Table 2. Values of Student's t.

| Percentage points for confidence limits | | 95% | 99% |
Percentage points for significance tests		5%	1%
Degrees of freedom:	1	12.70	63.70
	2	4.30	9.93
	3	3.18	5.84
	4	2.78	4.60
	5	2.57	4.03
	6	2.45	3.70
	8	2.31	3.36
	10	2.23	3.17
	12	2.18	3.06
	14	2.15	3.00
	16	2.12	2.92
	18	2.10	2.88
	20	2.09	2.85
	30	2.04	2.75
	100	1.98	2.62
	many	1.96	2.58

In some tables, the headings are in terms of probability values rather than percentages: 0.05 = 5%, 0.01 = 1%, etc.

The mean tells the reader the average value. The standard deviation tells him how variable the birds in the population are. The sample size allows him to work out how precise your estimates of the mean and standard deviation are likely to be and to carry out all the statistical calculations he is likely to want to perform on your data.

It is also helpful to the reader to provide him with a measure of precision by giving confidence limits (or standard errors) rather than leaving him to work them out for himself.

t table: an explanation

You will see that Table 2 is headed with two sets of "percentage points", those for confidence limits and those for significance tests. The latter are simply the complements of the former. Most published tables simply have the points for significance tests in their headings: for 95% confidence limits we need the column headed 5% in such tables. If in any doubt about which column to use, remember that the right one for 95% confidence limits is the one in which the values for the higher numbers of degrees of freedom are close to 2.

Introductory statistics 3

JEREMY J.D. GREENWOOD

Department of Biological Sciences, The University, Dundee, Scotland, UK

Citation: Greenwood, J.J.D. 1979. Introductory statistics 1. *Wader Study Group Bull.* 26: 19–22.

Estimating the difference between two means

Consider the summary data of Table 3. It is fairly clear that the males in the population from which the sample was drawn are larger than the females on average: the sample means are 115.40 mm (males) and 112.14 mm (females). Since these are the best estimates of the population means, it seems reasonable to say that our best estimate of the mean difference between male and female winglengths in the population is $115.40 - 112.14 = 3.26$ mm.

Thus we can say that, on the evidence available, males have wings 3.26 mm longer than females in this population.

The standard error of the difference

Just as we can measure how precise are our estimates of the means, we can measure how precise is the estimate of the difference. The difference also has a standard error associated with it. This may be calculated using a formula that looks horribly complicated but is easy to use. If n_1 and n_2 are the two sample sizes and s_1^2 and s_2^2 are the two estimated variances, then s.e. of difference:

$$s_{diff} = \sqrt{\left[\frac{(n_1 - 1)\,s_1^2 + (n_2 - 1)\,s_2^2}{n_1 + n_2 - 2}\right]\left[\frac{n_1 + n_2}{n_1 n_2}\right]}$$

For the data of Table 3:

$$s_{diff} = \sqrt{\left[\frac{(15 - 1)\,1.45^2 + (21 - 1)\,1.60^2}{15 + 21 - 2}\right]\left[\frac{15 + 21}{15 \times 21}\right]}$$

$$= 0.521 \text{ mm}$$

Confidence limits of the difference

Confidence limits can be calculated in the usual way, using Student's t with $n_1 + n_2 - 2$ degrees of freedom. For the data of Table 3, $n_1 + n_2 - 2 = 34$. The value of Student's t for 95% confidence limits and 34 degrees of freedom is 2.03. Thus $t.s_{diff} = 2.03 \times 0.521 = 1.06$ mm: the 95% confidence limits of the difference are 2.20 mm and 4.32 mm.

Thus we can say that the best estimate of the difference

Table 3. Wing lengths. Summary data for sample of males and females from the same population (all in mm)

	Sample size	Estimated mean	95% confidence limits		Estimated standard deviation
Males	15	115.40	114.60	116.20	1.45
Females	21	112.14	111.41	112.85	1.60

in wing-length between males and females is 3.26 mm, with 95% confidence limits of 2.20 mm and 4.32 mm. The chances are 95% that the true average difference is between 2.20 mm and 4.32 mm.

What if the lower confidence limits is negative?

Suppose that we had samples taken in spring and autumn and that the mean difference in weights was 4.5 g (spring birds heavier), with $s_{diff} = 3.2$ g, and the sample-sizes were 28 and 34. The number of degrees of freedom is 60, so the relevant t value is 2.00, giving $t.s_{diff} = 6.4$ g.

The upper 95% confidence limit is $4.5 + 6.4 = 10.9$ g. The lower 95% confidence limit is $4.5 - 6.4 = -1.9$ g. What does this mean?

Our best estimate of the difference is that it is 4.5 g, spring birds being heavier than autumn birds. The upper 95% confidence limit is that spring birds are 10.9 g heavier, The lower limit is that spring birds are -1.9 g heavier: i.e. 1.9 g lighter.

The chances are 95% that the true difference lies somewhere in the range: spring birds 1.9 g lighter to spring birds 10.9 g heavier. We cannot say with any objective confidence whether spring birds are heavier or lighter.

