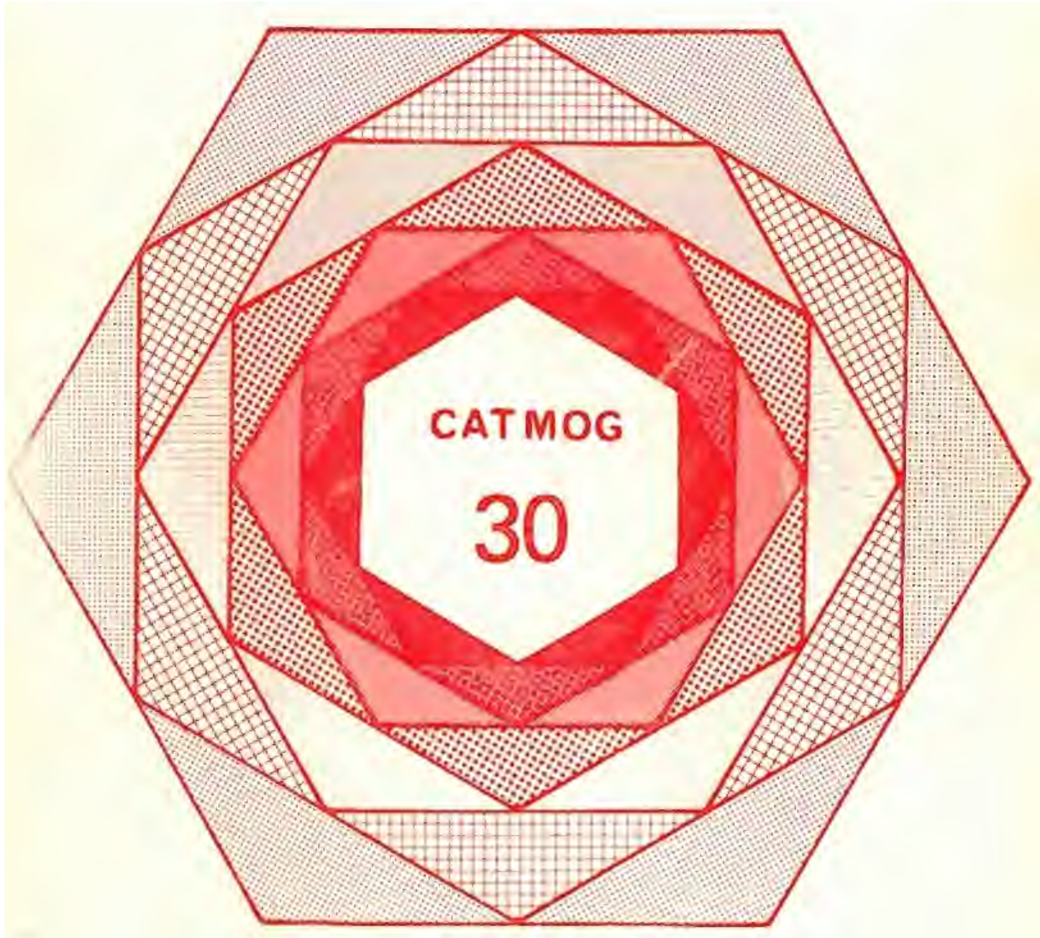
ISSN 0306 - 6142

ISBN 0 86094 011 3

Printed in Great Britain by Headley Brothers Ltd The Invicta Press Ashford Kent

THE ANALYSIS OF VARIANCE

John Silk



CATMOG

(Concepts and Techniques in Modern Geography)

CATMOG has been created to fill a teaching need in the field of quantitative methods in undergraduate geography courses. These texts are admirable guides for the teachers, yet cheap enough for student purchase as the basis of classwork. Each book is written by an author currently working with the technique or concept he describes.

1. An introduction to Markov chain analysis - L. Collins
2. Distance decay in spatial interactions - P.J. Taylor
3. Understanding canonical correlation analysis - D. Clark
4. Some theoretical and applied aspects of spatial interaction shopping models - S. Openshaw
5. An introduction to trend surface analysis - D. Unwin
6. Classification in geography - R.J. Johnston
7. An introduction to factor analytical techniques - J.B. Goddard & A.Kirby
8. Principal components analysis - S. Daultrey
9. Causal inferences from dichotomous variables - N. Davidson
10. Introduction to the use of logit models in geography - N. Wrigley
11. Linear programming: elementary geographical applications of the transportation problem - A. Hay
12. An introduction to quadrat analysis - R.W. Thomas
13. An introduction to time-geography - N.J. Thrift
14. An introduction to graph theoretical methods in geography - K.J. Tinkler
15. Linear regression in geography - R. Ferguson
16. Probability surface mapping. An introduction with examples and Fortran programs - N. Wrigley
17. Sampling methods for geographical research - C. Dixon & B. Leach
18. Questionnaires and interviews in geographical research - C. Dixon & B. Leach
19. Analysis of frequency distributions - V. Gardiner & G. Gardiner
20. Analysis of covariance and comparison of regression lines - J. Silk
21. An introduction to the use of simultaneous-equation regression analysis in geography - D. Todd
22. Transfer function modelling: relationship between time series variables - Pong-wai Lai
23. Stochastic processes in one-dimensional series: an introduction - K.S. Richards
24. Linear programming: the Simplex method with geographical applications - J.E. Killen

THE ANALYSIS OF VARIANCE

by

John Silk

(University of Reading)

CONTENTS

		Page
I	INTRODUCTION	3
	(i) Prerequisites and purpose	3
	(ii) Relationships between variables	3
II	RESEARCH DESIGN	5
	(i) General considerations	5
	(ii) Case study: investigation of mean valley-side slope angles in Wyoming	9
III	ONE-WAY ANALYSIS OF VARIANCE	11
	(i) The Analysis of Variance model and its assumptions	11
	(ii) Hypothesis testing	13
	1) Overall test of significance	13
	2) Comparisons of means	17
	3) Interval Estimation	20
	4) Transformations	22
IV	TWO-WAY ANALYSIS OF VARIANCE	23
	(i) Notation and terminology	23
	(ii) Carrying out the analysis	24
	1) Model and assumptions	24
	2) Application to valley-side slope angle data	25
	3) Additivity and interaction	27
	4) Testing for interaction	32
	(iii) A more complex example: a study of shopping knowledge in Oxford	34
	(iv) Two-way Analysis of Variance with one observation per cell	37
	(v) Two-way Analysis of Variance with unequal numbers of observations per cell	38
	(vi) Limitations	38

V RANDOM-EFFECTS OR COMPONENTS OF VARIANCE MODELS	39
VI SPATIAL MODELLING	40
(i) Simple trend surface construction	40
(ii) Components of variance models	40
1) Description of spatial hierarchies	41
2) The components of variance model	43
3) Estimation of variance components	44
4) Hypothesis testing	49
5) Limitations	49
VII APPLICATIONS AND FURTHER READING	51
VIII EXTENSIONS, AND CONNECTIONS WITH OTHER TECHNIQUES	52
IX CONCLUSION	53
BIBLIOGRAPHY	54

ACKNOWLEDGEMENTS

To Sophie Bowlby for making available data from her study of shopping behaviour in Oxford, and to Derek Pike for his much-valued advice and time in discussing technical matters; many thanks to Chris Holland for typing the text and Brian Rogers for drawing the diagrams.

I INTRODUCTION

(i) Prerequisites and purpose

The *Analysis of Variance* (ANOVA) is a highly developed branch of mathematical statistics, which is widely used throughout the natural and social sciences but which is particularly associated with carefully controlled experimental work in agriculture or psychology. As a number of commentators have pointed out (e.g. Johnston (1978)) it has been used relatively infrequently by geographers. However, there appears to be considerable scope for use of the technique in physical geography - geologists, for instance, have made frequent use of it - as well as in the analysis of survey research data and of the results of 'quasi-experimental' work carried out on spatial perception and knowledge by behavioural geographers. The utility of the technique increases when we realise that it shares certain features in common with multiple regression analysis, and that the two techniques may be combined in order to deal with complex data sets, including those which cannot be handled by classical ANOVA methods because of features such as unequal sample sizes. However, these are matters which can only be touched on here, and fuller reference is made later in this monograph. A further benefit flowing from discussion of ANOVA is that some consideration of research design is essential. Many of the principles involved apply not only under experimental conditions, but also in research settings commonly encountered by geographers.

In this monograph we can consider only the basic principles of research design, together with elementary concepts in the ANOVA itself. No previous knowledge of the techniques is assumed, and worked examples with graphical illustrations are provided throughout. No mathematics beyond simple algebra is required. References to more advanced treatments which involve calculus and matrix algebra are listed in the bibliography. It is assumed that the reader has taken or is taking an introductory statistics course covering elementary descriptive statistics, correlation analysis, statistical hypothesis testing and estimation, and sampling techniques.

Non-parametric procedures are not considered because the simpler ones are well-covered in elementary statistical texts for geographers (see Section VIII for further details), while the more complex ones need an entire monograph to themselves (see Wrigley (1979, 315-355) and Section VIII).

(ii) Relationships between variables

ANOVA consists of a set of techniques for describing and exploring relationships between one or more sets of explanatory or 'causal' variables and a single dependent or 'response' variable. Each explanatory variable is measured on a nominal or categorical scale and defined in terms of the mutually exclusive categories into which observations fall, and the response variable is normally measured on an interval or ratio scale, these scales defining what we generally understand to be 'measurements' such as those made on the Centigrade or Fahrenheit temperature scales (interval), scales of distance of height in metres (ratio), or cost in £ or $ (ratio). Often, an explanatory categorical variable is referred to as a factor.

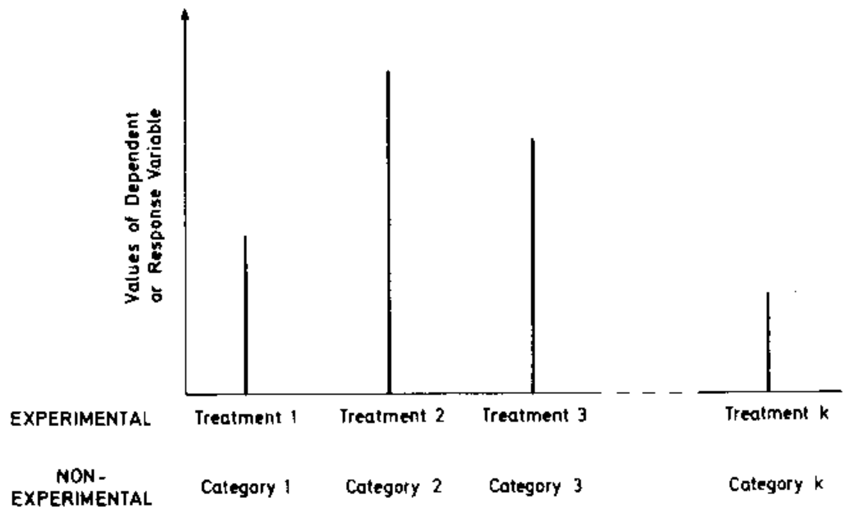


Fig. 1 Hypothetical results of experimental and non-experimental designs.

It is useful to distinguish experimental situations in which values (categories) of the explanatory variables may be manipulated by the researcher, from the non-experimental situations which are far more common in geographic research. For example, the explanatory variable (for simplicity, only one such variable is considered here) may consist of a number of experimental treatments devised by the researcher, as when student volunteers are divided into three groups, members of the first being required to drive along a given section of a motorway into a city centre, and those in the second and third travelling along the same route as front and back seat passengers respectively. In Figure 1 these groups are shown as treatments 1, 2 and 3. Every driver and passenger is then asked to recall as many features of the route as possible, the number of features named representing a value of the dependent or response variable. The heights of the vertical lines in Figure 1 represent the arithmetic means of the number of features recalled by each group. Other treatments may be envisaged, as shown on the right-hand side of Figure 1.

On the other hand, the explanatory variable may consist of mutually exclusive categories adopted or devised by the researcher such as socio-economic groups, life-cycle stages, or administrative or physiographic regions.

In ANOVA then, we estimate and compare the values of the response variable for different treatments or categories. As Ferguson (1978, 3) suggests, such analyses may be important in three ways. They many confirm or refute hypotheses about cause-effect relationships in the system under study, be used in an exploratory manner to uncover relationships not predicted by existing theory and thus stimulate further research or, finally, allow us to predict the responses liable to occur under specified conditions.

II RESEARCH DESIGN

(i) General Considerations

A highly desirable feature of all empirical research is that we should arrive at conclusions which are both valid and reliable - results are valid if they really mean what we think they mean, and reliable if they are not subject to unnecessary error. Our investigations must therefore be carefully planned and structured so as to permit control over many sources of variation - in other words, attention must be paid to research design.

The necessary control is most readily achieved under experimental conditions which, as already mentioned, are rarely encountered in geographic research. However, so-called 'quasi-experimental' situations (Campbell and Stanley, 1966), in which the investigator can actively manipulate certain variables, and allow for or control some sources of variation, are more common. These points may be examined in detail with reference to a classical research design illustrated in Figure 2.

