



Stat Note 5

In the fifth of a series of articles about statistics for biologists, **Anthony Hilton** and **Richard Armstrong** ask:
is one set of data more variable than another?

THERE MAY BE occasions when it is necessary to test whether the variability of two or more sets of data differ.

An investigator, for example, may wish to test whether a new treatment reduces the variability of a particular microbial response compared with an older treatment. In addition, an

important assumption for the use of the 't' test (Hilton & Armstrong, 2005) or analysis of variance (ANOVA) (Armstrong & Hilton, 2004) is that the variability of the different groups being compared is similar, i.e., that they exhibit *homogeneity of variance*. Replicate measurements within a control and a treated group, however, often exhibit different degrees

of variation and the assumption of *homogeneity of variance* may need to be explicitly tested. This Statnote describes four such tests, *viz.*, the variance-ratio (F) test, Bartlett's test, Levene's test, and Brown and Forsythe's test.

The scenario

We return to the scenario first described in Statnote 3

(Hilton & Armstrong, 2005). A hypothetical experiment was carried out to investigate the efficacy of two novel media supplements (S1 and S2) in promoting the development of cell biomass. Three ten-litre fermentation vessels were sterilised and filled with identical growth media with the exception that the media in two of the vessels was supplemented with ten ml of

either medium supplement S1 or S2. The vessels were allowed to equilibrate and were subject to identical environmental / incubation conditions. The vessels were then inoculated with a culture of *Bacterium x* at an equal culture density and the fermentation allowed to proceed until all the available nutrients had been exhausted and bacterial growth had ceased. The entire volume of culture media in each fermentation vessel was then removed and filtered to recover the bacterial biomass, which was subsequently dried and the dry weight of cells measured. This experiment was repeated 25 times and the dry weight of biomass produced in each of the three groups recorded in Table 1.

The variance-ratio test

If there are only two groups involved, then their variances can be compared by a two-tail variance ratio test (F-test) (Snedecor & Cochran, 1980).

How is the test done?

The larger variance is divided by the smaller and the resulting F ratio compared with the value in a table of the variance ratio to obtain a P-value, entering the table for the number of degrees of freedom (DF) of the numerator and denominator. This test uses the *two-tail probabilities* of F because we are testing whether or not the two variances *differ* rather than whether variance A is greater than variance B. Hence, this calculation differs from that carried out during a typical ANOVA, since in the latter, it is whether the treatment variance is *larger than* the error variance that is being tested (Armstrong & Hilton, 2004). Published statistical tables of the F ratio (Fisher & Yates, 1963; Snedecor & Cochran, 1980) are usually in the form of one-tail tables. Hence, the 2.5%

Table 1. Dry weight of bacterial biomass under unsupplemented (US) and two supplemented (S) growth conditions (S1 and S2) in a sample of 25 fermentation vessels.

US	S1	S2	US	S1	S2	US	S1	S2
461	562	354	506	607	556	518	617	714
472	573	359	502	600	578	527	622	721
473	574	369	501	603	604	524	626	722
481	581	403	505	605	623	529	628	735
482	582	425	508	607	644	537	631	754
482	586	476	500	609	668	535	637	759
494	591	511	513	611	678	542	645	765
493	592	513	512	611	698			
495	592	534	511	615	703			

Variances: US = 463.36. S1 = 447.88. S2 = 18695.24

Variance-ratio test comparing US and S1: $F = 463.36/447.88 = 1.03$
(2-tail distribution of F, $P > 0.05$)

probability column has to be used to obtain the 5% probability.

Interpretation of the results

When the unsupplemented and S1 data are compared (Table 1), a value of $F = 1.03$ was obtained. This value is less than the F value in the 2.5% column ($P > 0.05$) and consequently, there is no evidence that the addition of the medium S1 increased or decreased the variance in replicate flasks.

Bartlett's test

If there are three or more groups, then the different groups could be tested in pairs using the F-test

described above, but a better approach is to test all the variances simultaneously using Bartlett's test (Snedecor & Cochran, 1980).

How is the test done?

If there are equal numbers of observations in each group, calculation of the test statistic is straight-forward and a worked example is shown in Table 2. If the three variances do not differ from each other, then the ratio M/C is a member of the chi-square (χ^2) distribution with $(a - 1)$ degrees of freedom (DF), where 'a' is the number of groups being compared. If the groups have different numbers of observations in each (unequal 'n'), then the

Table 2. Comparison of the variances of three groups with equal observations ($v = 25$) in each by Bartlett's test.

Group	Variance	ln (variance)
Unsupplemented	436.36	6.1385
S1	447.88	6.1045
S2	18695.24	9.8360
Total	19606.48	22.079

$M = \sqrt{a} (\ln s^{*2} - \sum \ln s_i^2)$ where s^{*2} is the mean of the variances, 'a' the number of groups, v = DF of each group, and \ln = logarithms to base e.
Hence, $M = 102.62$
 $C = 1 + (a + 1)/(3av) = 1.018$
 $\chi^2 = M/C = 102.62/1.018 = 100.8$ (DF = $a - 1$, $P < 0.001$)

calculations are slightly more complex and are given in Snedecor and Cochran (1980).

