly simple mathematically. It is also very useful that the results obtained by assuming a normal distribution are often applicable even when the distribution differs somewhat from the normal.

The shape of the normal distribution curve

The shape of the normal distribution curve is illustrated in Figure 14. It is bell-shaped and is symmetrical about its mean. In other words, if one imagines a vertical line drawn through the mean, then the shapes on either side of the line are identical. The median and mode occur at the same value as the mean. The curve falls away relatively slowly at first on either side of the mean, i.e. there

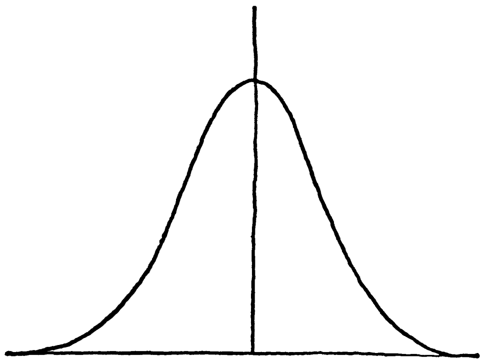

Figure 14 The normal distribution curve

is a high probability of scores occurring just a little above or just a little below the mean. When one gets to the 'tail' of the distribution, either above or below the mean, the curve approaches the horizontal axis 'asymptotically'. That is, the slope of the curve decreases continually for values further and further from the mean so that although the axis is approached it is never actually reached (although this is impossible to show on a drawing such as Figure 14). What this means is that there will be some very small probability of getting values a long way from the mean.

The standard deviation is associated with the curve in the following way. If we consider an upper limit obtained by going one standard deviation above the mean, and a lower limit obtained by going one standard deviation below the mean, a certain proportion of the cases, scores or whatever makes up the distribution, will be contained within these limits. For all normal distributions, this proportion is 0·6826, that is 68·26 per cent of the scores in any normal distribution fall within the limits of one standard deviation above and below the mean. In other words just over two-thirds of the scores are within a standard deviation of the mean.

If we consider wider limits, say two standard deviations above and below the mean, then the shape of the normal distribution is such that 95·44 per cent of the scores fall within these limits. For plus and minus three standard deviations, this percentage rises to 99·73 per cent.

To take an example: if it is known that the mean of a population is 100 (say the population is one of IQs which are standardized to a mean of 100) and the standard deviation is 15, then we know 68·26 per cent of the population will lie within limits of 115 to 85; 95·44 per cent within limits of 130 to 70; and 99·73 per cent within limits of 145 to 55.

Standard normal distribution

As was pointed out in the last chapter, standard scores (z-scores) can be obtained by expressing deviations from the mean in terms of standard deviation units,

$$\text{i.e. standard score } z = \frac{\text{deviation score } x}{\text{standard deviation}}$$

Tables of the normal distribution are usually given in this standard form. Table D (p. 163) is an example which shows the fractional area under the standard normal curve. It can be seen from this table that for a z-score of 1·0, the fractional area enclosed between the mean of the distribution (where $z = 0$) and $z = 1{\cdot}0$ is 0·3413. Figure 15 illustrates this.

As the curve is symmetrical, the area enclosed between z-scores of $-1{\cdot}0$ and $+1{\cdot}0$ is therefore $2 \times 0{\cdot}3413 = 0{\cdot}6826$. Another way of expressing this is to say that a proportion of 0·6826 (i.e. 68·26 per cent) is contained within these limits – which is what was stated in the last section.

Taking a different example of the use of Table D, look at the area corresponding to a z-score of 1·96. The table shows that this is