	PRE-TREATMENT		POST-TREATMENT	DIFFERENCE
Experimental Group	O_1	X	O_2	$O_2 - O_1 = d_e$
Control Group	O_3		O_4	$O_4 - O_3 = d_c$

Fig. 2 Structure of a classical experimental design.

The basis of the design is that there should be two comparable groups, one known as the experimental group, the other as the control group. Measurements are obtained on the dependent variable of interest at the pre-treatment stage for both groups. In Figure 2, O_1 and O_3 represent, say, the mean scores for the experimental and control groups respectively. Members of the experimental group are then treated, as shown by the X, but those of the control group are not. Measurements on the dependent variable are then obtained once more, O_2 representing the mean post-treatment score for the experimental group, and O_4 the corresponding mean score for the control group. The differences between the two mean scores are denoted $d_e = O_2 - O_1$ (for the experimental group) and $d_c = O_4 - O_3$ (for the control group). If the hypothetical results of such an experiment are as plotted in Figure 3, we see that, in terms of our earlier discussion, the explanatory variable consists of just two categories or treatments. It also appears that the experimental treatment has a striking effect because d_e is so much greater than d_c. However, such a conclusion is tenable only if the two groups were equivalent i.e. comparable in all respects, before the experiment began. Otherwise, it could be argued that initial differences between the two groups have caused the difference between d_e and d_c, rather than the experimental treatment per se.

The two methods most commonly used to ensure equivalence of groups in all respects other than those specified by the treatments of the explanatory variable are matching and randomization. Let us consider the merits of these methods with respect to a hypothetical case study. A small group of third

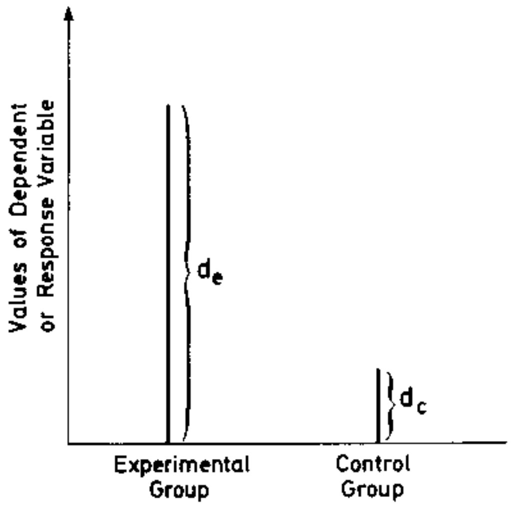


Fig. 3 Hypothetical results of a classical experiment.

year undergraduates studying behavioural geography at Reading University wish to know whether the 'Reading Walk' organised by the Geography Department for first year undergraduates taking the Human Geography course markedly increases student knowledge of the town centre - during the walk, land use and other activities in and around the town centre are observed and discussed in detail. The dependent variable, knowledge of the town centre, is measured by asking each student to name each of, say, 40 street scenes so that the number of correct answers gives each student's score. First thoughts might produce the research design of Figure 4(a), which simply compares the group of first year Human Geography students with a group of other (non-Human Geography) first year students. However, the two groups may not be equivalent in terms of characteristics likely to influence the behaviour under investigation. For example, all students choosing Human Geography have a background in academic geography and may learn about places more readily anyway than other students, or the proportion of males amongst the Human Geography students may be far lower than in the rest of the student body. The design of Figure 4(b) is far more satisfactory because Human Geography students are compared with two control groups, the first consisting of Physical Geography students and the second of Other students. This matches the experimental group with a more or less equivalent group in terms of geographic background, but may not be satisfactory in terms of other relevant variables. However, apart from the possible presence of one or two mature students, notice that the design automatically provides matching on the basis of age.

If we wish to rely further on <u>matching,</u> it becomes necessary to match each individual in one group with an individual in the other group or groups. A female in the experimental group who lives in lodgings near the town centre and owns a car, and a male in the same group living in a hall of residence who always travels by bicycle should each have a counterpart with similar characteristics in the control group(s), and so on for all individuals in each group. However, this method of achieving equivalence is difficult to attain on a large number of variables.

(a)

Human Geography Students (Experimental Group)	O_1	X	O_3	$d_e = O_3 - O_1$
Non-Human Geography Student (Control Group)	O_2		O_4	$d_c = O_4 - O_2$

(b)

Human Geography Students (Experimental Group)	O_1	X	O_4	$d_e = O_4 - O_1$
Physical Geography Students (Control Group 1)	O_2		O_5	$d_{c1} = O_5 - O_2$
Other Students (Control Group 2)	O_3		O_6	$d_{c2} = O_6 - O_3$

(c)

Human Geography Students (HG Experimental Group)	O_1	X	O_7	$d_{eH} = O_7 - O_1$
Human Geography Students (HG Control Group)	O_2		O_8	etc.
Physical Geography Students (PG Experimental Group)	O_3	X	O_9	
Physical Geography Students (PG Control Group)	O_4		O_{10}	
Other Students (O Experimental Group)	O_5	X	O_{11}	
Other Students (O Control Group)	O_6		O_{12}	

Fig. 4 More complex research designs.

<u>Randomization</u> provides a better method of constructing broadly equivalent groups. Consider yet another research design in which each Human Geography student is randomly allocated either to the experimental group or to the control group Figure 4(c). Once again, the X's denote treatments. It is argued that the effects of various variables should cancel out because each group has an equal chance of receiving individuals with characteristics tending to produce high, medium or low knowledge scores. This method also allows for the influence of variables which the investigator cannot anticipate. Nevertheless, if one variable is thought to be particularly important, a combination of randomization and matching may be employed. Suppose we want equal, or near equal, numbers of females and males in each group, and that 44 of the 80 first year Human Geography students are female. Modifying the randomization procedure to ensure that 22 females and 18 males are each randomly allocated to the experimental and control groups overcomes the problem. If the design also included Physical Geography and Other students in groups constructed on

similar lines, the results could be examined for differential impact of the Reading Walk on the knowledge scores of the various groups.

To summarize, basic features of the classical research design are

i) Comparison of each group with itself, before and after the treatment is given, and of the experimental and control groups.

ii) Manipulation so that it is clearly established that the causal variable or treatment, X, occurs before the effect on the response variable is measured.

iii) Control, by matching and/or randomization, so that other possible causal factors may be ruled out.

The features of the hypothetical example described in detail above do not have relevance only for studies in behavioural geography. Investigation of the effects of planting different kinds of vegetation on the stabilization of valley-side slopes could benefit from application of the same principles. A number of sites should be selected in equivalent erosional environments, and rates of erosion, runoff etc. measured on each (Figure 5). Treatments (i.e. X_1, X_2 and X_3 in Figure 5) should then be randomly assigned to sites (including the no treatment' treatment), and further measurements taken once each vegetation type has become firmly established. Differences may be calculated as before. Notice that in this case the explanatory variable has four categories or treatments.

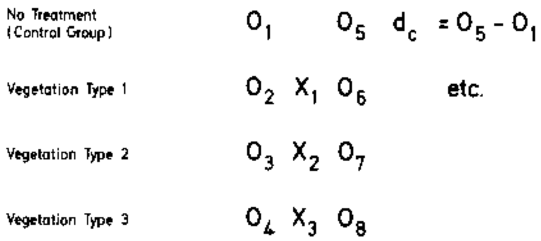


Fig. 5 Hypothetical research design for vegetation study.

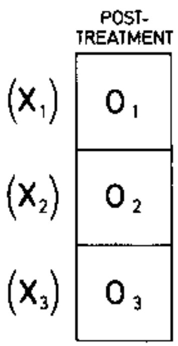


Fig. 6 Structure of a cross-sectional research design.

'Before and after' studies of this kind are not always possible, and so-called cross-sectional designs are very common in the survey research carried out by many human geographers. The structure of these is illustrated in Figure 6. An important feature is that the researcher arrives on the scene after the 'experiment' has been conducted. If the three treatments (X_1, X_2, X_3) in Figure 6 represent three different socio-economic groups, and the dependent variable is some measure of the frequency of visits to city centre shops, then there is no way in which the researcher can randomly assign individual households to those groups - for example, the average length of residence in the city of households in each croup may differ appreciably, or those in one group may live, on average, very much further away from the city centre, thus influencing their shopping behaviour. It is highly unlikely that any researcher would ever be permitted to randomly allocate households to these socio-economic groups (although cynics may argue that they do this already). However, a degree of control can be exercised by, for example, matching subgroups in terms of the distance at which they live from the city centre, or in terms of household composition, and various forms of statistical control can be exercised using methods beyond the scope of this monograph (see, for example, Ferguson (1978, 18-22), Silk (1979a, 4)).

So far, we have considered only questions of internal validity i.e. whether there are any factors, other than the explanatory variable of interest, which might affect the results and which therefore need to be identified and controlled for. The question of external validity is more familiar to most readers of this monograph, and has two aspects. First, is the sample of persons or items representative of some wider and identifiable population? Randomization does not ensure this, as geography students at Reading may not be representative of British students in general, and valleys in which test sites are located may not be representative of valleys developed on differing bedrock or under different climatic conditions. Second, are the circumstances under which the measurements were obtained unduly contrived and artificial (this is known as the problem of ecological validity by psychologists)? Showing photographs of street scenes to students is rather different from real-life in which a given scene is almost always encountered in the context of having passed through other scenes which are known to lead into it. The whole area of survey research, in which information on knowledge, attitudes and preferences is sought, abounds with such difficulties. The physical geographer seems to encounter fewer problems on this score.

More detailed treatment of the issues of internal and external validity may be found in Nachmias and Nachmias (1976, Ch. 3) and an introduction to sampling techniques in elementary textbooks such as Hammond and McCullagh (1978, Ch. 5) or Silk (1979b, Ch. 10).

ii) Case Study: The investigation of mean valley-side slope angles in Wyoming

Concepts of research design and ANOVA will initially be illustrated by reference to a study carried out by Melton (1960). Details of a study by Bowlby (1979), which are of interest to human geographers, will be given at a later stage when more advanced concepts in two-way ANOVA are discussed.

Melton investigated the effects of two explanatory variables or factors, erosional environment and valley-side orientation, on the dependent variable valley-side slope angle. One-way ANOVA allows us to examine the relationship between a single factor and the dependent variable. The relationship examined here will be that between erosional environment (the explanatory variable which we will call factor A) and mean valley-side slope angle (the dependent variable). Melton hypothesized that proximity to intermittently or continuously active tributary streams would influence gradational processes and so distinguished three erosional environments in which alluvial fans were absent from the neighbourhood of the slope (Fan absent), those in which the slope faced an alluvial fan i.e. faced a tributary on the opposite side of the valley (Fan Opposite), and those where the slope stood above an alluvial fan i.e. on the same side as a tributary channel and sufficiently close to it for the foot of the slope to intercept the fan (Above Fan) (Table 1).

Table 1 Observations on Valley-Side Slope Angles in the Laramie Mountains, Wyoming (adapted from Melton (1960))

A_1 Fan Opposite		A_2 Fan Absent		A_3 Above Fan	
Y_{1j}	Y^2_{1j}	Y_{2j}	Y^2_{2j}	Y_{3j}	Y^2_{3j}
19.42	377.1364	19.12	365.5744	16.84	283.5856
20.29	411.6841	18.38	337.8244	14.46	209.0916
23.21	538.7041	21.00	441.0000	18.58	345.2164
16.48	271.5904	14.75	217.5625	13.46	181.1716
15.54	241.4916	12.38	153.2644	13.21	174.5041
16.37	267.9769	15.33	235.0089	13.96	194.8816
Σ111.31	2108.5835	100.96	1750.2346	90.51	1388.4509

$n_1 = 6$	$n_2 = 6$	$n_3 = 6$	$N = 18$
$\bar{Y}_1 = 18.552$	$\bar{Y}_2 = 16.827$	$\bar{Y}_3 = 15.085$	$\bar{Y} = 16.821$
$\hat{\alpha}_1 = +1.731$	$\hat{\alpha}_2 = +0.006$	$\hat{\alpha}_3 = -1.736$	
$s_1 = 2.953$	$s_2 = 3.207$	$s_3 = 2.150$	

However, no specific *a priori* hypotheses about the relationship between erosional environment and the dependent variable were advanced, and this will be an important point to bear in mind when handling the ANOVA results.

In terms of our earlier discussion this is a cross-sectional research design. Control of, other variables (i.e. internal validity) was achieved almost entirely by matching. All slope angle measurements were obtained from similar east-west valleys in the Laramie Mountains, Wyoming, with low gradients, relatively homogeneous bedrock in which no influence of structural planes on valley-side slope angles could be detected, and little evidence of extensive modern gullying or channel trenching. Measurements of slope angles were taken at points along segments of each valley such that channel gradient, width of main valley and main slope angle appeared approximately uniform from

end to end. As Melton points out, external validity is, strictly speaking, limited to the study area and others similar to it. However, he believes the conclusions might well prove generally applicable in the subhumid to arid regions in middle latitudes of the northern hemisphere.

The measurements obtained in each of the erosional environments are recorded in Table 1. Each value represents an average of 4 to 6 slope angles measured at sites in three valleys. Erosional environment is denoted as factor A, A_1 representing category or level Fan Opposite, A_2 Fan Absent and A_3 Above Fan.

