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

Citation: Greenwood, J.J.D. 1979. Introductory statistics 4. *Wader Study Group Bull.* 27: 31–33.

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 population means takes any value outside these limits (including zero) is less than 5%. Thus if t_s is greater than the 95% value of Student's t for the relevant number of degrees of freedom, we say that "the difference is significant at the 5% level" or "$P < 0.05$".

If t_s is less than the 1% value of Student's t, then "the difference is significant at the 1% level" or "$P < 0.01$", and so on.

Clearly, the smaller the percentage level of significance (the more "highly significant", to use the usual jargon), then the less likely is it that the true difference between the population means is zero. For this reason, it is common to quote significance levels rather than just say that the test gave a significant result. The significance level tells us the degree of confidence we can have in the conclusion that one mean is bigger than the other: the more highly significant, the more confident we can be.

Note that the level of significance is not a measure of the size of the difference between the population means. The value of t_s depends not only on how large is the difference between the sample means but also on how small is the standard error of the differences.

The Null Hypothesis

I have written of the possibility that there is no difference between the population means. This possibility is known technically as the Null Hypothesis and it is this hypothesis that our procedure tests. If the test is significant, we have reason to reject the null hypothesis. If the test is not significant, we have no reason to reject the null hypothesis and we therefore accept it. We do accept it, not because we have shown that it is probably true, but because we have not shown that it is probably false. Clearly, therefore, it is only sensible to carry out a significance test when it is based on a null hypothesis that seems *a priori* reasonable. If our null hypothesis is *a priori* unreasonable, carrying out a test that turns out to be non-significant results in us having to accept, through "the rules of the game", this unreasonable hypothesis.

This last point may make one think that statistics runs counter to common-sense. But it is not statistics which is at fault here. What has happened is that statistics have been misused. One should avoid performing statistical tests that are based on unreasonable null hypothesis.

This is one reason for my assertion that the usefulness of statistical tests has been exaggerated. In many, perhaps most, situations where such tests are applied in ornithology the null hypothesis is in fact unreasonable. Before you carry out a test ask yourself: "Is it likely that there is *no* difference between the population means?" If it is unlikely, abandon the idea of a test and calculate the confidence limits of the difference. This is arithmetically just as easy and is more sensible.

Another advantage of confidence limits

Consider the following two estimates of differences between means:

1. + 0.03 cm, 95% C.L. – 0.01 to + 0.07 cm
2. + 5.93 cm, 95% C.L. – 4.64 to + 16.50 cm

Both pairs of confidence limits include zero: neither difference is significantly different from zero at the 5% level. Had we not estimated confidence limits but carried out statistical tests, this is all that we could have concluded. But the confidence limits tell us more. They tell us that the second difference could be zero or it could be quite large – anywhere between 4.5 cm one way and 16.5 cm the other. In contrast, even if the first difference is not zero it is unlikely to be very large. Thus confidence limits, measuring the precision of our estimates, tell one more than significance tests.

When are significance tests appropriate?

Significance tests are appropriate if one is genuinely uninterested in the magnitude of a difference but simply wishes to know whether one exists or not. This is rarely the case in ornithology.

They are also appropriate in situations where confidence limits cannot be calculated. This is the case, for example, in the analysis of variance. In such an analysis, we can estimate the variance components but usually cannot put confidence limits on the estimates. We can, however, test the null hypothesis that the variance between populations is zero.

Significance tests in the analysis of variance

In the analysis of variance we saw that the MS between groups is an estimate of

$$s^2 + n_o.s_A$$

where s^2 is the variance between individuals, s_A is the additive variance between groups, and n_o is a measure of average sample size. The MS within groups is an estimate of s^2 alone.

If the null hypothesis, that the variance between groups (s_A) is zero, is true, then the two MS will be more or less the same. They are unlikely to be identical because, although both are estimates of the variance between individuals, they are estimates based on slightly different information. Slight difference will occur by change, just as slight differences will occur between the means of two samples drawn from the same population.

If, in contrast, the null hypothesis is false – i.e. if there are real average differences between the populations, then MS (between) will be appreciably larger than MS (within). We can judge how large by using the variance-ratio, usually symbolised by F in honour of Sir Ronald Fisher, who invented the analysis of variance. We calculate the ratio:

$$F_s = \text{MS (between)/MS (within)}$$

and compare it with tabulated values of F. If it is larger, then we conclude that MS (between) is significantly greater than MS (within) – i.e. that there is a significant variance between groups (s_A).

