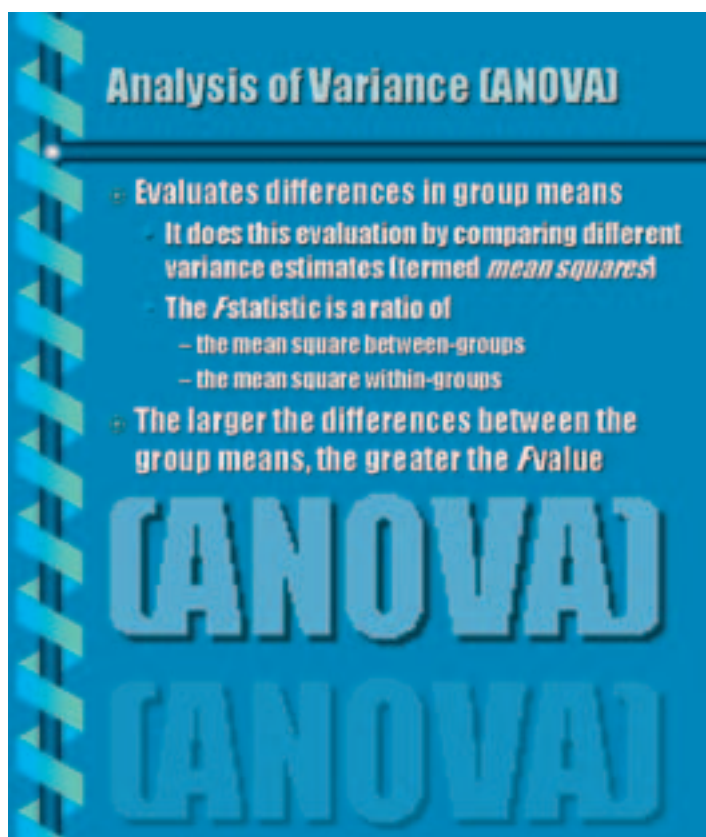


# The use of Analysis of Variance (ANOVA) in applied microbiology

Richard Armstrong and Anthony Hilton discuss the uses of this important tool



**S**TUDIES IN applied microbiology often involve comparing either several different treatments or the influence of two or more factors at the same time.

For example, an investigator may be interested in the degree of microbial contamination on coins collected from three different premises and the data analysis might involve a comparison of microbial numbers from the three different locations. In another example, one may wish to compare the pattern of transfer of bacteria from different dishcloths, either rinsed or not, onto a food preparation surface such as a cutting board (Hilton & Austin 2000). In this case, two factors may influence

microbial numbers, *viz.*, the type of dishcloth and rinsing treatment. An investigator may wish to establish whether both factors influenced microbial numbers independently or whether there was an interaction between them, *i.e.*, does rinsing have the same effect on the numbers of bacteria transferred when different dishcloths are used? The most appropriate method of statistical analysis of such experiments is analysis of variance (ANOVA) (Snedecor & Cochran 1980, Armstrong *et al.*, 2000, 2002).

Analysis of variance is the most effective method of analysing more complex data sets. It is, however, a method comprising many different variations, each of which apply in a particular

experimental context. Hence, it is possible to apply the wrong type of ANOVA and to draw erroneous conclusions from the results. This article is an introduction to ANOVA and describes first, how ANOVA came to be invented, the logic on which it is based, and the basic assumptions necessary for its correct use. Second, the application of the method to the analysis of three data sets drawn from experiments in applied microbiology is described.

## The origin of ANOVA

An experiment was set up to measure the degree of bacterial contamination on 2p coins collected from three different premises, *viz.*, a butcher's shop, a sandwich shop, and a newsagent. A sample of four coins was collected at random from each premises. The number of

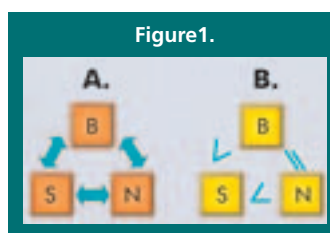
(Armstrong & Eperjesi, 2001). A significant value of Student's 't' indicates that the null hypothesis should be rejected and that there is a real difference in bacterial numbers between the two premises.

This method of analysis could be extended to three or more different premises. To compare all pairs of premises, three t-tests would be necessary (Fig 1). There is a problem, however, in making these tests because not all of the comparisons are independent. For example, if bacterial numbers from the butcher's shop were significantly greater than those from the sandwich shop but similar to those from the newsagent, it follows that numbers in the newsagent should be greater than those in the sandwich shop (Fig 1).