Calculation of mean valley-side slope angles provides an interesting summary of the data, as do the sample standard deviations of the observations within each category (Table 1). Such values should be carefully examined in the light of any pre-existing knowledge, whether theoretical or empirical, which will aid in their interpretation, before any inferential considerations are entertained.

III ONE-WAY ANALYSIS OF VARIANCE

(i) The Analysis of Variance Model and its assumptions

No matter how instructive examination of a pattern of results may be, viewing them in terms of the ANOVA model is also desirable. First, error limits about individual means (or with respect to differences between means) may be stated in the form of confidence intervals. Second, tests of significance may be carried out. Finally, the model provides a framework for assessing the relative importance of two or more factors or explanatory variables in their effect on the response variable. Both the independent and joint effects of factors may be examined.

The <u>fixed-effects model</u> is by far the most widely used by geographers, and is described in detail here. Let the mean values of the populations of measurements in categories A_1, A_2 and A_3 be represented by μ_1, μ_2 and μ_3 respectively, and the 'grand' or 'overall' mean of all measurements by μ. The difference between the grand mean μ and, say, the i^{th} category mean μ_i is represented by α_i such that

$$\alpha_i = \mu_i - \mu \qquad (1)$$

Rearrangement of (1), and study of Figure 7, shows that we may express each category mean as a deviation from the overall mean:

$$\mu_i = \mu + \alpha_i \qquad (2)$$

α_i is known as the <u>effect</u> of being in the i^{th} category. Within any given category, measurements are assumed to deviate about μ_i due to a whole host of minor factors, including measurement error, which tend to cancel each other out and follow a normal distribution (Figure 7). Any given observation, Y_{ij} (the j^{th} in the i^{th} group), may be expressed in terms of its deviation, ε_{ij}, from the mean of the i^{th} group:

$$Y_{ij} = \mu_i + \varepsilon_{ij} \qquad (3)$$

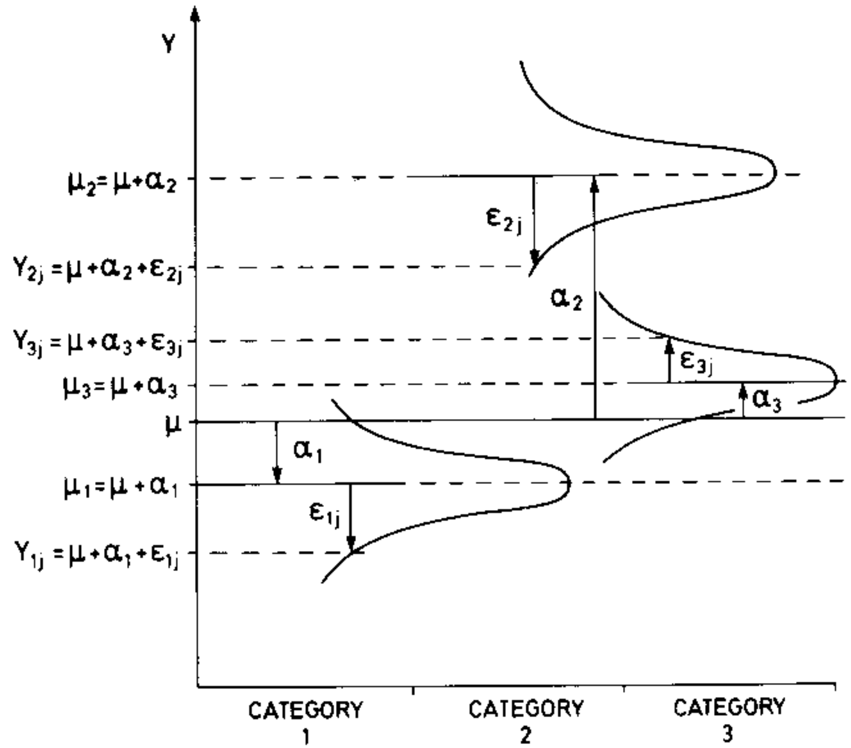


Fig. 7 Major features of the one-way ANOVA model.

If we substitute for μ_i from (2):

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad (4)$$

an individual value or score may be characterized in terms of two additive deviations from μ. The first denotes the effect of the region or group in which it falls (α_i), the second summarizes the combined influence of all the relatively minor factors which happen to affect it and is known as the population error or disturbance term (ε_{ij}).

All terms on the right hand side of (4) (other than ε_{ij}) are population parameters, whose values are known only if all possible measurements are obtained.

For the fixed-effects model, it is assumed either that all possible categories of the independent variable have been included, or that categories have been selected in a non-random manner from a larger set, so that findings can be generalized only to those particular categories in the overall population of measurements. In the study of valley-side slope angles, the three kinds of erosional environment were clearly not randomly selected, so a fixed-effects model is appropriate.

A further assumption is that the population means for each category are fixed or constant. It follows immediately that the values of the grand mean and of the effects must also be fixed. Thus μ, the μ_i and the α_i are non-random or non-stochastic elements. The weighted sum of the effects is assumed to be zero i.e. $\Sigma\, w_i\, \alpha_i = 0$, where w_i is usually taken to be the number of observations in the i^{th} group, n_i. As we shall see, this also implies that the weighted sum of the estimated effects, $\Sigma\, w_i \hat{\alpha}_i$, equals zero.

The remaining assumptions apply to the disturbance term, ε_{ij}. In decreasing order of importance, these are (Horton, 1978, 37):

1) The ε_{ij} are mutually independent, implying that they show no pattern either within or between categories, nor do they show any association with independent variables (which may be factors or continuous variables (see Silk (1979a, 24)) omitted from the analysis. This stipulation applies also where the ε_{ij} show a non-random spatial or mapped distribution (Cliff and Ord, 1973).
2) Within each category, the ε_{ij} show equal variance σ^2_ε. This is known also as the homescedasticity assumption.
3) Within each category, the ε_{ij} follow a normal distribution.

The first assumption is crucial, and far more likely to be fulfilled if random sampling and/or randomization techniques, other than matching, are employed. Assumption (2) becomes important if the sample sizes in each category are markedly unequal, and assumption (3) is not generally of major importance, particularly if the deviation from normality is not pronounced. For a discussion of checks on assumptions, and the related issue of transforming data, see Sections III(iv) and IV(iii).

(ii) Hypothesis Testing

1) Overall Test of Significance

If sample observations are obtained from k groups ($k \geq 2$), and we wish to simultaneously compare two or more sample means based on interval or ratio data, the statistical hypotheses are:

$$\left.\begin{array}{l} H_o: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k \;(= \mu) \\ H_1: \text{At least two of the } u_i \text{ are unequal} \end{array}\right\} \qquad (5)$$

For k 3, the population of observations would appear as in Figure 8(a) if H. were true, and might appear as in Figure 8(d) if H were true (and Ho false). Notice that (5) may also be written as

$$\left.\begin{array}{l} H_o: \alpha_1 = \alpha_2 = \alpha_3 = \ldots = \alpha_k = 0 \\ H_1: \text{At least one of } \alpha_i \text{ is not equal to zero} \end{array}\right\} (6)$$

i.e. under H_o (if H_o is true), all the effects will be zero; under H_1, at least one of the effects is nonzero. Under H_o, the frequency distributions of sample observations indicated schematically in Figure 8(c) are much more likely to occur in chance sampling than those in Figure 8(b), in which differences between sample means are greater. The logic of ANOVA depends upon a

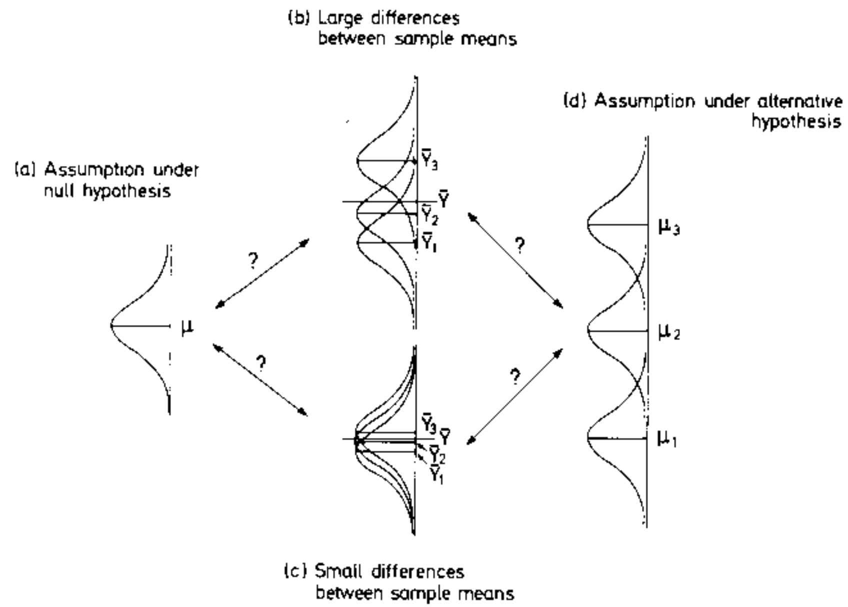


Fig. 8 Hypothesis-testing in one-way ANOVA.

comparison of the variation between groups, as measured by differences between the sample means, and the variation within groups, as measured by the amount of variation about each sample mean averaged over all groups. The procedure can also be viewed as one which splits up the total variation in the data into components ascribable to different causes (factors) by partitioning the total sum of squares according to the model specified. In one-way ANOVA, only one factor is involved so that the total variation is split simply into two components - explained by factor A and unexplained by factor A.

The total variation in the sample observations may be expressed in terms of the total sum of squares (SS_T) (N.B. SS = sum of squares), a quantity which in turn may be split into two components, the between-group sum of squares (SS_A) (i.e. the SS due to factor A) and the within-group or error sum of squares (SS_E) according to

$$\sum_i \sum_j (Y_{ij} - \bar{Y})^2 = \sum_i n_i (\bar{Y}_i - \bar{Y})^2 + \sum_i \sum_j (Y_{ij} - \bar{Y}_i)^2 \qquad (7)$$

$$(SS_T) \quad = \quad (SS_A) \quad + \quad (SS_E)$$

SS_T represents the sum of the squared deviations of each sample observation, Y_{ij}, about the sample grand mean of all the observations $\bar{Y}$, SS_A the sum of the squared deviations between each category sample mean and the sample grand mean weighted by the sample size (n_i) of each group, and SS_E the sum of the squared deviations between individual observations and their sample means. Geometrically, SS_E is obtained if each set of sample observations in Figure 8(b), or in Figure 8(c), is moved bodily until the sample means coincide,

and the squared deviations of the individual observations about the point of superimposition calculated and summed.

Assuming H_o to be true, the sample means $\bar{Y}_1$, $\bar{Y}_2$ and $\bar{Y}_3$ each represent independent estimates of μ, and SS_A provides an unbiased estimate of the variance, σ^2, of the population of observations shown in Figure 8(a) in terms of the between group variance.

$$s_B^2 = SS_A/(k-1) \qquad (8)$$

where k is the number of groups, and (k-1) the degrees of freedom associated with this variance estimate. The quantity SS_E provides another unbiased estimate (independent of that given by s_B^2) of σ^2 in terms of the error variance

$$s_B^2 = SS_E/(N-k) \qquad (3)$$

where N is the total number of observations, and (N-k) the degrees of freedom.

It can be shown that the ratio

$$F_{\nu_1, \nu_2} = s_B^2 \ / \ s_E^2 \qquad (10)$$

follows an F distribution with ν_1 = k-1 and ν_2 = N-k degrees of freedom when H_o is true. For a problem in which there are six observations in each group, and three groups, the degrees of freedom would be ν_1 = (3-1) = 2 and ν_2 = (18-3) = 15, and an F ratio exceeding 3.68 is to be expected at most 5% of the time.

Intuitively, it seems reasonable that the estimate of σ^2 given by the between-group variance s_B^2 is biased upward if H_o is false because deviations between category sample means and the sample grand mean will normally be increased if H_1 is true (compare Figures 8(b) and 8(c)). The estimate of σ^2 given by the error variance s_E^2 is unbiased no matter whether H_o is true or false (this follows from the assumption of approximately equal variances within each category). To reject H_o, therefore, we always expect s_B^2 to exceed s_E^2 and so the F test must be one-tailed because we are concerned only with the large ratios to be found in the right-hand tail of the distribution.

For the valley-side slope angle data shown in Table 1, the statistical hypotheses for an overall test of significance are

H_o : $\mu_1 = \mu_2 = \mu_3$ $(=\mu)$

or $\alpha_1 = \alpha_2 = \alpha_3 = 0$

H_1 : At least two of the μ_i are unequal

or At least one of the α_i is non-zero

Computational formulae will be used for SS_T and SS_A and (7) allows us to calculate SS_E by subtraction i.e.

$$SS_E = SS_T - SS_A$$

Results of the computations are conveniently entered in an ANOVA table (Table 2).