Interpretation of the results

In the worked example in Table 2, the value of χ^2 was highly significant ($P < 0.001$) suggesting real differences between the variances of the three groups. The previous F-test suggested, however, that the variance of the unsupplemented data was similar to that of the growth medium S1. Therefore, it is the effect of the growth medium S2 that has substantially increased the variance of bacterial biomass. Hence, if these data were to be analysed by ANOVA (Armstrong & Hilton, 2004), the assumption of homogeneity of variance would not hold and it may be necessary to transform the data to logarithms before analysis to stabilize the variance. Data transformation is described in more detail in Statnote 4 (Hilton & Armstrong, 2006).

The use of the χ^2 distribution to test the significance of M/C is questionable if the DF within the groups are less than five and in such a case, there are special tables for calculating the significance of the statistic (Pearson & Hartley, 1954). Bartlett's test is used less today and may not normally be available as part of a statistics software package. This is because the test is regarded as being too 'sensitive' resulting in too many significant results especially with data from long-tailed distributions (Snedecor & Cochran, 1980). Hence use of the test may raise unjustified concerns about whether the data conform to the assumption of homogeneity of variance. As a consequence, Levene (1960) developed a more robust test to compare three or more

variances (Snedecor & Cochran, 1980).

Levene's test. How is the test done?

Levene's test makes use of the absolute deviation of the individual measurements from their group means rather than the variance to measure the variability within a group. Avoiding the squaring of deviations as in the calculation of variance results in a measure of variability that is less sensitive to the presence of a long-tailed distribution. An ANOVA (Armstrong & Hilton, 2004) is then performed on the absolute deviations and if significant, the hypothesis of homogeneous variances is rejected.

Interpretation of the data

A Levene's test on the data in Table 1 using STATISTICA software, for example, gave a value of $F = 52.86$ (DF 2,72; $P < 0.001$) confirming the results of Bartlett's test.

More recently, Levene's test has also been called into question since the absolute deviations from the group means are likely to be highly skewed and therefore, violate another assumption required

for an ANOVA, that of normality (Armstrong and Hilton, 2004). This problem becomes particularly acute if there are unequal numbers of observations in the various groups being compared. As a consequence, a modification of the Levene test has been proposed by Brown and Forsythe (1974).

Brown-Forsythe test. How is the test done?

This differs from Levene's test in that an ANOVA is performed not on the absolute deviations from the group means but on deviations from the group medians. This test may be more accurate than Levene's test even when the data deviate from a normal distribution. Nevertheless, both Levene's and the Brown-Forsythe tests suffer from the same defect in that to assess differences in variance requires an ANOVA, and an ANOVA requires the assumption of 'homogeneity of variance,' which some authors consider to be a 'fatal flaw' of these analyses.

Conclusion

There may be circumstances where it is necessary for microbiologists to compare variances rather

than means, e.g., in analysing data from experiments to determine whether a particular treatment alters the degree of variability or testing the assumption of homogeneity of variance prior to other statistical tests.

All of the tests described in this Statnote have their limitations. Bartlett's test may be too sensitive but Levene's and the Brown-Forsythe tests also have problems. We would recommend the use of the variance-ratio test to compare two variances and the careful application of Bartlett's test if there are more than two groups.

Considering that these tests are not particularly robust, it should be remembered that the homogeneity of variance assumption is usually the least important of those considered when carrying out an ANOVA.

If there is concern about this assumption and especially if the other assumptions of the analysis are also not likely to be met, e.g., lack of normality or non additivity of treatment effects (Armstrong & Hilton, 2004) then it may be better either to transform the data or to carry out a non-parametric test on the data.

References

- Armstrong RA & Hilton A (2004) The use of analysis of variance (ANOVA) in applied microbiology. *Microbiologist*, vol 5: **No.4** 18.
- Brown MB & Forsythe AB (1974) Robust tests for the equality of variances. *J Am Stats Assoc* 69: 264-267.
- Fisher RA & Yates F (1963) Statistical tables. Longman, London.
- Hilton A & Armstrong RA (2005) Statnote 3: Comparing the difference between two groups. *Microbiologist*, vol 6: **No.4** 30.
- Hilton A & Armstrong RA (2006) Statnote 4: What if the data are not normal? *Microbiologist*, vol 7: **No.1** 34.
- Levene H (1960) In: *Contributions to Probability and Statistics*. Stanford University Press, Stanford, California.
- Pearson ES & Hartley HO (1954) *Biometrika Tables for Statisticians*, vol1. Cambridge University Press.
- Snedecor G W & Cochran W G (1980) *Statistical Methods*, 7th Ed. Iowa State University Press, Ames Iowa.

Dr Anthony* Hilton and Dr Richard Armstrong**

*Pharmaceutical Sciences and
**Vision Sciences, Aston University, Birmingham, UK

Instant access!

Did you know that previous Stat Notes are available for download from the website in Adobe Acrobat PDF format?

Simply click the articles you wish to view and/or right click a link to save a copy of the PDF to your hard disk.



Simply visit: <http://www.sfam.org.uk/features.php>