However, the latter comparison would not have been tested independently but follows inevitably from the first two comparisons. To overcome this problem, ANOVA was developed by Sir Ronald Fisher in the 1920s and provides a single statistical test of the null hypothesis that the means of the bacterial numbers on coins from the three premises are identical.

## The logic of ANOVA

The data from the experiment described above are shown in Table 1. Each of the three premises is represented by measurements of bacterial contamination on four coins and therefore, the experiment is described as a one-way ANOVA with four replications in a randomized design. In an ANOVA, the total



bacterial colonies present on each coin was estimated by dilution plating. If only two types of premises were involved then the null hypothesis that there is no significant mean difference in the numbers of bacteria at the two locations could be tested using Student's 't' test (Snedecor & Cochran, 1980). The statistic 't' is the ratio of the difference between the means and a measurement of the variation between the counts on the individual coins pooled from both groups

variation between the observations ( $x_{ij}$ ) is calculated and then partitioned into portions associated with differences between the three premises and the variation between the replicate coins within premises. The calculations involved are shown in Table 1. The sum of squares of the deviations of the  $x_{ij}$  from their mean ( $X^*$ ) is used as a measure of total variation while the sum of squares of the three premises means from their overall mean is a measure of the treatment effect and is calculated from the column totals ( $T_i$ ). Variation between the coins within each location (residual or error variation) is calculated as the sum of squares of the  $x_{ij}$  in each column from their column mean and then added together to give the error sum of squares or can be obtained by subtraction.

If there are no significant differences between the means of the three premises, the twelve observations are distributed about a common population mean ' $\mu$ '. If this is the case, then the variance (also called the mean square) calculated from the between premises sum of squares and the error sum of squares should be estimates of the same quantity. Testing the difference between these two mean squares is the basis of an ANOVA and the statistics are set out in an ANOVA table (Table 1). To compare the between premises and error mean squares, the sums of squares are divided by their appropriate degrees of freedom (DF). The DF of a quantity is the number of observations minus the number of parameters estimated from the data required to calculate that quantity. Hence, the total and between premises sum of squares each have eleven and two DF respectively, one less than the number of


observations or groups. This is because the mean of the ' $x_{ij}$ ' values and the mean of the three group totals were calculated from the data to obtain the sums of squares. The error sum of squares has 9 DF because the column means are used in the calculation, i.e., there are three DF in each of the three columns making nine in total.

The between premises mean square is then divided by the error mean square to obtain the variance ratio. This statistic was named 'F' (hence, 'F-test') in honour of Fisher by G.W. Snedecor (Snedecor & Cochran, 1980). The value of 'F' indicates the number of times the between premises mean square exceeds that of the error mean square. The probability of obtaining a statistic of this magnitude by chance, i.e., from data with no significant differences between the group means, is obtained from the F distribution. If the value of 'F' is equal to or greater than the value at the 5% level of

probability, then the null hypothesis that the three premises means are identical is rejected. In this case, a value of  $F = 4.89$  was obtained which has a P value of 0.037, i.e., there is less than a 5% of chance of obtaining an F ratio of this magnitude by chance. This result indicates a real difference between the bacterial counts from the three premises. Note that this analysis relates only to the three premises studied. It would not be possible to make a general statement about all premises of this type from these data. This would require a random sample of each premises to be obtained so that an estimate could be made of the variation in bacterial counts between similar premises.

### Comparison of group means

The F-test of the group means is only the first stage of the data analysis. The next step involves a more detailed

examination of the differences between the means. In many circumstances, this involves making *post-hoc* tests between the group means. The subject of *post-hoc* tests a complex topic and a variety of methods are usually available for making such tests in statistical packages. The most commonly used tests and the assumptions necessary for their correct application have been discussed in detail previously (Armstrong *et al.*, 2000) and only two examples, *viz.*, Fisher's protected least significant difference (PLSD) and Scheffe's test, will be considered in this article. The results of the *post-hoc* tests applied to the present data are shown in Table 1. These tests vary in the degree of error control they provide and in particular their sensitivity to making a 'type one' error, i.e., rejecting a null hypothesis when it is true. By contrast, a type 2 error is accepting the null hypothesis when a real difference is present. Fisher's PLSD is the most 'liberal' of the methods available and the most likely to result in a Type 1 error. All possible comparisons of the group means are tested and the method uses the t-distribution to determine the critical value to be exceeded for any pair of means. This test indicates that bacterial counts were significantly higher on 2p coins from the butcher's compared with the sandwich shop and newsagents but that counts from the later two premises were similar. By contrast, Scheffe's test is one of the most conservative of the *post-hoc* tests giving maximum protection against making a type 1 error but increasing the probability of making a type 2 error. In this case, the test does not indicate any significant differences between the group means. Which test should be used in each circumstance depends on the objectives of the 