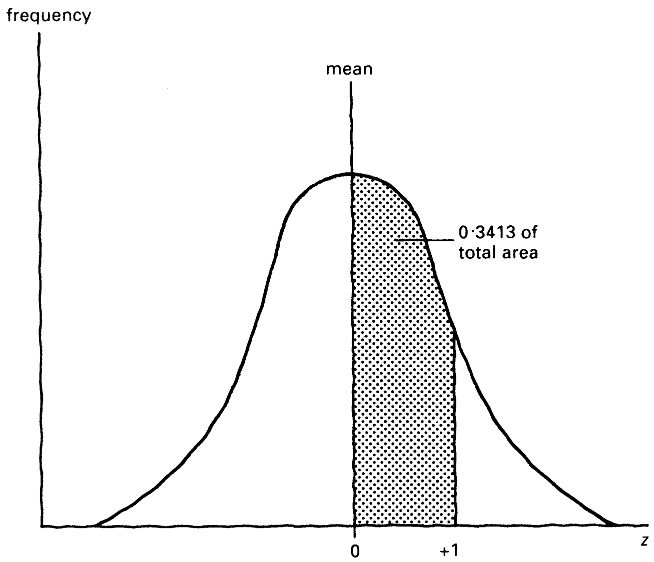

Figure 15 Fractional area enclosed between $z = 0$ and $z = +1$

0·4750. Therefore between the limits of $z = -1{\cdot}96$ and $z = +1{\cdot}96$ a fractional area of $2 \times 0{\cdot}4750 = 0{\cdot}95$ (95 per cent) is contained. This means, of course, that 5 per cent of the population exceeds these limits. Figure 16 shows this.

Hence, in an experimental situation where we can establish that we are dealing with the normal distribution, a z-score exceeding 1·96 can be used to demonstrate that the IV had an effect on the DV at the 5 per cent level of significance. Use the table to find the z-score corresponding to the 1 per cent level of significance.

Samples and populations

In an experiment, what we are doing is collecting a set of scores. Typically, these scores are considered as a sample taken from some population. By appropriate randomization techniques, we try to make sure that each member of the population has an equal chance

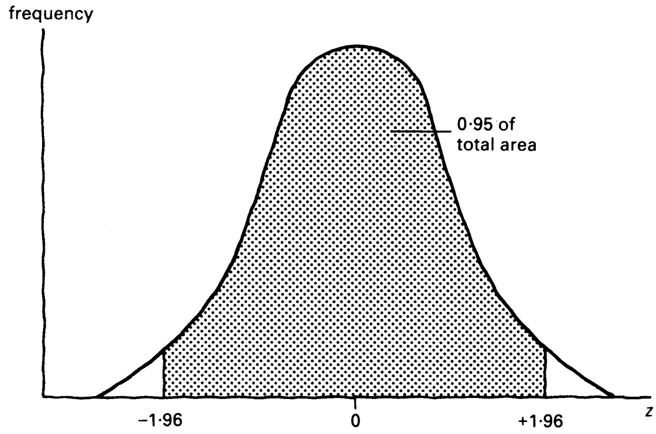

Figure 16 Fractional area enclosed between $z = +1{\cdot}96$ and $z = -1{\cdot}96$

of appearing in the sample. This then means that the results obtained from the sample can be generalized to the population.

You may be confused about the way in which we have been using the words 'sample' and 'population'. In statistics they are not limited to references to people. Thus, one talks about samples and populations of scores, as well as samples and populations of people.

When standard deviation was being discussed earlier, it was pointed out that the formula we used was appropriate for obtaining estimates of the population standard deviation from the scores in the sample. Thus, if from our sample we obtain a value of 80 for the mean and 10 for the standard deviation, then our estimate is that in the population from which the sample is drawn, 68 per cent of the scores will lie between the limits of 70 to 90.

This is only an estimate, of course. We have no certainty that it is correct. But it does mean that if, at a later stage, we obtain a score of 56 it would be quite improbable that it came from the same population (after all we estimate that approximately 95 per cent of the scores lie between 60 and 100, i.e. between plus and minus two standard deviations). It would not be impossible, however, for the score of 56 to come from the population – remember those tails approaching the horizontal axis asymptotically.

Comparing two samples

A problem which we are much more likely to be concerned with in attempting to evaluate the results of our experiments is the decision as to whether the scores obtained under one experimental treatment or condition differ from the scores obtained under another experimental treatment or condition. We have already considered one way in which a decision could be reached – by using the sign test – but there ought to be a way which would make direct use of the actual scores which we obtain. What we would really like to know is whether the difference in the mean scores of the two experimental samples can be taken as evidence that there is a genuine difference between the two experimental conditions. The question is a familiar one: Is the observed difference in means sufficiently unlikely when random errors alone are involved that we are willing to decide that something else apart from the random errors is having an effect? In other words that the IV is affecting the DV?

In order to attack this problem, we have to consider a special kind of standard deviation called the **standard error**.