Table 2 Analysis of Variance for Valley-Side Slope Angle Data

Source	Sum of squares (SS)	Degrees of freedom (df)	Variance or mean square (MS)	F ratio
BETWEEN GROUPS	36.054	2	18.027	2.29
ERROR (WITHIN GROUP)	118.119	15	$S_E^2 = 7.875$ ($s_E = 2.806$)	
TOTAL	154.173	17		

By the computational formulae, we have

$$SS_T = \sum_i \sum_j Y_{ij}^2 - (\sum_i \sum_j Y_{ij})^2/N \qquad (11)$$

$$SS_A = \sum_i (\sum_j Y_{ij})^2/n_i - (\sum_i \sum_j Y_{ij})^2/N \qquad (12)$$

where n_i is the number of observations in the i^{th} group, and $N(= \Sigma n_i)$ is the total number of observations. Note that equation (11) represents the SS of all the individual observations minus a 'correction factor' which allows for deviations about $\bar{Y}$. SS_T sets an upper bound on all the other SS values to be calculated, in both one- and, as we see later, two-way ANOVA. Applied to the data in Table 1:

$$SS_T = (19.42)^2 + (20.29)^2 + \ldots + (13.21)^2 + (13.96)^2 - (302.78)^2/18$$
$$= 5247.269 - 5093.096 = \underline{154.173}$$
$$SS_A = (111.31)^2/6 + (100.96)^2/6 + (90.51)^2/6 - (302.78)^2/18$$
$$= 5129.1495 - 5093.096 = \underline{36.054}$$

and so, by subtraction

$$SS_E = 154.173 - 36.054 = \underline{118.119}$$

These values are entered in the first column of the ANOVA table (Table 2), and degrees of freedom in the second column. For SS_A, df $= \nu_1 = (k-1) = (3-1) = 2$, and for SS_E, df $= (N-k) = (18-3) = 15$. Notice that the df for SS_T is always given by $N-1$ i.e. the total number of observations minus one, and that the df for SS_A and SS_E should sum to $N-1$.

Adopting a preset significance level of 5% ($\alpha = 0.05$), the critical value of F is 3.68. The calculated F ratio of 2.29 based on 2 and 15 df is therefore not significant. We conclude that there is little evidence against the hypothesis of no difference between the mean slope angles observed in the three different erosional environments.

If the conclusion is stated in terms of effects, then the three estimated effects:

$$\hat{\alpha}_1 = \bar{Y}_1 - \bar{Y} = 18.552 - 16.821 = +1.731$$
$$\hat{\alpha}_2 = \bar{Y}_2 - \bar{Y} = 16.827 - 16.821 = +0.006$$
$$\hat{\alpha}_3 = \bar{Y}_3 - \bar{Y} = 15.085 - 16.821 = -1.736$$

do not differ significantly from zero at the 5% level (N.B. This is not the same as saying that any individual effect fails to differ significantly from zero at the 5% level, for reasons to be given below. Notice that the sum of effects is 1.731 + 0.006 - 1.736 = 0.001 and fails to equal zero only because of error due to rounding - weighting of effects is unnecessary here because sample sizes are equal.

2) Comparisons of Means

The overall test of significance given by the F ratio can only tell us whether a real difference between means is likely to exist, and does not specify where the difference (or differences) occur. However, comparison of means - whether pairwise or in a more complex form - must also be considered in the light of whether planned (a priori) comparisons or unplanned (a posteriori) comparisons are involved. The making of such comparisons is part of the general problem of multiple comparisons, for which many methods are employed - for instance, at least seven are listed in the SPSS manual (Nie et al (1975)).

Before looking closely at comparisons between means, the term comparison must be defined. The term contrast may also be used to denote a comparison. Definition of Comparisons. If there are three groups under study, then all possible pairwise comparisons i.e. comparisons between individual pairs of means, are

$$\bar{Y}_1 - \bar{Y}_2; \quad \bar{Y}_1 - \bar{Y}_3; \quad \bar{Y}_2 - \bar{Y}_3$$

Each of these comparisons has the general form

$$C_i\bar{Y}_i + C_j\bar{Y}_j = (1)\,\bar{Y}_i + (-1)\,\bar{Y}_j$$

the values of the coefficients being $C_i = 1$ and $C_j = -1$. Comparisons may also involve more than two means as in

$$\frac{\bar{Y}_1 + \bar{Y}_2}{2} - \bar{Y}_3; \quad \frac{\bar{Y}_2 + \bar{Y}_3}{2} - \bar{Y}_1$$

Here, we compare the mean of groups 1 and 2 with that of group 3, and the means of group 2 and 3 with that of group 1, respectively. For these comparisons, the general form is

$$C_i\,\bar{Y}_i + C_j\,\bar{Y}_j + C_k\,\bar{Y}_k = \tfrac{1}{2}\bar{Y}_i + \tfrac{1}{2}\bar{Y}_j + (-1)\bar{Y}_k \qquad (13)$$

where $C_i = \frac{1}{2}$, $C_j = \frac{1}{2}$, $C_k = -1$. Notice that the sum of the coefficients is zero in every case, and discussion in this monograph will be restricted to such cases. All these comparisons, and the associated statistical hypotheses, may be presented as in Table 3. It should be clear that an expression of the kind shown in (13) may be readily generalised to include any desired number of means.

Planned Comparisons. A planned comparison is one which the investigator has specified before collecting the data. If only one such comparison is specified, a t-test may be carried out using the statistic

$$t_{N-k} = \frac{\bar{Y}_i - \bar{Y}_j}{s_E\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \qquad (14)$$

Table 3 Statistical Hypotheses and Associated Coefficients

Comparison	Hypotheses	Coefficients		
		C_1	C_2	C_3
1.	$H_0: \mu_1 = \mu_2$	1	-1	0
	$H_1: \mu_1 \neq \mu_2$			
2.	$H_0: \mu_1 = \mu_3$	1	0	-1
	$H_1: \mu_1 \neq \mu_3$			
3.	$H_0: \mu_2 = \mu_3$	0	1	-1
	$H_1: \mu_2 \neq \mu_3$			
4.	$H_0: (\mu_1 + \mu_2)/2 = \mu_3$	$\frac{1}{2}$	$\frac{1}{2}$	-1
	$H_1: (\mu_1 + \mu_2)/2 \neq \mu_3$			
5.	$H_0: (\mu_2 + \mu_3)/2 = \mu_1$	-1	$\frac{1}{2}$	$\frac{1}{2}$
	$H_1: (\mu_2 + \mu_3)/2 \neq \mu_1$			

assuming that a simple pairwise comparison is involved, where s_E is the square root of the error mean square s_E^2, n_i and n_j the number of sample observations in groups i and j, and $\bar{Y}_i$ and $\bar{Y}_j$ the means in question. This expression may be generalised to include any number of means. If three means are involved, the general form of the test statistic is given by

$$t_{N-k} = \frac{C_i\bar{Y}_i + C_j\bar{Y}_j + C_k\bar{Y}_k}{s_E\sqrt{\frac{C_i^2}{n_i} + \frac{C_j^2}{n_j} + \frac{C_k^2}{n_k}}} \qquad (15)$$

For the particular comparison specified in (13), the statistic is

$$t_{N-k} = \frac{\frac{1}{2}\bar{Y}_i + \frac{1}{2}\bar{Y}_j - \bar{Y}_k}{s_E\sqrt{\frac{1}{4n_i} + \frac{1}{4n_j} + \frac{1}{n_k}}}$$

No matter how many means are involved in a single comparison, use of the t-test does not require that the overall F ratio should be significant if the comparison is planned.

If several comparisons have been planned matters are more complex. The comparisons must be orthogonal (for a definition see Kirk (1968, 70-78), a property that can only be readily achieved if the number of observations in each group is equal. Assuming these conditions are fulfilled, a t-test based on (14) or (15) may be used. Furthermore, the comparisons in which we are interested are not necessarily those which are orthogonal. Finally, it must be pointed out that setting a significance level of a for each of a number of individual (orthogonal) comparisons does not mean that the same level holds for the entire collection of comparisons. It is for this reason that an overall

F ratio may be insignificant while, holding a constant, one or more orthogonal comparisons may yield a significant ratio. For further discussion of planned comparisons, and of techniques for handling non-orthogonal comparisons, see Kirk (1968, 79-86).

Unplanned or a posteriori comparisons are usually carried out to track down the source of differences in the data where the overall test of significance has led to rejection of H_0. This process is sometimes known as 'data snooping' and raises problems because comparisons are not independent. Suppose it is decided to compare the largest and smallest of k>2 means, having obtained the overall result. For k = 3, the observed t ratio for the largest difference will exceed the tabulated 5% value approximately 13% of the time, for k = 6 approximately 40% and for k = 10 approximately 60%. If enough t ratios are calculated, a 'false positive' is almost assured (see also Kirk (1968, 82-83)).

Given a significant overall F ratio, Scheffé's method can be used to make all possible comparisons (pairwise or more complex) between means. The error rate for the entire set of possible pairwise comparisons is given by a, and for a pairwise comparison to be significant, it must exceed

$$S = \sqrt{(k-1)\, F_{\nu_1,\, \nu_2,\, \alpha}}\;\; s_E\sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \qquad (16)$$

where k is the total number of groups, ν_1 and ν_2 are the degrees of freedom associated with the between-group SS and the error SS respectively, and F the critical value of the F distribution based on ν_1 and ν_2 degrees of freedom at significance level α.

For non-pairwise comparisons, the expression for S must be modified and for three means takes the form

$$S = \sqrt{(k-1)\, F_{\nu_1\,\, \nu_2\,\, \alpha}}\;\; s_E\sqrt{C_i^2/n_i + C_j^2/\,n_j + C_1^2/\,n_1} \qquad (17)$$

By now, you will have gathered that comparison of means is not a straightforward topic, and for more comprehensive coverage the reader is referred to Kirk (1968) and O'Neill and Wetherill (1971). In general, the investigator should try to identify those comparisons which are of particular interest in advance by specifying planned comparisons. In most cases, however, some 'data snooping' will be inevitable, and the investigator may even wish to test planned comparisons and explore further relationships within the same data set (Kirk, 1968, 113-114).

If Melton had planned, in advance, to make a single pairwise comparison between Fan Opposite ($\bar{Y}_1$) and Above Fan ($\bar{Y}_3$) sample means, then the appropriate test statistic, based on (14), is

$$t_{18-3} = \frac{18.552 - 15.085}{2.806\sqrt{1/6 + 1/6}} = 2.141$$

(Note: the question of orthogonality does not arise because only one comparison is involved).

The difference ($\bar{Y}_1 - \bar{Y}_3$) is found to be just significant at the 5% level ($t_{15,0.05}$ = 2.131). This result apparently contradicts that of the overall

F test. As noted earlier, however, the results are not incompatible. Each is valid in carefully specified situations.

If the $(\bar{Y}_1 - \bar{Y}_3)$ contrast had been unplanned, then application of (16) gives a critical value for S, at the 5% level, of

$$S = \sqrt{(3-1)\, F_{2,15,0.05}} \quad s_E \sqrt{1/n_1 + 1/n_3}$$

$$= \sqrt{7.34} \times 1.619 = \underline{4.386}$$

The contrast $(\bar{Y}_1 - \bar{Y}_3) = 3.467$ and so fails to exceed the calculated S value. There is insufficient evidence to reject H_o: $\mu_1 = \mu_3$ at the 5% level. Yet again we reach a different conclusion because of the conditions under which the contrast has been specified.

A comparison suggested by inspection of the data, and clearly unplanned by Melton, involves all three means - the results show that the mean of the angles in the Fan Absent environment approximately equals the mean of the angles in the Opposite Fan and Above Fan environments. The appropriate hypotheses are

$$H_o: (\mu_1 + \mu_3)/2 = \mu_2$$

$$H_1: (\mu_1 + \mu_3)/2 \neq \mu_2$$

Using Scheffé's procedure, the critical difference obtained from (17) gives

$$S = \sqrt{(3-1)F_{2,15,0.05}} \quad s_W \sqrt{(\tfrac{1}{2})^2/6 + (-1)^2/6 + (\tfrac{1}{2})^2/6}$$

$$= \sqrt{2 \times 3.68} \times 2.806 \times \sqrt{0.25} = \underline{3.806}$$

The comparison in question is

$$(\bar{Y}_1 + \bar{Y}_3)/2 - \bar{Y}_2 = (18.552 + 15.085)/2 - 16.827 = -0.008$$

and cannot be regarded as significant at the 5% level.

Frequently, the numbers of observations in each group are unequal, and we end this section by showing how the calculations for t and S are modified if $n_1 = 10$, $n_2 = 6$, $n_3 = 8$ and $N = 24$.

The t statistic now becomes

$$t_{21} = \frac{18.552 - 15.085}{2.806\sqrt{1/10 + 1/8}} = \frac{3.467}{1.329} = \underline{2.609}$$

and

$$S = \sqrt{(3-1)F_{2,21,0.05}} \quad s_E \sqrt{(\tfrac{1}{2})^2/10 + (-1)^2/6 + (\tfrac{1}{2})^2/8}$$

$$= \sqrt{2 \times 3.47} \times 2.806 \times \sqrt{0.2229} = \underline{3.489}$$

3) Interval Estimation

We may not only need to know whether population means differ, but also by how much they differ. The likely magnitude of error associated with any one population mean may also be determined.