Such a result will be obtained when the difference between the means is small compared with the precision with which it has been estimated. Thus, although we may feel disgruntled that we cannot really tell whether spring or autumn birds are the heavier, we may at least conclude that any difference is probably small.

More than two means

Suppose one has samples from a number of populations. One could take all possible pairs of samples and examine the differences between the means in the way I have been discussing. There are, however, various drawbacks to such a procedure. One is the time and effort involved if the number of samples is at all large: there are 10 different pairs for 5 samples, 45 pairs for 10 samples, and 190 for 20.

Another drawback is that examining only two samples at once ignores the information available in all the other samples. The other samples can, in particular, give us information about the amount of variation between individuals: this is useful in assessing how precise our estimates are.

Finally, difficulties of interpretation of confidence limits arise when we are making a whole series of comparisons. I shall discuss these a little when considering significance tests.

Fortunately, there is an elegant and powerful technique available for dealing with several samples: the analysis of variance.

Table 4. Some illustrative data for analysis of variance

	Population 1	Population 2	Population 3	
	1	2	3	
	2	3	3	
	2	3	3	
	3	4	4	
		5	4	
			5	
			6	Total
n	4	5	7	16
Σx	8	17	27	52
Σx^2	18	63	115	196
$(\Sigma x)^2$	64	289	729	2704

$$\Sigma \left[\frac{(\Sigma x)^2}{n} \right] = \frac{64}{7} + \frac{289}{5} + \frac{729}{7} = 177.94$$

The idea of analysis of variance (anova)

Let us continue to think of our several populations. There will be differences between individual birds within populations, which can be measured by the "variance within populations" (s^2). There may also be differences between the means of the populations: this can be measured by the "added variance among populations" (s_A^2). Anova allows us to estimate these.

The differences between birds within samples are the result simply of individual variation. The differences between the sample means are the result of both the variance between populations and of differences between individual birds making up the samples. These two sets of differences may be measured by the "mean squares".

Expressed formally:

Mean square within samples $= s^2$

Mean square between samples $= s^2 + n_0 . s_A^2$

where n_0 is a measure of average sample size.

Thus if we can calculate mean squares we can estimate s^2 and s_A^2. The mean squares are calculated from sums of squares.

Calculation of sums of squares

Consider Table 4, which gives hypothetical data for samples from three populations. Beneath each set are given n, Σx, Σx^2, and $(\Sigma x)^2$. To the right are given the same things for all the data taken together:

i) the total sample size. This is, of course, the sum of the three individual sample sizes. We may write it Σn.

ii) The grand sum of the data values (x values). Again, this is the sum of the three values of Σx. We may write it as $\Sigma\Sigma x$, to indicate that we have added up all the values of Sx.

iii) The grand sum of all the values of x^2. Again, this is the sum of the three values of Σx^2 and we may write it as $\Sigma\Sigma x^2$.

iv) The square of $\Sigma\Sigma x$. We write this as $(\Sigma\Sigma x)^2$. Note that this is not the same as the sum of the three values of $(\Sigma x)^2$, which we could write as $\Sigma(\Sigma x)^2$.

For a single sample, the sum of squares is
$$\Sigma(x - x)^2 = \Sigma x^2 - (\Sigma x)^2/n.$$
Similarly, for the combined data:

Sum of squares (total) = $\Sigma\Sigma x^2 - (\Sigma\Sigma x)^2/\Sigma n$
$$= 196 - 2704/16 = 27$$
This total SS is made up of a sum of squares between samples and a sum of squares within samples. These are:

SS (between) = $\Sigma[(\Sigma x)^2/n] - (\Sigma\Sigma x)^2/\Sigma n$
$$= (64/4 + 289/5 + 729/7) - 2704/16$$
$$= 8.94$$
SS (within) = SS (total) – SS (between) = 18.06

Calculation of mean squares

The mean squares may be calculated from the SS by dividing each by the number of degrees of freedom associated with it:

MS (between) = SS (between)/(k – 1)
MS (within) = SS (within)/(Σn – k)
where k is the number of samples.
In our case:

MS (between) = 8.94/(3 – 1) = 4.47
MS (within) = 18.06/(16 – 1) = 1.20

The anova table

It is helpful to lay out anovas in tabular form, starting with a table like Table 4 and then filling up an "anova table". Table 5 is such a table and summarises the formulae used so far.

Estimating the variance components

We learned earlier that:
MS (within) = s^2

MS (between) = $s^2 + n_0 \cdot s_A{}^2$
We now know the two MS, so the variance can be estimated if we know n_0. This is not an arithmetic mean of the sample sizes but is given by:
$$n_0 = [\Sigma n - (\Sigma n^2)/\Sigma n]/(k - 1)$$
In our example: $\Sigma n^2 = 4^2 + 5^2 + 7^2 = 90$
$$n_0 = [16 - 90/16]/2 = 5.19$$
Hence: s^2 = MS (within) = 1.20
$$s_A{}^2 = [\text{MS (between)} - s^2]/n_0$$
$$= (4.47 - 1.20)/5.19$$
$$= 0.63$$
We are usually interested in the relative size of these variances rather than their absolute sizes. These are commonly measured as:

% of variation between samples = $100 \times s_A{}^2/(s^2 + s_A^2)$
$$= 63/1.83$$
$$= 34\%$$
% of variation within samples = $100 \times s^2/(s^2 + s_A^2)$
$$= 120/1.83$$
$$= 66\%$$
Hence in our example, about two-thirds of the variation was due to differences between birds within populations, about one-third was due to differences between population means.