Standard error

Suppose that we take as a population the actual population of male Chinese in the world. And further suppose that we take a random sample of 1000 of them and measure their heights. (The actual procedure necessary to get a random sample is left to the fertile imagination of the reader, as I am not sure that I could do it. Remember that each individual must have an equal chance of ending up in the sample.) The distribution might look something like Figure 17. Flushed by our success, we gather in a second sample of 1000 and measure again. This time we might find that the mean of the distribution was slightly higher. Repeating the process again with a third sample would lead to a slightly different mean again, and so on. The point I am trying to make is that we would not obtain identical means from successive samples. There

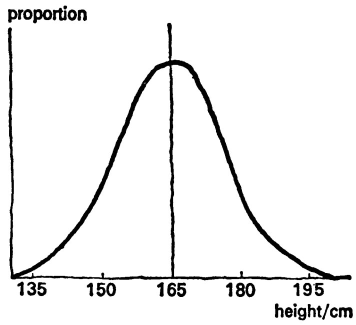

Figure 17 Distribution of heights for a sample of 1000 people (fictitious data)

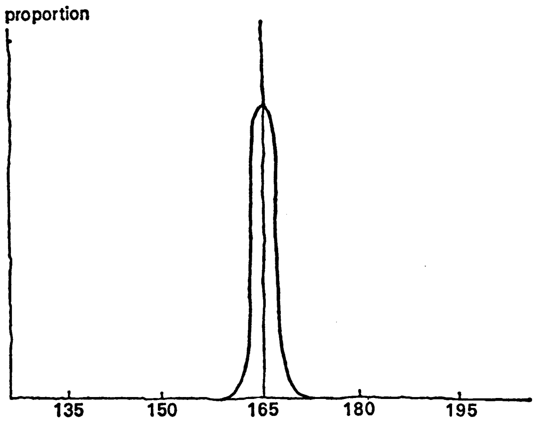

Figure 18 Distribution of *mean* heights of samples of 1000 people

would be a certain amount of variation. If one were to sample repeatedly in this way, taking 1000 people at a time, it would be possible to build up a **sampling distribution of the means**. This is a curve showing the frequency of occurrence of different mean scores, and it might look something like Figure 18. If this is compared with the sampling distribution of the scores themselves, it can be seen that there is much less variability in the mean scores than in the individual scores. An alternative way of putting this is to say that the standard deviation of the means is considerably less than the standard deviation of the scores themselves. *The standard deviation of the means is given a special name – the* **standard error**.

It is common sense that the means will vary much less than the individual scores. Mathematically there is a very simple and neat relationship between the standard error (SE) of the means, and the standard deviation (SD) of the scores:

$$\text{SE} = \frac{\text{SD}}{\sqrt{N}},$$

where N is the size of sample (1000 in the example quoted above).

The *t*-test

Let us say that, in an experiment, the mean score for condition A exceeded that for condition B, where A and B are two levels of a particular independent variable. Are we justified then in saying that the IV is affecting the DV? This is, once again, the question of generalizing from the sample of experimental results to the population. Two factors which would influence our decision are, firstly, the size of the difference in means and, secondly, the amount of variability in the scores. The bigger the difference in means, the more confidence we have that the sample difference reflects a real difference between the experimental conditions. But the larger the variability in scores, the less is our confidence.

These two factors are taken into account in the t-test, where

$$t = \frac{\text{difference in means}}{\text{standard error of the difference in means}}$$

Previously, we have talked about the standard error as being the name given to the standard deviation of the mean. This can be generalized to indicate the standard error of the difference in two means. If we draw two samples from a single population and look at the means of the two samples, we will find that almost always there will be some difference in the mean scores. Sometimes the difference will be negligible.

By taking repeated pairs of samples and each time noting the

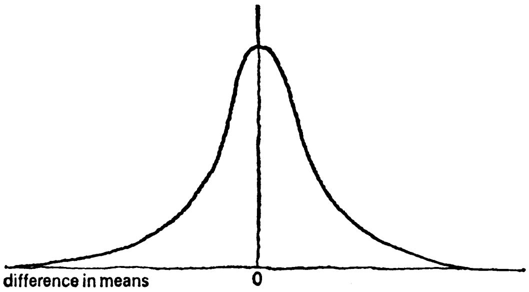

Figure 19 Sampling distribution of difference in means of two samples taken from the same population

difference in the means of the two samples, it will be possible to plot the sampling distribution of the difference in means in a similar manner to that in which the sampling distribution of the means was built up in Figure 18. An example of what this might look like is shown in Figure 19.

By the way, do not worry that you are going to have to spend the rest of your life painstakingly building up sampling distributions by taking sample after sample. By using statistical theory it is possible, after making certain assumptions about the population and sample, to derive formulae for the sampling distributions.

Do not be confused by Figure 19 looking very different from Figure 18. By choosing an appropriate scale for Figure 19, both figures could be made to look very much the same. The point made by drawing Figures 17 and 18 to the same horizontal scale is that the sampling distribution of individual scores is much wider than the sampling distribution of mean scores. Similarly, the sampling distribution for difference in individual scores would be much wider than the sampling distribution for difference in means if it were drawn to the same horizontal scale as Figure 19.