Taking the second point first, a confidence interval for an individual mean at the $(1 - \alpha) \times 100\%$ level may be determined from

$$\bar{Y}_i \pm t_{\alpha/2,N-k} \left(s_E / \sqrt{n_i}\right) \qquad (18)$$

where $t_{\alpha/2,N-k}$ is the critical $(\alpha/2) \times 100$ upper percentage point of a t distribution with N-k degrees of freedom, s_E is the square root of the residual mean square s_E^2, and n_i is the number of observations in the i^{th} group.

For the slope angle data of group 1 (Opposite Fan), the 95% confidence interval is therefore

$$18.552 \pm 2.131 \times (2.806/\sqrt{6})$$

$$= 18.552 \pm 2.442 = \underline{(20.994,\ 16.106)}$$

For differences between means, the procedure for constructing the confidence interval depends (as in hypothesis testing) on whether the differences are associated with planned or unplanned comparisons. The formulae that follow apply only to simple pairwise comparisons.

For planned (orthogonal) comparisons, a $(1 - \alpha) \times 100\%$ confidence interval is given by

$$\bar{Y}_i - \bar{Y}_j \pm t_{\alpha/2,N-k}\, s_E \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \qquad (19)$$

Under this assumption, the 95% confidence interval for the difference $\bar{Y}_1 - \bar{Y}_2$ is

$$18.552 - 16.827 \pm 2.131 \times 2.806 \times \sqrt{1/6 + 1/6}$$

$$= 1.725 \pm 3.450 = \underline{(5.175,\ -1.725)}$$

A comparable interval for unplanned comparisons is calculated from

$$\bar{Y}_i - \bar{Y}_j \pm S \qquad (20)$$

where S is as defined in (16).

Applied to the difference $\bar{Y}_1 - \bar{Y}_2$, this gives

$$18.552 - 16.827 \pm \sqrt{2 \times 3.47} \times 2.806 \times \sqrt{1/6 + 1/6}$$

$$= 1.725 \pm 4.269 = \underline{(5.994,\ -2.544)}$$

As we might expect, in order to counteract the bias introduced by data snooping, the confidence interval based on the S statistic is wider than that based on t.

All the above expressions for confidence intervals may be generalised by extending the notation employed in this section. For example, a comparison involving 5 means would require an expression of the form

$$s_E \sqrt{\frac{C_1^2}{n_1} + \frac{C_2^2}{n_2} + \frac{C_3^2}{n_3} + \frac{C_4^2}{n_4} + \frac{C_5^2}{n_5}}$$

to be included in (17).

4) Transformations

It may be desirable to transform the raw observations submitted to ANOVA so as either to fulfil the assumption of equal variances and/or to ensure that the observations within each group approximately follow a normal distribution.

Where group means and variances tend to be proportional, a square root transformation may be appropriate, the transformed score Y' being given by

$$Y' = \sqrt{Y}$$

The transformation

$$Y' = \sqrt{Y + 0.5} \text{ or } Y' = \sqrt{Y + 1}$$

should be used if any of the values of Y is less than 10. Such a transformation may be appropriate for example where the dependent variable represents the number of errors made in identifying each of a series of photographs (assuming the errors to be relatively rare) in a perception study.

For data sets in which group means and standard deviations tend to be proportional, a logarithmic transformation

$$Y' = \log_{10} Y \text{ or } Y' = \log_{10} (Y + 1)$$

may be used, the latter when some scores are zero or very small. This is because log (0) is undefined, and logarithms of very small numbers are 'very large'. This transformation is also used when observations show relatively strong right skewness, and for a good example see Schumm (1956).

A reciprocal transformation of the form

$$Y' = 1/Y \text{ or } Y' = 1/(Y + 1)$$

is used when the squares of the group means are proportional to the standard deviations. The latter formula should be used if any scores equal zero.

The angular transformation given by

$$Y' = 2 \arcsin \sqrt{Y}$$

is employed when means and variances are proportional and the distribution of Y is binomial in form. Y should be expressed as a proportion and the value of Y' may be obtained from tables such as those in Kirk (1968, 539). Haggett (1964) employs an arcsin transformation with respect to data on the proportion of forest cover in different regions in Brazil.

As means and variances are unrelated for normally distributed populations, a plot of the relationship between group means and variances for each set of transformed values may be used as a guide to choice of transformation. A procedure which applies each transformation to the largest and smallest score in each group may also be used. The range within each group is then calculated and the ratio of the largest to the smallest range determined. The transformation yielding the smallest ratio is chosen. Applied to the observations of Table 1, the values of Table 4 are obtained. There is little to choose between the transformations and the original observations in this case, although the logarithmic transform comes out best. If transformed, the observations should still be checked for normality, using graphical techniques or a goodness-of-fit test.

Table 4. Transformations Applied to Largest and Smallest Scores in Table 1

	CATEGORY 1	2	3	Largest range / Smallest range
Largest score (L)	23.21	21.00	18.58	
Smallest score (S)	15.54	12.38	13.21	
Range	7.67	8.62	5.37	8.62/5.37 = 1.605
$\sqrt{L}$	4.82	4.58	4.31	
$\sqrt{S}$	3.94	3.52	3.63	
Range	0.88	1.06	0.68	1.06/0.68 = 1.559
log (L)	1.3656	1.3222	1.2691	
log (S)	1.1914	1.0927	1.1209	
Range	0.1742	0.2295	0.1482	0.2295/0.1482 = 1.549
1/L	0.043	0.048	0.054	
1/S	0.064	0.081	0.076	
Range	0.021	0.033	0.022	0.033/0.021 = 1.571

IV TWO-WAY ANALYSIS OF VARIANCE

(i) Notation and Terminology

If a second factor or criterion of classification is introduced, our original notation must be extended. For the Melton data, the three types of erosional environment are now represented by three rows of Table 5. The second factor to be introduced, valley-side orientation, is known as factor B (you will recall that erosional environment is known as factor A). Factor B consists of two groups - north-facing and south-facing - and is represented by the two columns of Table 5. This structure for the data is known as a factorial design, because for every category or level of factor A there are observations for every level of factor B, and vice versa.

To distinguish the two sets of population means, the row means are represented by μ_{i+} (i=1,3) and the column means by μ_{+j} (j=1,2). The grand mean is still denoted by μ_{++}. The corresponding sample means are $\bar{Y}_{i+}$ for rows, $\bar{Y}_{+j}$ for columns, and $\bar{Y}_{++}$ for the grand mean. The + notation is used to show that observations are summed over all categories of the factor represented by the +. Thus $\bar{Y}_{i+}$ denotes the mean of all sample observations in the ith row of the table (summing over all columns), and $\bar{Y}_{+j}$ the mean of all sample observations in the jth column of the table (summing over all rows). The (sample) grand mean $\bar{Y}_{++}$ results from averaging the sum of observations over all rows and all columns. An individual observation must show the row and column within which it falls. If there are several observations in each cell forming the intersection of a row and column, as is the case in Table 5, then each observation is denoted by Y_{ijk}, where i denotes the row, j the column,

Table 5. Observations on Valley-Side Slope Angles in Tableau for Two-Way Analysis of Variance

		ASPECT		
		NORTH FACING	SOUTH FACING	
EROSIONAL ENVIRONMENT	FAN OPPOSITE	19.42 20.29 23.21	16.48 15.54 16.37	$\bar{Y}_{1+}$ = 18.552 $\hat{\alpha}_1$ = +1.731
	FAN ABSENT	19.12 18.38 21.00	14.75 12.38 15.33	$\bar{Y}_{2+}$ = 16.827 $\hat{\alpha}_2$ = +0.006
	ABOVE FAN	16.84 14.46 18.58	13.46 13.21 13.96	$\bar{Y}_{3+}$ = 15.085 $\hat{\alpha}_3$ = 1.736
		$\bar{Y}_{+1}$ = 19.033 $\hat{\beta}_1$ = 2.212	$\bar{Y}_{+2}$ = 14.609 $\hat{\beta}_2$ = -2.212	$\bar{Y}_{++}$ = 16.821

and k the position within the ijth cell. Thus, in Table 5 the second observation in cell (2,1) is $Y_{2,1,2}$ = 18.38. Just as the difference between μ_{++} and the ith row mean μ_{i+} is represented by $\alpha_i = \mu_{i.} - \mu_{++}$, so the difference between μ_{++} and the jth column mean is represented by

$$\beta_j = \mu_{+j} - \mu_{++} \qquad (21)$$

and β_j is estimated by $\hat{\beta}_j = \bar{Y}_{+j} - \bar{Y}_{++}$. In order to distinguish them from interaction effects that are considered later, we point out here that the α_i and β_j are conventionally known as the main effects due to factors A and B respectively.

Finally we note that the population mean in the ijth cell is denoted by μ_{ij}, and the corresponding estimator by $\bar{Y}_{ij}$.

(ii) Carrying Out the Analysis

1) Model and Assumptions

The two-way ANOVA model may be written

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \qquad (22)$$

any given observation now being expressed both in terms of its deviation on factor A, 'row-wise', from μ_{++}, and on factor B, 'column-wise', from μ_{++}, as well as in terms of a within-cell error term ε_{ijk}. For the fixed-effects model it is assumed that the population means of rows or columns are fixed or constant. The assumptions outlined with respect to the ε_{ij} also hold for the ε_{ijk} (i.e. independence, equal variance, normality).

Each cell should contain the same number of observations as this ensures no relationship in the sample data between the two nominal-scale variables representing the factors. Only if there is no relationship can we unambiguously distinguish the sums of squares explained by the 'row factor' (factor A) and the 'column factor' (factor B), without having to resort to more advanced techniques of analysis (a conventional analysis may be carried out if a particular proportionality exists among the sample sizes in different groups. See Kirk (1968, 200-202) for details).

A further assumption implicit in the statement (22), is that the row and column effects are additive in their influence on values of the dependent variable. However, the implications of this assumption can be more meaningfully discussed later.

2) Application to Valley-side Slope Angle Data

Provided the number of observations in each cell (n_{ij}) is equal (which is the case in Table 5 where $n_{ij} = n = 3$) the sum of squares explained by the difference between columns can be regarded as coming completely from the within-row or unexplained (by row) sum of squares. A corollary of this is that the two factors are statistically independent of one another in terms of their influence on the dependent variable, and so the sum of squares explained by row differences remains the same, no matter whether row or column differences are extracted first, and the sum of squares explained by column differences has the same properties. Thus, the problem of multicollinearity, common in multiple regression analysis, is avoided. The total variation can now be divided as follows

Total SS = between-row SS + between-column SS + unexplained (error)SS

or

$$SS_T = SS_A + SS_B + SS_E \qquad (23)$$

The notation in (23) is desirable because it may be easily generalised to take account of additional factors in a 3- or higher-way ANOVA design, and of the interactions between them. The sums of squares may be expressed algebraically as

$$\Sigma\Sigma\Sigma(Y_{ijk} - \bar{Y}_{++})^2 = bn\sum_i(\bar{Y}_{i+} - \bar{Y}_{++})^2 + an\sum_j(\bar{Y}_{+j} - \bar{Y}_{++})^2 + \Sigma\Sigma\Sigma(Y_{ijk} - \bar{Y}_{ij})^2 \qquad (24)$$

where a is the number of rows, b the number of columns, and n the number of observations per cell (N.B. we can write n for n_{ij}, since all n_{ij} are equal) The between-row SS (originally referred to as SS_A in the section on one-way ANOVA) due to factor A is estimated using the computational formula

$$SS_A = \sum_i(\sum_{jk}\Sigma Y^2_{ijk})/bn - (\Sigma\Sigma\Sigma Y_{ijk})^2/N \qquad (25)$$

the between-column SS due to factor B using

$$SS_B = \sum_j(\sum_{ik}\Sigma Y^2_{ijk})/an - (\Sigma\Sigma\Sigma Y_{ijk})^2/N \qquad (26)$$

The new within-cell, unexplained or error SS is obtained by substruction from

$$SS_E = SS_T - SS_A - SS_B \qquad (27)$$

where SS_T is the total SS originally calculated using (11).

For the slope angle data column differences relate to the categories 'north-facing' and 'south-facing' of the aspect factor, so that the number of columns is b = 2, and we know from the section on one-way ANOVA that the number of rows is a = 3.

As we have already calculated SS_A in one-way ANOVA, it remains only to compute

$$SS_B = (171.30)^2/9 + (131.48)^2/9 - 5093.0960$$
$$= 5181.1866 - 5093.0960 = \underline{88.091}$$

and

$$SS_E = 154.173 - 36.054 - 88.091 = \underline{30.028}$$

The degrees of freedom are (a-1) for SS_A, (b-1) for SS_B, and (N-1)-(a-1)-(b-1) for SS_E. The ANOVA table for the two-way analysis is given in Table 6. The mean squares associated with the differences between rows and columns each give independent and unbiased estimates of the total variance in the population, σ^2, when the appropriate null hypothesis

$$H_o: \mu_{1+} = \mu_{2+} = \mu_{3+} \; (=\mu)$$
$$(\alpha_1 = \alpha_2 = \alpha_3 = 0)$$

or

$$H_o: \mu_{+1} = \mu_{+2} = \mu_{+3} \; (=\mu)$$
$$(\beta_1 = \beta_2 = \beta_3 = 0)$$

is true. Otherwise, inflated estimates are to be expected and we refer to the upper tail of the appropriate F distribution as in the one-way case.