**TABLE 1.** The number of bacteria isolated from 2p coins collected from three types of premises

Replicates	Butcher's	Sandwich Shop	Newsagent
1	140	2	40
2	108	21	5
3	76	0	5
4	400	42	0
Mean	181	16	13
SE	74.16	9.80	9.24

Total sum of squares (SS) =  $\sum(x_{ij} - X^*)^2 = 142238.917$ , Between groups (premises) SS =  $\sum(\sum T_i - H^*)^2 = 74065.167$ , Error SS = Total SS - Between groups SS = 68173.75

#### ANOVA table

Source	DF	SS	MS	F	P
Between Premises	2	74065.167	37032.583	4.89	0.037
Error	9	68173.75	7574.861		

#### Post-hoc tests

Comparison	Mean diff.	Fishers PLSD	Scheffe's Test
Butcher's/Sandwich	164.75	P < 0.05	NS
Butcher's/Newsagent	168.5	P < 0.05	NS
Sandwich/Newsagent	3.75	NS	NS

**Abbreviations:** SE = Standard error of the mean,  $x_{ij}$  = Individual counts,  $X^*$  = overall mean of all 12 bacterial counts,  $T_i$  = Column total,  $H^*$  = mean of the three column totals, DF = degrees of freedom, MS = Mean square, F = variance ratio, P = probability, NS = not significant

experiment and on the relative consequences or costs of making a Type 1 or Type 2 error.

### Assumptions of ANOVA

ANOVA makes certain assumptions about the nature of the experimental data that have to be at least approximately true before the method can be validly applied. An observed value  $x_{ij}$  can be considered to be the sum of three parts: 1) the overall mean of the observations ( $\mu$ ), 2) a treatment or class deviation, and 3) a random element drawn from a normally distributed population. The random element reflects the combined effects of natural variation between subjects and errors of measurement. ANOVA assumes first, that these errors are normally distributed with a zero mean and standard deviation 's', second, that although the means may vary from group to group, the variance is constant in all groups, and that effects of individual treatments are additive and not multiplicative. Failure of an assumption affects both the significance levels and the sensitivity of the F-tests. Experiments are usually too small to test whether these assumptions are likely to be true. In many biological and medical applications, in which a quantity is being measured, however, the assumptions hold well (Cochran & Cox, 1957; Ridgeman, 1975). In many applications in applied microbiology in which bacterial numbers are being estimated, the assumptions may not hold. There are two common problems when the data comprise numbers of microbes. First, small whole numbers with many zeros are unlikely to be normally distributed and second, a wide range of bacterial numbers may be present leading to heterogeneous variances. The

latter problem is evident in the example analysed in Table 1 where the standard errors for the three premises vary markedly. If there is doubt about the validity of the assumptions, significance levels and confidence limits must be considered to be approximate rather than exact. If there is only a single significant figure, the assumptions are more doubtful and if the data are small whole numbers, then the assumptions are unlikely to hold. If the assumptions do not hold, then a transformation of the  $x_{ij}$  into

effects of a number of different factors can be studied at the same time. Combining factors usually requires fewer replications than studying each factor individually in a separate experiment. In addition, the analysis reveals the possible synergistic or interactive effects between the factors and is often the most interesting information from a factorial experiment.

To illustrate the analysis, an investigator wished to study the influence of type of dishcloth (cloth or sponge) and prior rinsing of the

varied both with dishcloth and rinsing and whether the two factors had an independent influence on numbers of bacteria.

The resulting ANOVA (Table 2) is more complex than that of a one-way ANOVA because the between groups or treatments sums of squares is partitioned into three factorial effects, *viz.*, the main effects of dishcloth type and rinsing and the interaction between the two factors. In the present example, there is a main effect of dishcloth type ( $F = 20.99$ ,  $P < 0.01$ ) and of rinsing ( $F = 28.92$ ,  $P < 0.001$ ) suggesting significantly more bacteria were transferred from the cloth than the sponge and significantly fewer after rinsing both materials. In addition, there is significant interaction between the factors ( $F = 20.97$ ,  $P < 0.01$ ) suggesting that the effect of rinsing is not the same for the two types of dishcloth.