It is perhaps surprising that if the samples drawn from the population are relatively small (say 50 or less) then the sampling distribution of t is not normal, although the underlying population is itself normal. The distribution is known as the t-distribution and differs slightly in shape from the normal distribution. However, it

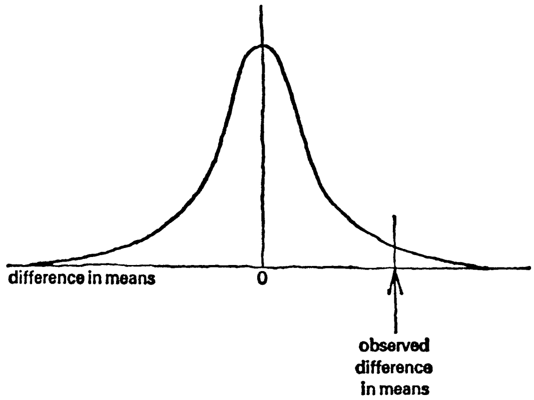

Figure 20 Observed difference in means superimposed on the sampling distribution of Figure 19

gets closer and closer to the normal distribution as the sample size increases.

Let us say that the observed difference in means in a particular experiment is as shown in Figure 20. If the samples had been drawn from the same populations this difference in means is unlikely to occur. As you can see, it takes us into one of the tails of the distribution. But is it sufficiently improbable for us to come to the decision that the independent variable did have an effect on the dependent variable?

The way in which we come to this decision is identical to the way in which we came to a decision for the sign test. We decide on a significance level. Remember that the significance level is the probability of making a type 1 error, i.e. the probability of deciding that the independent variable had an effect on the dependent variable when this is not the case. If we choose the conventional significance level of 5 per cent this amounts to cutting off 5 per cent of the *t*-distribution. This has been done in Figure 21. Here the distribution is divided into a central region, where the decision is made that the IV had no effect on the DV; and the two tails (shaded in the figure) where the decision is made that the IV did have an effect on the DV. What value of difference in means do we take as the critical value, i.e. the one which is exceeded by 5 per cent of the population? This is given, in terms of *t*, in Table E (p. 164).

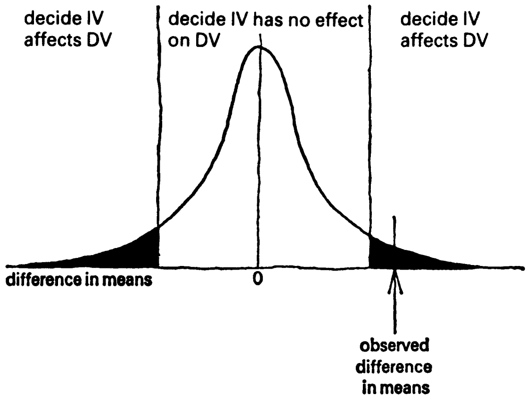

Figure 21 Cut-off points added to Figure 20

As discussed above, the t statistic is the difference in means divided by the standard error. An alternative way of saying the same thing is by regarding it as a difference in means measured in units of standard error. A t of 2 indicates that the difference in means is twice the standard error, and so on. A feature of the t-distribution is that it changes shape somewhat as the size of the sample changes. This means that the critical 5 per cent value changes with the sample size (which is related to the d.f. – degrees of freedom – of Table E).

Computation of *t*

The computation of t differs according to whether we are using an independent samples design on the one hand; or a matched pairs or repeated measures design on the other hand (see pp. 18–19).

Assumptions underlying the *t*-test

In deriving the t-distribution, certain assumptions have to be made about the populations from which the samples are drawn. These are that the population distributions are **normal** and of the **same variance** (sometimes called the **homogeneity of variance** assumption). It is possible to test for the reasonableness of these assumptions for particular sets of scores. The **chi-square test** (Chapter 6) can be used as a test of 'goodness of fit' to a normal distribution. The details of how this is done are not given in Chapter 6. They are available in more advanced texts, e.g. Hayes (1981). The variance-ratio test (discussed later in this chapter) can be used to test the homogeneity of variance assumption.

However, statisticians have demonstrated that the *t*-test is extremely **robust** with respect to violation of these assumptions. This means there can be considerable deviation from normality and/or homogeneity of variance without the result of the *t*-test being affected. An exception to this is with the independent sample design when there are different numbers of scores under the two experimental conditions. Here, violations of the homogeneity of variance assumptions can be serious, and it is worthwhile to test this assumption (using the variance-ratio test) before carrying out the *t*-test.