Table 6. Analysis of Variance for Main Effects in Two-Way Table - Valley-side slope angle data

Source	Sum of squares (SS)	Degrees of freedom (df)	Variance or mean square (MS)	F ratio
BETWEEN ROWS	36.054	2	18.027	8.40
BETWEEN COLUMNS	88.091	1	88.091	41.07
ERROR	30.028	14	2.145	
TOTAL	154.173	17		

Notice that the difference between rows (erosional environments) is now significant at the 5% level, as the calculated F of 8.40 exceeds $F_{2,14,0.05} = 3.74$. It is quite possible for such a change to occur, compared with one-way ANOVA, the reason being that we have identified and subtracted out the systematic variation due to between-column differences which was originally left within the error component of the one-way ANOVA, and so obtained a more accurate estimate of the error variance. In other words, we have specified a better model. Introducing another factor does not always reduce this variance estimate as the degrees of freedom lost from the error line (only 1 df in this example) may more than cancel out any reduction in the error sum of squares.

For the aspect factor, the sample mean for north-facing slopes is $\bar{Y}_{+1} = 19.033$, and for south-facing slopes is $\bar{Y}_{+2} = 14.609$, so that the estimated effects are

$$\hat{\beta}_1 = \bar{Y}_{+1} - \bar{Y}_{++} = 19.033 - 16.821 = \underline{2.212}$$
$$\hat{\beta}_2 = \bar{Y}_{+2} - \bar{Y}_{++} = 14.609 - 16.821 = \underline{-2.212}$$

(Compare these calculations with those for the $\hat{\alpha}_i$ on p. 16).

3) Additivity and Interaction

Although the two-factor model (which includes both erosional environment and aspect as independent categorical variables) provides a better fit to the data than the one factor model (based on erosional environment alone), the former model is not necessarily the 'best model'. The form of the models fitted to the slope data by ANOVA may be visualized in terms of a 'response surface'. Figure 9(a) shows the values of the $\bar{Y}_{i+}$ as 'heights' within each category of erosional environment and also the estimated effects ($\hat{\alpha}_i$) about the grand mean $\bar{Y}$. A similar response surface with respect to aspect alone is shown in Figure 9(b).

We have seen how it is possible to estimate row and column effects. The two-way ANOVA model considered is based on the assumption that row and column effects are additive, so that estimates of cell means are given by

$$\hat{\mu}_{ij} = \bar{Y}_{++} + \hat{\alpha}_i + \hat{\beta}_j$$

and, substituting values of the effects

$$\hat{\mu}_{11} = 16.821 + 1.731 + 2.212 = \underline{20.764}$$
$$\hat{\mu}_{12} = 16.821 + 1.731 - 2.212 = \underline{16.340}$$
$$\hat{\mu}_{21} = 16.821 + 0.006 + 2.212 = \underline{19.039}$$
$$\hat{\mu}_{22} = 16.821 + 0.006 - 2.212 = \underline{14.615}$$
$$\hat{\mu}_{31} = 16.821 - 1.736 + 2.212 = \underline{17.297}$$
$$\hat{\mu}_{32} = 16.821 - 1.736 - 2.212 = \underline{12.873}$$

Plotting these values in the form of the response surface in Figure 10 yields a pattern which represents the addition of the two sets of effects shown in Figures 9(a) and 9(b) to the estimate of the grand mean.

If the relationship between the dependent variable and the two factors was truly additive, the observed values of the $\bar{Y}_{ij}$ should coincide perfectly with the estimates provided by the $\hat{\mu}_{ij}$. However, because of minor measurement errors and the combined influence of other factors not explicitly included in the analysis, minor deviations between the $\bar{Y}_{ij}$ and $\hat{\mu}_{ij}$ are almost always to be expected, even if our additive model was the correct one. These differences are shown in Figure 10, and represented by

$$\widehat{\alpha\beta}_{ij} = \bar{Y}_{ij} - \hat{\mu}_{ij} \qquad (28)$$

The deviations are listed beside Figure 10 and reinforce the impression given by the diagram that the observed cell means ($\bar{Y}_{ij}$) differ little from those estimated under the assumption of additivity.

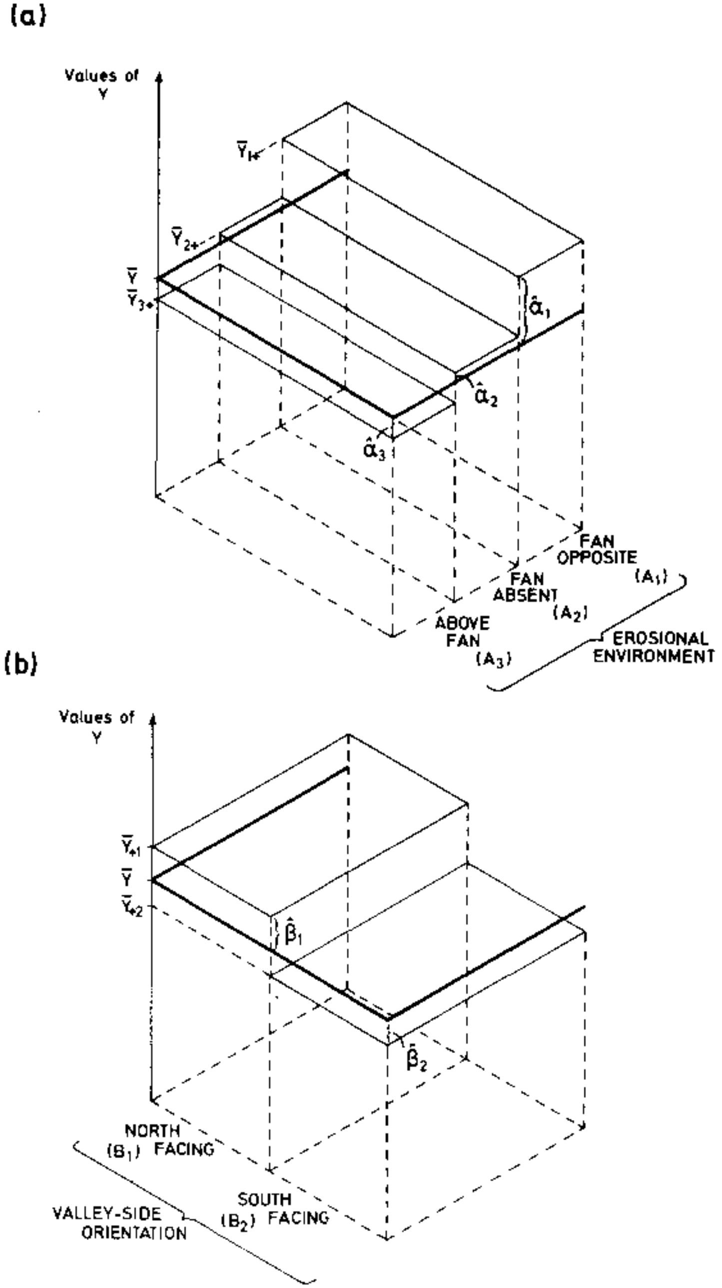


Fig. 9 ANOVA estimates in the form of response surfaces - one-way design.

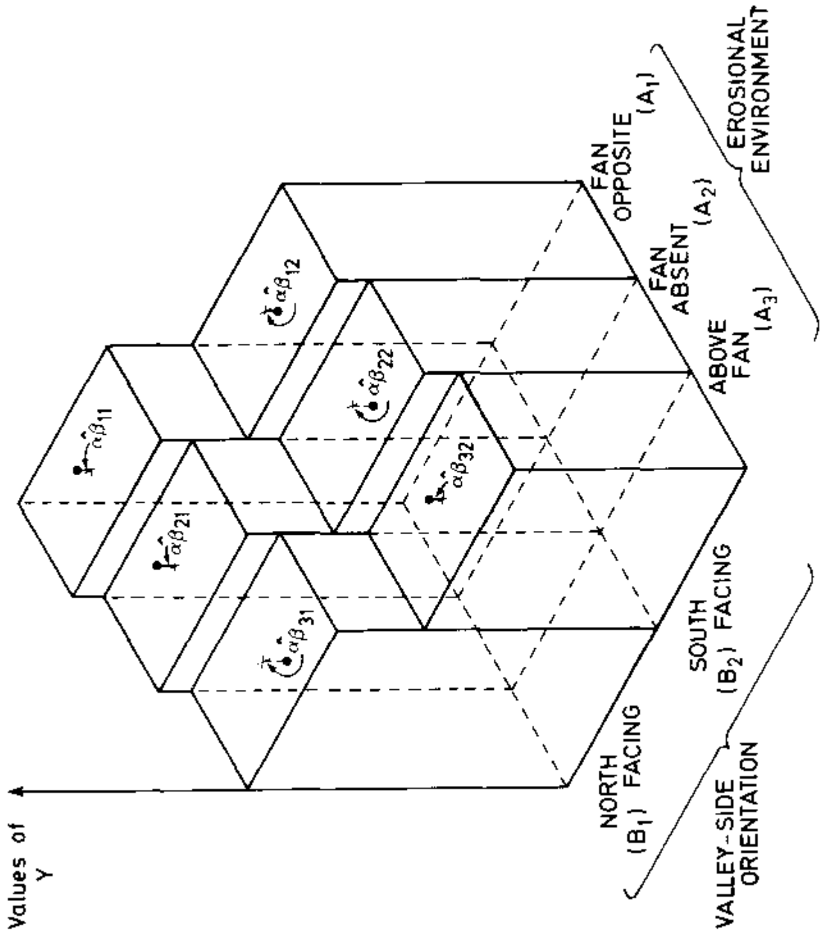


Fig. 10 ANOVA estimates in the form of a response surface - additive model in a two-way design.

Table 7. Cell mean values estimated from the two-way additive model for slope angle data

		ASPECT		
		NORTH FACING	SOUTH FACING	
EROSIONAL ENVIRONMENT	FAN OPPOSITE	20.764	16.340	Difference 1.725
	FAN ABSENT	19.039	14.615	Difference 1.742
	ABOVE FAN	17.297	12.873	
		Difference 4.424		

The pattern of relationships between the $\hat{\mu}_{ij}$ suggests another method of checking for additivity. In any given row of Table 7, we find that the difference between the means in the two columns is always 4.424 i.e. the difference between the two column means in row 1 is 20.764-16.340 = 4.424, and for rows 2 and 3 is 19.039-14.615 = 4.424 and 17.297-12.873 = 4.424 respectively. A similar relationship holds for differences between pairs of row means, holding the column constant. If the cell means of Table 7 are plotted as in Figure 11(a), it is clear that the response variable Y behaves in the same manner when in a given category of one factor, relative to other levels of that same factor, no matter what the category of the other factor in the analysis e.g. north-facing slopes are steeper than south-facing slopes, and by the same magnitude, no matter which of the three erosional environments is present. Graphically, additivity is represented by perfectly parallel lines - the lines need not be straight, however, as illustrated by the pattern of cell means in Figure 12. A plot of the observed cell means ($\bar{Y}_{ij}$) from Melton's data therefore strongly suggests that the additive model is adequate.

Where the response variable behaves differently in a particular category of one factor, according to the category of the other factor into which it falls, the two factors are said to interact, and the effect of such behaviour is described as interaction. Essentially, the response variable is held to behave in a distinctive manner in one or more of the cells of the two-way table, so much so that one or more cells must be considered individually rather than in terms of the additive model which involves averaging over all cells in a given row or column. An example in which strong interaction is present is considered in Section IV(iii). If the interaction between two factors A and B is being considered, we usually write this as AxB. In the case of two-way ANOVA we sometimes write rows x columns or rxc, or use factor names e.g. (erosional environment) x (aspect).

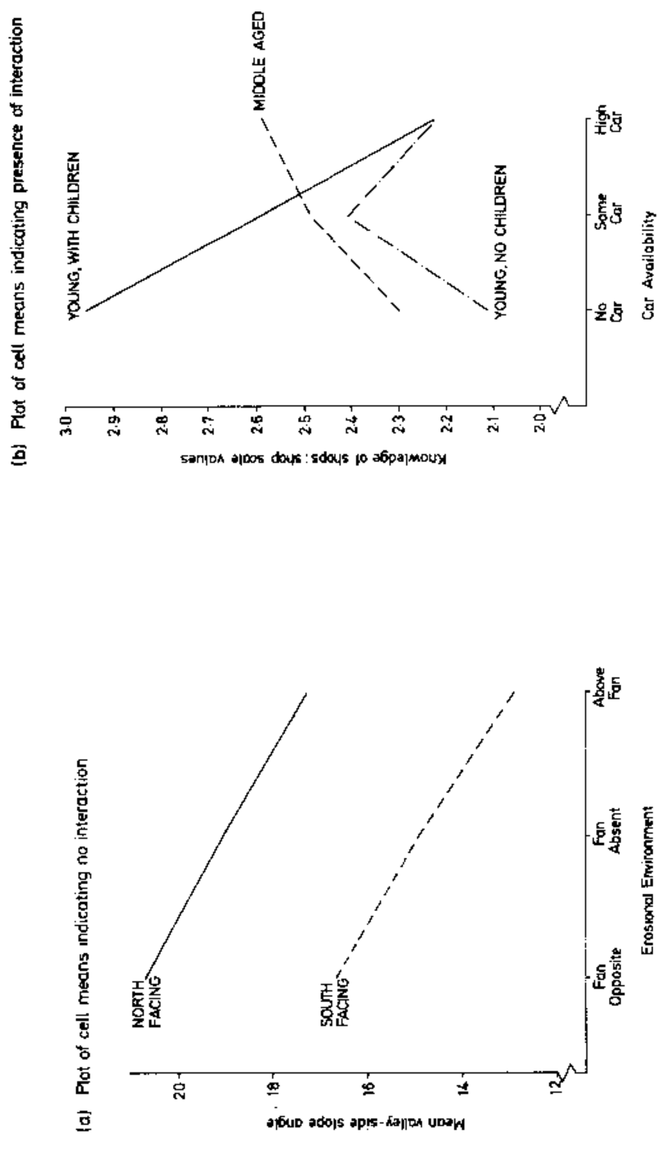


Fig. 11 Checks for interaction using graphical plots.