Examination of the treatment means suggests that the sponge transfers a smaller proportion of its bacterial load to the food preparation surface compared with the cloth. This effect is probably attributable to organisms being more exposed on the surface of the cloth and therefore more liable to be transferred compared with the more cavernous sponge (Hilton & Austin, 2000).

### A more complex factorial experiment

An investigator wished to examine the pattern of survival of bacteria on £5 notes. The data comprise numbers of bacteria of two species, *viz.*, *E. coli* and *S. epidermidis*, inoculated on to the surface of £5 notes and subsequently measured at ten time intervals (Table 3). Survival of the two bacteria on cover-slips was examined as a control. The objectives of the experiment were first, to determine whether there was a

**Table 2.** Influence of type of dishcloth and rinsing treatment on the number of bacteria transferred to a food preparation surface

Replicates	Cloth		Sponge	
	Rinsed	Not rinsed	Rinsed	Not rinsed
1	$1.0 \times 10^5$	$7.8 \times 10^7$	$3.9 \times 10^5$	$8.0 \times 10^5$
2	$2.3 \times 10^4$	$5.0 \times 10^7$	$9.0 \times 10^3$	$4.0 \times 10^5$
3	$3.9 \times 10^5$	$4.1 \times 10^7$	$8.5 \times 10^4$	$2.0 \times 10^5$
Mean	$1.7 \times 10^5$	$5.6 \times 10^7$	$1.6 \times 10^5$	$4.7 \times 10^5$

#### ANOVA table

Source	DF	SS	MS	F	P
Material	1	2002.83	2002.83	20.99	$P < 0.01$
Rinsing	1	2760.42	2760.42	28.92	$P < 0.001$
Interaction	1	2001.33	2001.33	20.97	$P < 0.01$
Error	8	763.49	95.44		

**Abbreviations:** SE = Standard error of the mean,  $x_{ij}$  = Individual counts,  $X^*$  = overall mean of all 12 bacterial counts,  $T_i$  = Column total,  $H^*$  = mean of the three column totals, DF = degrees of freedom, MS = Mean square, F = variance ratio, P = probability, NS = not significant

another scale will often allow an ANOVA to be carried out. For example, in Table 1, a transformation of the data to a logarithmic scale is likely to equalize the variances for the three premises. For a more detailed discussion of the use of transformations see Snedecor & Cochran (1980) and Armstrong & Eperjesi (2001).

### Factorial experiments

The ANOVA described above is an example of a single factor experiment, the variable involved being type of premises. In a factorial experiment, however, the

material on the number of bacteria transferred to a food preparation surface (Hilton & Austin, 2000). Each dishcloth was inoculated with 1 ml of a  $10^8$  cfu ml<sup>-1</sup> *E. coli* culture and after 10 min, the cloth was wiped over an appropriate area of cutting board. Additional pieces of cloth were inoculated but rinsed in sterile running water before wiping. The data comprise the number of bacterial colonies obtained on nutrient agar from two types of dishcloth, rinsed and unrinsed, and are shown in Table 2. The objectives of the experiment were to determine whether bacterial loading

difference in survival of bacteria on control surfaces as against £5 notes and second, to determine whether the two species exhibit different survivorship curves on the various surfaces.

This experiment is considerably more complex than the previous examples. First, three factors are involved, *viz.*, bacterial strain, type of surface, and time interval and this results in a 'three-factor' factorial. Second, the fact that repeated measurements of bacterial numbers are made on each surface leads to a 'repeated measures' design (Armstrong *et al.*, 2002). In the ANOVA, there are two major factors (bacteria and type of surface) with time constituting the repeated measures factor. A common mistake made by investigators is to analyse such an experiment as if it comprises three completely independent factors.