The strategy recommended for all other cases is to examine the data and then, unless there are glaring deviations from either normality or from homogeneity of variance, go ahead with the *t*-test.

One-tail and two-tail tests

There are some situations where we have a good reason for specifying the expected direction of the difference between the means (or, more generally, the direction of the effect of the IV). This reason may be theoretical, that is, it comes out as prediction from a theory. Or it may be from previous work done in the area, either by yourself or others. In some cases, it may simply be common sense. If, for example, you are working on some aspect of the abilities of persons who have recently suffered from strokes, it is a reasonable presupposition that they may be inferior to those of control non-stroke persons.

In situations such as this use can be made of what is called a **one-tailed test**. The reason for the name is obvious – we are dealing with only one of the tails of the distribution shown in Figure 21. In this case the null hypothesis that there is no difference in means is being tested against a directional alternative hypothesis where condition A (say) has a higher mean than condition B.

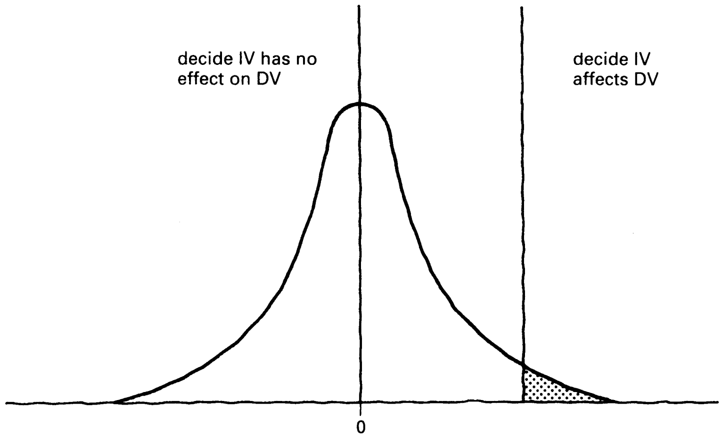

Figure 22 One-tailed test

We will, if we are using a one-tailed test, only decide that the IV has affected the DV if the experimental result falls at one end of the distribution – as illustrated in Figure 22.

The previous situation, as shown in Figure 21, is a **two-tailed test**. This is appropriate when the null hypothesis (of *no* difference) is being tested against a non-directional alternative hypothesis (simply that there *is* a difference).

There is an important difference in interpretation of significance levels for one-tailed and two-tailed tests. **The 5 per cent *t*-value for a two-tailed test becomes a 2·5 per cent value if a one-tailed test is used**. This is because the 5 per cent value in the table refers to 5 per cent of the distribution occurring in the two tails taken together; hence there is 2·5 per cent in each of the separate tails. So, if we wish to use the conventional 5 per cent significance level in conjunction with a one-tail test, we must use the 10 per cent value in the table. These values are not shown in Table E but are available in books of Statistical Tables.

It may have occurred to you that, because of this, it can happen that a result which would not be statistically significant if a two-tail test is used may become statistically significant if a one-tail test is used. Whilst this might seem somewhat fishy, remember that the decision about what kind of hypothesis you are dealing with should be made *before* doing the experiment and not after. There are people who argue that two-tailed tests should always be used and, starting out on experimentation as you are, it is wise to deal almost exclusively in terms of two-tailed tests. Remember that if you have a one-tailed test and the result comes out in the opposite direction to that hypothesized, then you cannot conclude that the IV had an effect on the DV, even in cases where the difference in means is large.

Significance revisited

By this time you should be getting more of a feel of the meaning of statisical significance. If you are not you should skip back to p. 32 and review this. It is perhaps appropriate to add a word of warning here. In common usage 'significant' means something like 'important'. However, remember that statistical significance simply tells us that something is unlikely to have happened by chance. This in itself tells us little about the practical importance of the effect which we have found. In particular you should note that by increasing the sample size, that is by collecting more data, you are going to make it more likely to get a statistically significant result. This appears intuitively clear in the case of the *t*-test where increasing the sample size (N) has the effect of decreasing the standard error ($= SD/\sqrt{N}$) and hence of increasing the value of t (= difference in means/SE).