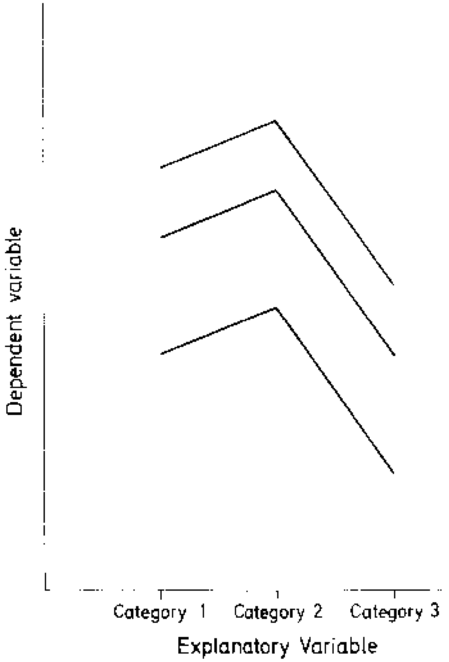


Fig. 12 Plot of cell means for a perfectly additive model (no interaction effects)

Plots of cell means are always very useful as they not only suggest whether interaction is occurring, but also allow us to pinpoint its source(s). As will be shown later, the latter is essential if results are to be interpreted.

4) Testing for Interaction

A formal test of significance for the presence of interaction may be carried out if the additive model of (22) is elaborated to include a term representing the difference between cell mean values based on the additive model and those based on the non-additive model

$$\mu_{ij} = \mu_{++} + \alpha_i + \beta_j + \alpha\beta_{ij} \qquad (29)$$

where $\alpha\beta_{ij}$ represents the interaction effect in the ijth cell. In Figure 10, estimates of the $\alpha\beta_{ij}$ are denoted $\widehat{\alpha\beta}_{ij}$ because quantities in this diagram are based on sample data. If there is no interaction, all values of $\alpha\beta_{ij}$ should be zero, and the corresponding estimates provided by the $\widehat{\alpha\beta}_{ij}$ should be 'small'. This suggests an overall test for interaction based on the hypotheses

$$H_o: \alpha\beta_{11} = \alpha\beta_{12} = \ldots = \alpha\beta_{32} = 0$$

H_1: At least one of the $\alpha\beta_{ij}$ is nonzero

As with the row and column effects, the overall test does not isolate any particular interaction effect as significant, but shows whether the variance due to all such effects is significantly greater than the estimated error variance.

Provided the number of observations per cell is equal, the sum of squares due to interaction can be extracted from the error SS (SS_E) about the two-way ANOVA. The value of an interaction term in the population is given by

$$\alpha\beta_{ij} = \mu_{ij} - \alpha_i - \beta_j - \mu_{++} \qquad (30)$$

and estimated by

$$\widehat{\alpha\beta}_{ij} = \bar{Y}_{ij} - \hat{\alpha}_i - \hat{\beta}_j - \bar{Y} = \bar{Y}_{ij} - \hat{u}_{ij}$$

where $\hat{\mu}_{ij}$ is the estimated cell mean value under the assumption of additivity. The computational formula for the interaction sum of squares (SS_{AB}), $n\sum_{ij}\sum \widehat{\alpha\beta}_{ij}$, is

$$SS_{AB} = \sum_{ij}\sum(\Sigma Y_{ijk})^2/n - \sum_i\{(\sum_{jk}\Sigma Y_{ijk})^2/bn\} - \sum_j\{(\sum_{ik}\Sigma Y_{ijk})^2/an + (\sum_{ijk}\sum\Sigma Y_{ijk})^2/N \qquad (32)$$

and has (a-1) x (b-1) degrees of freedom. For the ANOVA model which includes an interaction term, we estimate the error SS from

$$SS_E = SS_T - SS_A - SS_B - SS_{AB} \qquad (33)$$

a quantity which has N-ab degrees of freedom.

For the slope data

$$SS_{AB} = \{(62.92)^2 + (48.39)^2 + (58.50)^2 + (42.46)^2 + (49.88)^2 + (40.63)^2\}/3$$
$$- 5129.1495 - 5181.1866 + 5093.0960$$
$$= 5221.4766 - 5129.1495 - 5181.1866 + 5093.0960 = \underline{4.237}$$

All but one of the values in the line above have been calculated previously ; 5129.1495 and 5093.0960 are the uncorrected row SS and correction value respectively as calculated in the one-way ANOVA example, and 5181.1866 is the uncorrected column SS.

Finally

$$SS_E = 154.173 - 36.054 - 88.091 - 4.237 = \underline{25.791}$$

The hypotheses with respect to interaction effects are tested by comparing the interaction mean square and the error mean square from the new model (stated in (29)) using an F ratio as shown in Table 8. Overall, the interaction effects are clearly insignificant, and it is normal practice to re-absorb the sum of squares due to interaction within the error SS under these circumstances. Apart from giving us a simpler model, a further advantage of this practice is that we obtain more degrees of freedom on which to base our error estimate, and should obtain narrower confidence intervals and more sensitive tests (because the error mean square should be reduced). This is not the case for the slope angle data, since the error MS of Table 6 is 2.145 and that of Table 8, 2.149. In fact, Scheffé (1959, 126) cautions that if the interaction SS is not in fact zero, then the expected value of the pooled mean square will be biased upward, having the opposite of the desired effect. To obtain enough degrees of freedom for the error (say a minimum of 10), it is best to plan a study bearing this point in mind.

Table 8. Analysis of Variance for Main and Interaction Effects in a Two-way Table - Valley-side slope angle data

Source	Sum of squares (SS)	Degrees of freedom (df)	Variance or mean square (MS)	F ratio
BETWEEN ROWS	36.054	2	18.027	8.39
BETWEEN COLUMNS	88.091	1	88.091	40.99
INTERACTION (ROWS x COLUMNS)	4.237	2	2.119	0.97
ERROR	25.791	12	2.149	
TOTAL	154.173	17		

(iii) <u>A More Complex Example: A Study of Shopping Knowledge in Oxford</u>

The following example is included because transformation of the dependent variable, and interpretation of interaction effects, are required, making it typical of many research problems to which ANOVA is applied.

A survey of sample households was carried out in Oxford in the winter of 1975 and the spring of 1976 (Bowlby (1979)). A psychological scaling technique was used to derive scores representing shoppers' perceived knowledge of the location of grocery shops. Respondents were placed in one of three car availability categories:- (1) no car availability (2) some car availability i.e. a car was occasionally available for shopping (3) full car availability i.e. car always available for shopping when required, and one of three life cycle categories:- (1) middle aged - households whose head is between the ages of 50 and 65 (2) younger people without children - households whose head was 49 years old or less and without children (3) younger people with children - households whose head was 49 years old or less and where there were preschool or school age children. In all life-cycle groups, more than 80% of the respondents were women.

To simplify the analysis, only isolated corner shops, or shops in small centres, are included, so that scaled knowledge scores are available for 21 shops. Sample means and variances for each of the cells in the 3x3 table (Table 9(a)) are thus based on 21 observations in each case. A plot of these data (Figure 13(a)) shows a marked inverse relationship between the sample means and variances. There is no standard transformation available to deal with an inverse relationship of this kind, and so the scores may be reversed (by multiplying by minus one). We now have a direct relationship between cell variances and means, but the dependent variable measures the level of ignorance, rather than the level of knowledge, about any particular store i.e. the higher the score, the greater the level of ignorance. Under these circumstances, the appropriate transformation is either a square root or a logarithmic transformation of the original variable. Results of these transformations are shown in Figures 13(b) and 13(c) and it is clear that the logarithmic transformation is far more satisfactory, as there is no apparent relationship in this case, but still a pronounced relationship for the values based on the square root transformation.

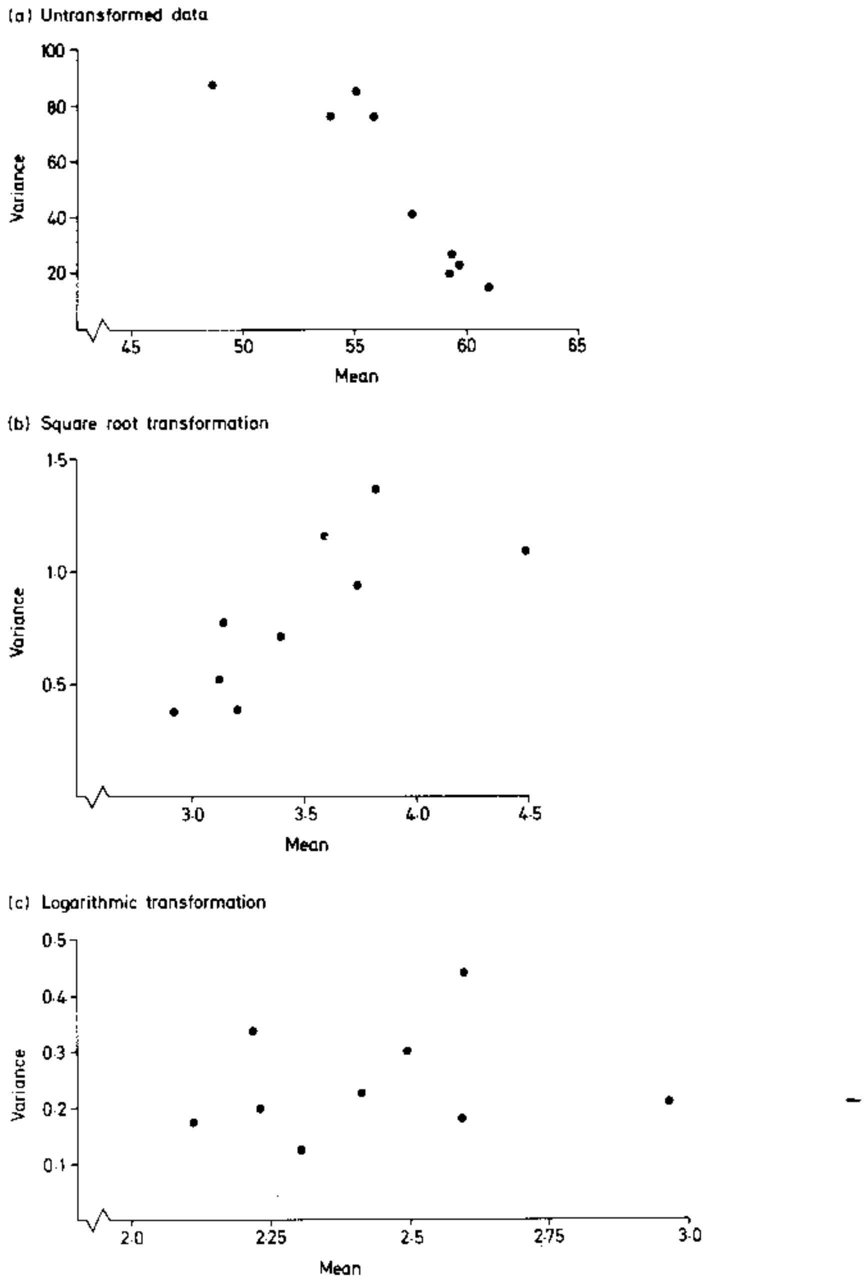


Fig. 13 Plots of cell variances against cell means for the Oxford shopping data.