The ANOVA appropriate to these data is shown in Table 3. There are significant main effects of bacterial strain ( $F = 140.81$ ,  $P < 0.001$ ) and type of surface ( $F = 111.48$ ,  $P < 0.001$ ) which suggests first, greater numbers of *S. epidermidis* at the beginning of the experiment and second, greater numbers of bacteria on the £5 notes compared with cover-slips. There is an interaction, which is only just significant, between type of surface and bacteria ( $F = 8.13$ ,  $P < 0.05$ ) indicating that differences between surfaces varied with species. The main effect of time ( $F = 5707.7$ ,  $P < 0.001$ ) reflects the general decline in numbers over the period of the experiment. This decline, however, varies with type of surface ( $F = 47.33$ ,  $P < 0.001$ ), a more rapid and pronounced decline in numbers being observed on cover-slips compared with £5 notes and with species ( $F = 168.7$ ,  $P < 0.001$ ); a much

**TABLE 3.** The survival of two strains of bacteria on two different surfaces (EC = *Escherichia coli*; SE = *Staphylococcus epidermidis*). Each figure is the mean of two replicates

Time (hrs)	Notes		Control		Notes		Control	
	EC	EC	EC	EC	SE	SE	SE	SE
0	3800	3800	3800	3800	5400	5400	5400	5400
1	800	430	0	0	600	800	7	22
3	500	351	6	0	560	560	6	10
5.5	446	249	1	0	700	764	0	2
24	0	1	0	0	272	171	0	0
27	0	0	0	0	54	2	0	0
30	0	0	0	0	79	42	0	0
48	0	0	0	0	124	105	0	0
51	0	0	0	0	7	14	0	0
54.5	0	0	0	0	2	10	0	0

Analysis of variance					
Source	DF	SS	MS	F	P
Bacteria	1	909298	909298	140.81	P < 0.001
Control/notes	1	719911	719911	111.48	P < 0.001
Bacteria x control/notes	1	52480	52480	8.127	P < 0.05
Main-plot error	4	25480	6457.6		
Time	9	146528276	16280919	5707.7	P < 0.001
Time x	9	4332765	481418	168.7	P < 0.001
Bacteria Time x	9	1215098	135011	47.33	P < 0.001
Control/notes					
3-factor interaction	9	65389	7265	2.54	P < 0.5
Sub-plot error	36	102687	2852		

**Abbreviations:** SE = Standard error of the mean,  $x_{ij}$  = Individual counts,  $X^*$  = overall mean of all 12 bacterial counts,  $T_i$  = Column total,  $H^*$  = mean of the three column totals, DF = degrees of freedom, MS = Mean square, F = variance ratio, P = probability, NS = not significant

more marked decline being observed in *E. coli* compared with *S. epidermidis*. There is also a significant three-factor interaction, but higher order interactions are usually too complex to interpret easily.

## Conclusion

Experiments combining different groups or factors and which use ANOVA are a powerful method of investigation in applied microbiology. ANOVA enables not only the effect of individual factors to be estimated but also their interactions; information which cannot be obtained readily when factors are investigated separately. In addition, combining different

treatments or factors in a single experiment is more efficient and often reduces the number of replications required to estimate treatment effects accurately. Because of the treatment combinations used in a factorial experiment, the DF of the error term in the ANOVA is a more important indicator of the 'power' of the experiment than the number of replicates (Ridgman, 1975; Armstrong *et al.*, 2002). Finally, it is important to consider the design of the experiment because this determines the appropriate ANOVA. Some of the most common experimental designs used in the biosciences and their relevant ANOVAs are discussed in detail by Armstrong *et al.*, (2002). If there is doubt about the design or which ANOVA to use, the researcher should seek advice from a statistician with experience of research in applied microbiology. Once committed to an inappropriate experimental design there may be little that a statistician can do to help.

**A C Hilton and  
R A Armstrong**  
Aston University, Birmingham

## References

- Armstrong R A, Slade S V and Eperjesi F (2000) An introduction to analysis of variance (ANOVA) with special reference to data from clinical experiments in optometry. *Ophthal Physiol Opt* **20**: 235-241.
- Armstrong R A and Eperjesi F (2001) Data analysis methods in optometry: Is there a difference between two samples? *Optometry Today*, Jan **26**: 27-31.
- Armstrong R A, Eperjesi F and Gilmartin B (2002) The application of analysis of variance (ANOVA) to different experimental designs in optometry. *Ophthal Physiol Opt* **22**: 248-256.
- Cochran, W G and Cox, G M (1957) *Experimental designs*. Second Ed., John Wiley, New York, London and Sydney.
- Hilton, A C. and Austin, E (2000) The kitchen dishcloth as a source of and vehicle for foodborne pathogens in a domestic setting. *Inter J Env Health Res* **10**: 257-261.
- Ridgman, W J (1975) *Experimentation in Biology*. Blackie, London.
- Snedecor, G W and Cochran, W G. (1980) *Statistical methods*. 7th edition, Iowa State University Press, Ames, Iowa.