Another way of putting this is to appreciate that with a sufficiently large sample size one can get a significant value of t with a very small difference in means. In practical terms this difference in means may be trivial, particularly if we have a complex situation where other variables are more important. So, the general message is that statistical significance is not all-important and, once again, that you cannot afford to switch off your common sense when interpreting the results of experiments.

t-Test – independent samples

Use with independent samples design
(NB **steps A1–5** and **B1–5** are identical to **steps 1–5** of the standard deviation procedure, p. 52.)

Step A1	Add all A observations together	ΣX_A
Step A2	Divide **A1** (i.e. the result of **step A1**) by the number of A observations N_A	$\frac{\Sigma X_A}{N_A} = \bar{X}_A$
Step A3	(a) Square each of the A observations	X_A^2
	(b) Add all the squares together	ΣX_A^2
Step A4	(a) Square **A1**	$(\Sigma X_A)^2$
	(b) Divide **A4a** by N_A	$\frac{(\Sigma X_A)^2}{N_A}$
Step A5	Subtract **A4b** from **A3b**	$\Sigma X_A^2 - \frac{(\Sigma X_A)^2}{N_A}$

Steps B1–5 Repeat the above 5 steps for the B observations

Worked example

t-Test – independent samples

A observation	**Step A3a** (*A observation*)^2	*B observation*
3	9	6
5	25	5
2	4	7
4	16	8
6	36	9
2	4	4
7	49	7
—	—	8
Step A1 $\Sigma X_A = 29$	**Step A3b** $\Sigma X_A^2 = 143$	9
		7

Step A2 $\bar{X}_A = \dfrac{29}{7} = 4{\cdot}14$

Step A4a $(\Sigma X_A)^2 = 29^2$

Step A4b $\dfrac{(\Sigma X_A)^2}{N_A} = \dfrac{29^2}{7} = 120{\cdot}1$

Step A5 $\Sigma X_A^2 - \dfrac{(\Sigma X_A)^2}{N_A} = 143 - 120{\cdot}1 = 22{\cdot}9$

Step B1 $\Sigma X_B = (6 + 5 + 7 + 8 + 9 + 4 + 7 + 8 + 9 + 7)$
$= 70$

Step B2 $\bar{X}_B = \dfrac{70}{10} = 7$

Step B3 $\Sigma X_B^2 = (36 + 25 + 49 + 64 + 81 + 16 + 49 + 64 + 81$
$+ 49) = 514$

Step B4a $(\Sigma X_B)^2 = 70^2$

Step B4b $\dfrac{(\Sigma X_B)^2}{N_B} = \dfrac{70^2}{10} = 490$

Step B5 $\Sigma X_B^2 - \dfrac{(\Sigma X_B)^2}{N_B} = 514 - 490 = 24$

Step-by-step procedure – continued

t-Test – independent samples

Step 6 Add **A5** and **B5**

$$\left[\Sigma X_A^2 - \frac{(\Sigma X_A)^2}{N_A}\right] + \left[\Sigma X_B^2 - \frac{(\Sigma X_B)^2}{N_B}\right]$$

Step 7 Divide **6** by N_A minus 1 added to N_B minus 1

$$\frac{[\Sigma X_A^2 - (\Sigma X_A)^2/N_A] + [\Sigma X_B^2 - (\Sigma X_B^2)/N_B]}{(N_A - 1) + (N_B - 1)}$$

Step 8 Find the reciprocal of N_A and the reciprocal of N_B and add them together

$$\frac{1}{N_A} + \frac{1}{N_B}$$

Step 9 Multiply **7** by **8**

$$\frac{[\Sigma X_A^2 - (\Sigma X_A)^2/N_A] + [\Sigma X_B^2 - (\Sigma X_B)^2/N_B]}{(N_A - 1) + (N_B - 1)} \times \left(\frac{1}{N_A} + \frac{1}{N_B}\right)$$

Step 10 Take the square root of **9**

Step 11 Take the difference between **A2** and **B2**

$$\bar{X}_A - \bar{X}_B$$

Step 12 Divide **11** by **10**: the result is *t*!!

$$t = (\bar{X}_A - \bar{X}_B) \div \sqrt{\left[\frac{\{\Sigma X_A^2 - (\Sigma X_A)^2/N_A\} + \{\Sigma X_B^2 - (\Sigma X_B)^2/N_B\}}{(N_A - 1) + (N_B - 1)} \times \left(\frac{1}{N_A} + \frac{1}{N_B}\right)\right]}$$

with $(N_A - 1) + (N_B - 1)$ degrees of freedom

Step 13 Translate the result back in terms of the experiment

Worked example – continued

t-Test – independent samples

Step 6 $22{\cdot}9 + 24 = 46{\cdot}9$

Step 7 $\dfrac{46{\cdot}9}{(7-1)+(10-1)} = \dfrac{46{\cdot}9}{15} = 3{\cdot}13$

Step 8 $(\frac{1}{7} + \frac{1}{10}) = (0{\cdot}1429 + 0{\cdot}1000) = 0{\cdot}2429$