Confidence limits of the variance components

Unfortunately, statisticians do not seem to have developed a way of placing confidence limits on the estimated variance components, except in the special case where all the samples are of equal size.

Interpretation of the variance means (s_A^2)

The variance between means that we estimate is generally only applicable to the populations we have studied. Consider, for example, geographical variation of wing length in Dunhins (*Wader guide* p. 101). If we did an anova on samples from various parts of Europe, our estimate of the variance between population means would be fairly small. If we included samples from the whole breeding range it would be much larger. The first would be an estimate of the variance between European populations: the second would be an estimate of the variance between all populations.

We can only estimate the variance between all populations if we sample all populations or if we sample a random selection of all populations.

Table 5. Generalised anova table

Source of variation	Sum of squares (SS)	Degrees of freedom	Mean squares (MS)
Between samples	$\Sigma\left[\dfrac{(\Sigma x)^2}{n}\right] - \dfrac{(\Sigma\Sigma x)^2}{n}$	k – 1	$\dfrac{\text{SS (between)}}{(k - 1)}$
Within samples	SS (total) – SS (between)	Σn – k	$\dfrac{\text{SS (within)}}{(\Sigma k - 1)}$
Total	$\Sigma\Sigma x^2 - \dfrac{(\Sigma\Sigma x)^2}{\Sigma n}$	n – 1	

Thus we should always specify which populations have been studied before we quote a value of s_A^2.

Negative variances

The smaller a variance, the less variation there is. A variance of zero means that there is no variation – all individuals are identical. Thus a negative variance would mean less than no variation: negative variances are impossible.

For reasons connected with this, SS and MS values can never be negative – unless one has made an arithmetic error. However, it is possible to get a negative estimate of the between-population variance (s_A^2). This happens by chance and does not mean that such a negative variance is possible. In such a situation it is, of course, silly to say that the best estimate of the between-population variance is the calculated (negative) value: the best estimate is the closest possible value to that calculated – i.e. zero.

A warning: non-Normal samples

The calculations of confidence limits of means, the estimation of differences, and of between-population variance components all depend on the data in each sample being Normally distributed. They are unlikely to be seriously affected unless the data are markedly non-Normal but if in doubt take competent advice.

A warning: unequal variances

The estimation of confidence limits of differences and the anova assume that the variances of the different populations are the same. Space does not permit discussion of how large the difference between sample variances must be before we need to worry but in general, so long as one's samples are each larger than 50, one is safe if the largest variance is no more than twice as big as the smallest. If it is, take competent advice.

Introductory statistics 4

JEREMY J.D. GREENWOOD

Department of Biological Sciences, The University, Dundee, UK

Significance tests – at last!

So far, we have seen how to estimate means and standard deviations of populations, how to measure the precision of that estimate, and how to estimate the average difference between a whole set of populations by the analysis of variance. The latter may have seemed grossly unfamiliar to some readers, who may be asking why I have not yet covered more familiar ground such as significance tests. The reason is that I am convinced that the usefulness of significance tests has been greatly exaggerated. In many cases where such tests are applied in ornithology it would actually be more useful to carry out estimations and apply confidence limits.

Nonetheless, significance tests have their place. In this article I intend to explain that place and to show how certain tests may be performed.

The difference between two means

In Part 3, I considered what was the interpretation of a situation in which the confidence limits of the difference between two means included zero. We saw that this meant that one could not be sure which of the two populations had the greater mean: it might even be that the difference between them was zero. If we had no *a priori* reason for expecting a difference, we could not therefore disprove anyone's assertion that there was none.

In contrast, if the confidence limits did not include zero, we could be reasonably sure that there really was a difference between the two population means, basing our judgement on the difference between the sample means. We would say that the difference was statistically significant.

Whether a difference between two sample means is significant may be assessed without calculating confidence limits. If d is the estimated difference between the two means and sdiff is the standard error of the difference (calculated as in Part 3, page [19]), one calculates:

$$t_s = d/s_{diff}$$

If the confidence limits of the difference do not include zero, t_s will be greater than Student's t for $(n_1 + n_2 - 2)$ degrees of freedom. Thus, having calculated t_s we simply compare it with the t table to see if it is larger than the tabulated value. If it is, we conclude that the difference is significant.

If ts is less than the tabulated value of Student's t, the difference is "not significant" – i.e. we have no reason to reject the possibility that there is no difference between the population means. To put it another way, it is easily possible that the difference between the sample means has arisen by chance.

Levels of significance

If a test gives a significant result, it means that the 95% confidence limits do not include zero. To look at it the other way round, the probability that the true difference between the