As with Student's t, there are different values of F for different levels of significance. There are also different values for different numbers of degrees of freedom but here F is more complicated than t, for each value of F has a *pair* of degrees of freedom associated with it. The first of the pair is a number of degrees of freedom associated with MS (between): i.e. $(k - 1)$ in Table 5. The second of the pair is the number of degrees of freedom associated with MS (within): i.e. $(n - k)$ in Table 5.

Table 6 is the "top left" corner of an F table, to illustrate the usual format. I have included values for $(\Sigma n - k) = 13$ so that the use of the method can be illustrated with the example given in Part 3 (page [20] and Table 4). Here we found MS (between) = 4.47, MS (within) = 1.39, (k – 1) = 2, and (Σn – k) = 13. Hence:

$$F_s = \frac{4.47}{1.39} = 3.22, \text{ with 2 and 13 d.f.}$$

Consulting Table 6, we see that F for 2 and 13 d.f. is 3.81. Since our value is smaller than this, we have no reason to reject the null hypothesis and that $s_A - 0$. We have been unable to demonstrate a significant difference between groups.

It is, of course, true that the earlier part of the analysis of variance suggested that the variation between samples accounted for 34% of the total variation in this example. However, this figure was only an estimate of the true value. The variance ratio test tells us that the estimate is not significantly different from zero.

If we have more than two populations, with a sample from each, the analysis of variance allow us:

1. To estimate the percentage of the total variation that can be attributed to difference between populations (over and above differences between individuals).
2. To assess the statistical significance of the apparent differences between populations.

If the significance test is positive (F_s larger than tabulated F), then we can conclude that there are probably real differences between the populations. If it is not, then we can conclude that any differences between populations are too small to be demonstrable with reasonable sureness from the available data.

Prospect

The analysis of variance is an elegant and powerful statistical tool. It can be used for more complex analyses than I have shown here. For example, suppose we had samples from several locations, each divided into males and females. Differences in wing-length might arise from four basic sources:

Table 6. Partial table of F values for 5% significance level.

Second Degrees of Freedom (n–k)	First Degrees of Freedom (k–1)			
	1	2	3	4
1	161	299	216	225
2	18.5	19.0	19.2	19.3
3	10.1	9.55	9.28	9.12
4	7.71	6.94	6.59	6.39
.				
.				
13	4.67	3.81	3.41	3.18

1. Differences between individuals
2. Differences between sexes
3. Differences between locations
4. Differences between locations in the size of sex differences (or, to put it another way, differences between sexes in the size of locality differences)

All these could be estimated, and their significance tested, by the appropriate analysis of variance.

Even more complex analyses are possible. Their case depends very much on how the data are collected. It is always valuable, therefore, to consult a statistician before gathering the data. That way, one is less likely to amass a set of data that it is quite impossible to analyse – as happens all too often.

This series of articles has dealt with some basic statistical ideas and techniques. I have not dealt with the statistics of counts or with the examination of correlations. I hope, nonetheless, that the basic ideas presented have made it easier for readers to approach such matters. I hope also that they have shown that statistics is basically a matter of ornithological common sense and that the arithmetic involved is fairly trivial. What I intend to do in the next (and last) of the series is to discuss some of the traps into which the unwary often fall, so that the common sense and ability to use the formulae will be backed up with a sufficient degree of caution.

Introductory statistics 5

JEREMY J.D. GREENWOOD

Department of Biological Sciences, The University, Dundee, UK

Citation: Greenwood, J.J.D. 1980. Introductory statistics 5. *Wader Study Group Bull.* 28: 25–26.

Planning your research project

Some years ago, an eminent research director wrote "As a statistician, I still have to spend far too much of my time trying to find ways of making use of data for which the methods of collection were inefficient, if not invalid . . . Papers are still being published in reputable journals where the methods of data collection and statistical analysis are deplorable if not impossible . . . We still encounter the post-graduate degree student in despair, because he has spent two and a half years collecting his data, and he now finds that even his supervisor has very little idea about how these data can be analysed or presented, or even whether the method of data collection is relevant to the problem he is undertaking."

The position is no better today. What is worse, it is scarcely any better among professionals than among amateurs.

There are many reasons for this, few of which place those