Step 9 $3{\cdot}13 \times 0{\cdot}2429 = 0{\cdot}760$

Step 10 $\sqrt{0{\cdot}760} = 0{\cdot}872$

Step 11 $4{\cdot}14 - 7 = -2{\cdot}86$

Step 12 $t = -\dfrac{2{\cdot}86}{0{\cdot}872} = -3{\cdot}28$

with $(7 - 1) + (10 - 1) = 15$ degrees of freedom
From Table E, $t = 2{\cdot}13$ at the 0·05 level of significance (i.e. $p = 5$ per cent) with 15 degrees of freedom

Step 13 We therefore conclude that the IV had an effect on the DV, as the observed value of t is numerically greater than 2·13

Step-by-step procedure

t-Test – correlated samples

Use with matched pairs or repeated measures design

Step 1	Obtain the difference (d) between each pair of scores	$d = (X_A - X_B)$
Step 2	Add all the differences together	Σd
Step 3	Divide **2** (i.e. the result of **step 2**) by the number of pairs of scores (n)	$\frac{\Sigma d}{n} = \bar{d}$
Step 4	(a) Square each of the differences (b) Add all the squares together	d^2 Σd^2
Step 5	(a) Square **2**	$(\Sigma d)^2$
	(b) Divide **5a** by n	$\frac{(\Sigma d)^2}{n}$
Step 6	Subtract **5b** from **4b**	$\Sigma d^2 - \frac{(\Sigma d)^2}{n}$
Step 7	Divide **6** by $n(n - 1)$	$\frac{\Sigma d^2 - (\Sigma d)^2/n}{n(n - 1)}$
Step 8	Take the square root of **7**	
Step 9	Divide **3** by **8**: the result is t $t = \bar{d} \div \sqrt{\frac{\Sigma d^2 - (\Sigma d)^2/n}{n(n - 1)}}$ with $(n - 1)$ degrees of freedom	
Step 10	Translate the result of the test back in terms of the experiment	

Worked example

t-Test – correlated samples

The data represents scores obtained by 7 people in a certain test with and without the presence of a drug

Partici-pant	*Scores with*	*Scores without*	**Step 1** d	**Step 4a** d^2
1	3	6	3	9
2	8	14	6	36
3	4	8	4	16
4	6	4	−2	4
5	9	16	7	49
6	2	7	5	25
7	12	19	7	49

Step 2 $\Sigma d = 30$

Step 3 $\frac{\Sigma d}{n} = \bar{d} = \frac{30}{7} = 4{\cdot}29$

Step 4b $\Sigma d^2 = 188$

Step 5a $(\Sigma d)^2 = 30^2 = 900$

Step 5b $\frac{(\Sigma d)^2}{n} = \frac{900}{7} = 128{\cdot}57$

Step 6 $\Sigma d^2 - \frac{(\Sigma d)^2}{n} = 188 - 128{\cdot}57 = 59{\cdot}43$

Step 7 $\frac{\Sigma d^2 - (\Sigma d)^2/n}{n(n-1)} = \frac{59{\cdot}43}{7 \times 6} = 1{\cdot}41$

Step 8 $\sqrt{1{\cdot}41} = 1{\cdot}19$

Step 9 $t = \bar{d} \div \sqrt{\frac{\Sigma d^2 - (\Sigma d)^2/n}{n(n-1)}} = \frac{4{\cdot}29}{1{\cdot}19} = 3{\cdot}61$

with $(7 - 1) = 6$ degrees of freedom
From Table E, $\underline{t = 2{\cdot}45}$ at the 0·05 level of significance, with 6 degrees of freedom

Step 10 We therefore conclude that the IV had an effect on the DV, as the observed t is numerically greater than 2·45: the drug produced a significant decrease in mean score on this test ($t = 3{\cdot}61$ with 6 d.f. Significant at the 5 per cent level)

The variance-ratio test (*F*-test)

In the discussion above, we have been concerned with hypotheses about differences in means between two conditions. Although many behavioural hypotheses can be translated into statistical hypotheses about means, there are others where the appropriate statistical hypothesis is concerned with the relative dispersion of scores under two conditions. The variance-ratio test (or *F*-test) is suitable for these situations.

(What is variance? If you have forgotten or are not sure return to p. 48.)

As an example, consider performance in some task with the preferred hand as against the non-preferred hand. Suppose we get our participants to play shove ha'penny with the preferred hand on some occasions, the non-preferred hand on others. One obvious way of comparing performance with the two hands would be by comparing the variability in aiming by the two hands, i.e. by using the variance-ratio test to compare the two variances.

You should note that the values shown in Table F relate to significance for a two-tailed test. Many books of tables give values for a one-tailed test which are appropriate for other uses of the variance-ratio test.

Assumptions underlying the variance-ratio test

As with the *t*-test, the variance-ratio test makes assumptions about the underlying population distribution. Again the assumption is of normality. However, once again, the test is robust and the recommended action is to carry on with the variance-ratio test unless the distributions are very far from normal.

Step-by-step procedure

Variance-ratio test (*F*-test)

Step 1 Obtain the variance separately for each set of scores (use the standard deviation step-by-step procedure as far as **step 6**; see p. 52)

Step 2 Obtain $F = \frac{\text{larger variance}}{\text{smaller variance}}$

Step 3 Look up the significance of *F* in Table F. Note that you need two sets of degrees of freedom to do this. The columns in the table refer to the degrees of freedom of the top line of F ($N_1 = N_A - 1$, where N_A is the number of scores making up the larger variance). The rows in the table refer to the degrees of freedom of the bottom line of F ($N_2 = N_B - 1$, where N_B is the number of scores making up the smaller variance)

Step 4 Translate the results of the test back in terms of the experiment

Note The values in Table F are appropriate for a two-tail test (i.e. testing for a difference in the variances without *a priori* specifying the direction of the difference). Many tables of *F* refer to a one-tail test, which is useful in a different application of *F*

Worked example

Variance-ratio test (*F*-test)

The following error scores were obtained in an aiming test
Non-preferred hand (A): 3·3, 2·1, 4·7, 0·1, 5·6, 0·0, 4·7
Preferred hand (B): 5·6, 4·9, 6·2, 5·1, 5·8, 6·3

Step 1 (using the method on p. 52)

X_A	(3a) X_A^2	X_B	(3a) X_B^2
3·3	10·89	5·6	31·36
2·1	4·41	4·9	24·01
4·7	22·09	6·2	38·44
0·1	0·01	5·1	26·01
5·6	31·36	5·8	33·64
0·0	0·00	6·3	39·69
4·7	22·09		

(1) $\Sigma X_A = 20{\cdot}5$ **(3b)** $\Sigma X_A^2 = 90{\cdot}85$

(2) $\bar{X}_A = \frac{20{\cdot}5}{7} = 2{\cdot}93$

(4a) $(\Sigma X_A)^2 = 20{\cdot}5^2 = 420{\cdot}25$

(4b) $\frac{(\Sigma X_A)^2}{N_A} = \frac{420{\cdot}25}{7} = 60{\cdot}04$

(5) $\Sigma X_A^2 - \frac{(\Sigma X_A)^2}{N_A} = 90{\cdot}85 - 60{\cdot}04 = 30{\cdot}81$

(6) $\text{variance}_A = \frac{30{\cdot}81}{6} = \underline{5{\cdot}14}$

(1) $\Sigma X_B = 33{\cdot}9$ **(3b)** $\Sigma X_B^2 = 193{\cdot}15$

(2) $\bar{X}_B = \frac{33{\cdot}9}{6} = 5{\cdot}65$

(4a) $(\Sigma X_B)^2 = 33{\cdot}9^2 = 1149{\cdot}21$

(4b) $\frac{(\Sigma X_B)^2}{N_B} = \frac{1149.21}{6} = 191{\cdot}54$

(5) $\Sigma X_B^2 - \frac{(\Sigma X_B)^2}{N_B} = 193{\cdot}15 - 191{\cdot}54 = 1{\cdot}61$

(6) $\text{variance}_B = \frac{1{\cdot}61}{5} = \underline{0{\cdot}322}$

Step 2 $F = \frac{5{\cdot}14}{0{\cdot}322} = 15{\cdot}9$

Step 3 $N_1 = N_A - 1 = 6$; $N_2 = N_B - 1 = 5$
Table value of $F = 6{\cdot}98$ ($p = 0{\cdot}05$)
As the observed value of F exceeds the table value, there is a significant difference between the variances at the 5 per cent level

Step 4 Inspection of the results shows that the variance for the non-preferred hand exceeds that for the preferred hand. This difference is significant at the 5 per cent level