Table 9. Data and Two-Way ANOVA for Shopping Study

a) Cell means and variances -untransformed data

		LIFE CYCLE (LC) MIDDLE-AGED	YOUNG W'OUT CHILDREN	YOUNG WITH CHILDREN	
CAR AVAILABILITY (CA)	NO CAR	59.34 17.77	61.06 13.16	48.67 87.24	mean variance
	SOME CAR	55.89 75.19	57.62 39.23	55.10 84.57	mean variance
	FULL CAR	53.98 75.38	59.35 29.70	59.72 22.39	mean variance

b) Two-way ANOVA table- transformed (logarithmic) data

Source	Sum of squares (SS)	Degrees of freedom (df)	Variance or mean square (MS)	F ratio
BETWEEN ROWS (LIFE CYCLE, LC)	3.933	2	1.966	7.891
BETWEEN COLUMNS (CAR AVAILABILITY, CA)	0.751	2	0.376	1.507
INTERACTION (LCxCA)	6.664	4	1.666	6.686
ERROR	44.853	180	0.249	
TOTAL	56.200	188		

A two-way analysis of variance produced the breakdown of Table 9(b). Differences due to life cycle are significant at the 5% level, but those due to car availability are insignificant, even at this level. The interaction between the two factors, life cycle and car availability, is shown to be significant at the 5% level, which is hardly surprising when we look at the plot of observed cell mean values against car availability, distinguished by life cycle category in Figure 11(b). Thus, the relationship between car availability and ignorance differs radically between life-cycle categories - ignorance decreases consistently with increase in car availability for young households with children, but shows the opposite trend for middle-aged households. For young households without children, there is an initial rise in ignorance with increasing car availability, and then a fall. The relationship for young households with children is much as one would expect, as mothers with young children but without cars are particularly circumscribed in their activities. In the case of the middle-aged, however, it appears some other influence must also be at work - those with cars may tend to be relatively affluent and mobile members of the professional and managerial classes who have moved to Oxford at a certain stage of their careers and know the city

chiefly from car travel. In contrast, those without cars may tend to be working class, less affluent and less mobile, but more liable to have lived in Oxford all their lives and so knowing it far more intimately through walking, bus travel and local change of residence. These interpretations are all somewhat speculative, requiring further data and analysis to substantiate or refute them. However, it is important to recognise the power of graphical methods to display interesting structures in data and suggest further lines for investigation. Clearly, the relationships discussed would have remained hidden if the main effects alone had been considered. Striking differences between the behaviour of the slope angle data, considered earlier, and the shopping knowledge data are also apparent.

(iv) Two-Way ANOVA with One Observation per Cell

On occasion, we may have only one observation per cell. This may be because we lack the resources to obtain further information, or because there were originally unequal numbers of observations in each cell. In the latter case, the observation in each cell may represent the mean of the original observations.

Table 10. Two-Way Analysis of Variance based on One Observation per Cell

a) Raw data - mean slope angle values in each cell

		ASPECT NORTH FACING	SOUTH FACING
EROSIONAL ENVIRONMENT	FAN OPPOSITE	20.97	16.13
	FAN ABSENT	19.50	14.15
	ABOVE FAN	16.63	13.54

b) Two-Way Analysis of Variance for mean slope angle data

Source	Sum of squares (SS)	Degrees of freedom (df)	Variance or mean square (MS)	F ratio
BETWEEN ROWS	12.006	2	6.003	8.54
BETWEEN COLUMNS	29.393	1	29.393	41.81
ERROR	1.406	2	0.703	
TOTAL	42.805	5		

Table 10(a) shows the slope angle data with cell means substituted for the original values. The total sum of squares is given by

$$SS_T = \Sigma\Sigma Y_{ij}^2 - (\Sigma\Sigma Y_{ij})^2/N$$
$$= (20.97)^2+(16.13)^2+\ldots+(13.54)^2-(100.92)^2/6$$
$$= 1740.2788-1697.4743 = \underline{42.805}$$

the between-row SS (due to factor A) by

$$SS_A = \sum_i(\sum_j Y_{ij})^2/b - (\sum\sum Y_{ij})^2/N$$
$$= \{(37.1)^2+(33.65)^2+(30.17)^2\}/2 - 1697.4743$$
$$= 1709.4807 - 1697.4743 = \underline{12.006}$$

and the between-column SS (due to factor B) by

$$SS_B = \sum_j(\sum_i Y_{ij})^2/a - (\sum\sum Y_{ij})^2/N$$
$$= \{(57.1)^2+(43.82)^2\}/3 - 1697.4743$$
$$= 1726.8674 - 1697.4743 = \underline{29.393}$$

There is no within-cell sample variation as such because there is only one observation per cell. However, it is common practice to estimate the error variance from the interaction variance under these circumstances, provided there is no strong evidence of any interaction effects. A plot of the values from Table 10(a) showed little evidence of interaction, and so SS_E is obtained by subtraction as before

$$SS_E = SS_T - SS_A - SS_B$$
$$= 42.805-12.006-29.393 = \underline{1.406}$$

Results are summarized in Table 10(b). The row effects are shown to be insignificant, and the column effects significant, at the 5% level. However, we should be a little wary of these results, as the error mean square is based on only 2 degrees of freedom and may not be particularly reliable.

Because of the way in which the interaction SS(SS_{AB}) is used to estimate the error SS here, there is no formal method available for assessing the statistical significance of the interaction effects.

(v) Two-way ANOVA with Unequal Numbers of Observations per Cell

The estimation procedures outlined above work only if the number of observations per cell in a two- or higher-way fixed-effects ANOVA design is constant. If the observations have been obtained experimentally, and only one or two are missing from a relatively small number of cells - perhaps because a subject failed to turn up for an experiment, or because of faulty apparatus - then many researchers use a method which has the effect of minimizing the error SS (Kirk, 1968, 146-147). Alternatively, different estimation procedures may be used, such as that incorporated in GLIM (Baker and Nelder, 1978), in which missing values cause no particular problems. Note that the degrees of freedom for the total and error lines in the ANOVA table are each reduced by one for every missing value estimated.

(vi) Limitations

Variation in the number of observations per cell clearly imposes some limits on the estimation procedures that may be used for 2- or higher-way ANOVA, and makes compliance with basic assumptions of the ANOVA model more compelling than in the one-way case. As already mentioned (pp 17-20), comparison of means can bring the researcher up against all kinds of problems associated with sequential tests carried out on the same body of data e.g. overall

F test for significance, followed by tests involving selected pairs of means. Proceeding in this area can be like entering a minefield, and reference to a more detailed text such as that by Kirk (1968) is highly recommended, and, as far as possible, so is careful advance specification of the objectives of the study and identification of comparisons of particular interest. Four- or higher-way ANOVA designs may be difficult to handle because interpretation of interaction effects is not straightforward and there can be a bewildering number to examine. A three-way design, based on factors A, B and C, is manageable, giving 3 two-way interactions (AxB, AxC, BxC) and 1 three-way interaction (AxBxC). In the four-factor case, however, there can be up to 6 two-way, 4 three-way and 1 four-way interactions to consider. Satisfactory estimation of the error variance may require very large numbers of observations.

V RANDOM-EFFECTS OR COMPONENTS OF VARIANCE MODELS

If the categories or treatment levels in a one-way Analysis of Variance are a random sample from a larger population of categories or treatment levels, the model is known as a random-effects, Type II or components of variance model (The fixed-effects model is sometimes known as th- Type I model). This terminology may seem a little confusing at this stage, but all -hree descriptions are widely used. For example, if the erosional environment categories employed by Melton could be regarded as a random sample from a much larger (and preferably infinite) population of environments, a random-effects model may be employed. If α_i represents the effect of the ith treatment (or of being in the ith category), then it is assumed the population of α_i values is normally distributed with mean zero and variance σ_A^2. The values of the α_i for the categories in the study are further assumed to be independent of one another and of the error terms ε_{ij}. Finally, the grand mean is assumed to be fixed, and the ε_{ij} to be mutually independent and normally distributed with mean zero and constant variance σ_ε^2 as in the fixed-effects model.

Analysis based on the random-effects model may be extended to the two-way case (or beyond). If it made sense to suppose that very minor differences in orientation influenced valley-side slope angle, then valleys within each erosional environment could also be selected which fell into a small number of randomly selected orientation categories. The β_j, representing orientation effects, are assumed to show the same properties as the α_i, except that the variance of the population from which the former are drawn is σ_B^2.

Where random-effects models are used, the primary aim is to estimate the grand mean and the variance across populations of categories rather than category means and differences between them. For the one-way design, the entire ANOVA table and the F test are the same for both fixed- and random-effects models, although interpretation is of course rather different. For more complex designs, however, the F tests frequently differ, as will be evident when a particular kind of Type II design is considered in the next section. We also note that the random-effects model is far more sensitive to violation of the assumptions of normality and equal variance for the ε_{ij}. In general, the random-effects or Type II model is associated with a nested or hierarchical research design, a special case of which is discussed in Section VI on Spatial Modelling. However, we note here that nested fixed-effects models may also be employed.

Mixed models are appropriate when a study is concerned with, say, all I regions in a country, or I purposefully selected regions, but with only J subregions randomly selected from large populations of subregions within each of the I regions. Lack of space prevents further treatment of the mixed model here, but for an exposition see Scheffé (1959, Ch.8) or Snedecor and Cochran (1967, 288-289).

VI SPATIAL MODELLING

Analysis of variance techniques have also been used to analyse map data by constructing simple trend surfaces (for an introduction to trend surface analysis proper see Unwin (1975)) and by estimating the variance associated with different spatial scales in a set of observations.

(i) Simple Trend Surface Construction

Krumbein (1956) described a technique for distinguishing 'regional' and 'local' effects in mapped data. The observations must be available for points on a regular grid, and are arranged in rows and columns for each of which (row and column) means are then computed. 'Expected values' for each grid point are then calculated according to the two-way additive model i.e.

$$\hat{\mu}_{ij} = \bar{Y}_{++} + \hat{\alpha}_i + \hat{\beta}_j \qquad (34)$$

where $\hat{\mu}_{ij}$ is the estimated 'expected value' for the observation at the grid point in the ith row and jth column and the $\bar{Y}_{++}$, $\hat{\alpha}_i$ and $\hat{\beta}_j$ are as defined earlier. The contour map based on the expected values constitutes the 'regional map' of large scale trends in the observations, and the contour map derived by subtracting expected values from the corresponding observed values at each grid point depicts a residual surface of 'local effects'. As a quick and easy method, this approach can give important qualitative insights. However, it produces a relatively simple discontinuous surface based on the first-order polynomial terms of a trend surface analysis, and is not advisable with large maps with many rows and columns. All cross-product terms are consigned to the residual surface, and such terms are important if trends in the data do not happen to be aligned with the grid orientation.

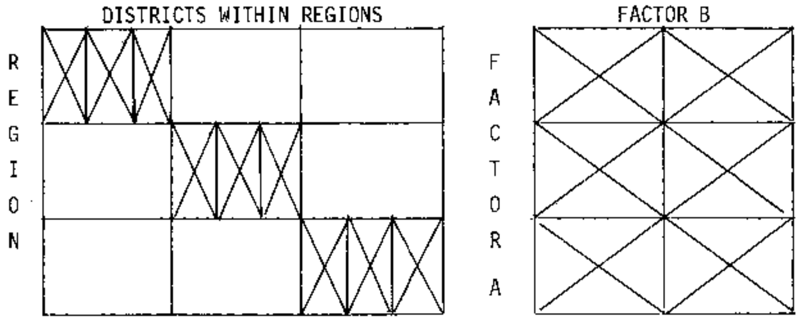
The two-way ANOVA design has also been adopted by geologists so that the 'rows' and 'columns' which intersect to yield the points on a spatial grid can be given a meaningful interpretation in terms of the processes being studied. For example, changes in sedimentary attributes due to the main down-stream current ('along the rows') and to cross-fluctuations in currents ('down the columns') may be compared with the residual variation. Miller and Kahn (1962, 409-418) provide detailed discussion of this and other examples. We must point out here that both row and column factors are 'quantitative' rather than 'qualitative' in such cases, and that attention should be paid to the possibility of fitting a continuous response surface either by eye or by using an appropriate technique such as trend surface analysis, otherwise valuable information may be lost (Mead and Pike (1975)).

(ii) Components of Variance Models

Yet another area in which ANOVA has been used is in the identification of the different spatial levels or scales at which a phenomenon is present, or

Table 11. Differences in Structure between Factorial and Hierarchical Designs

a) Hierarchical or Nested Design b) Factorial Design



at which a process is operating. Because, in these studies, a subregion at the lower level in a two level hierarchy cannot simultaneously exist within two different regions at the higher level, the structure for observations is in the form of a hierarchical or nested design (Table 11(a)), and should be contrasted with the factorial design described earlier (p.23)(Table 11(b)). We must emphasize that although most hierarchical designs are analysed using components of variance models, the two are not invariably associated. Before considering how this is done, we define terms and notation for describing spatial hierarchies and associated observations.

1) Description of Spatial Hierarchies

Data may be available for regularly-shaped spatial units, such as quadrats, or for irregularly shaped units such as administrative areas. In the regular case, the basic units can often be readily aggregated to for a balanced hierarchy in which the number of units nested at any particular level is constant. Figure 14(a) shows a case in which the entire area is divided into 4 regions, each of which is sub-divided into 2 districts, which are in turn subdivided into 4 parishes. The hierarchy may also be represented by a tree diagram (Figure 14(b)). We adopt the convention that the entire area represents level 0, and that the terms region, district and parish will be used to describe spatial units at levels 1, 2 and 3 respectively. This terminology will be used when illustrating the application of the components of variance model to hypothetical hierarchies and observations later in this section. The hierarchy of Figure 14(a) is even, which means that the units at any given level in the spatial hierarchy are of equal size. Thus, the system in Figure 14(a) represents a balanced even hierarchy. In Figure 15, the hierarchy is both unbalanced and uneven - at any given level, spatial units may be subdivided into varying numbers of units of varying size. Most systems of administrative regions are of this form. If we consider England, for example, the entire country represents level 0, standard regions might represent level 1, counties and county boroughs level 2, parishes and wards level 3.

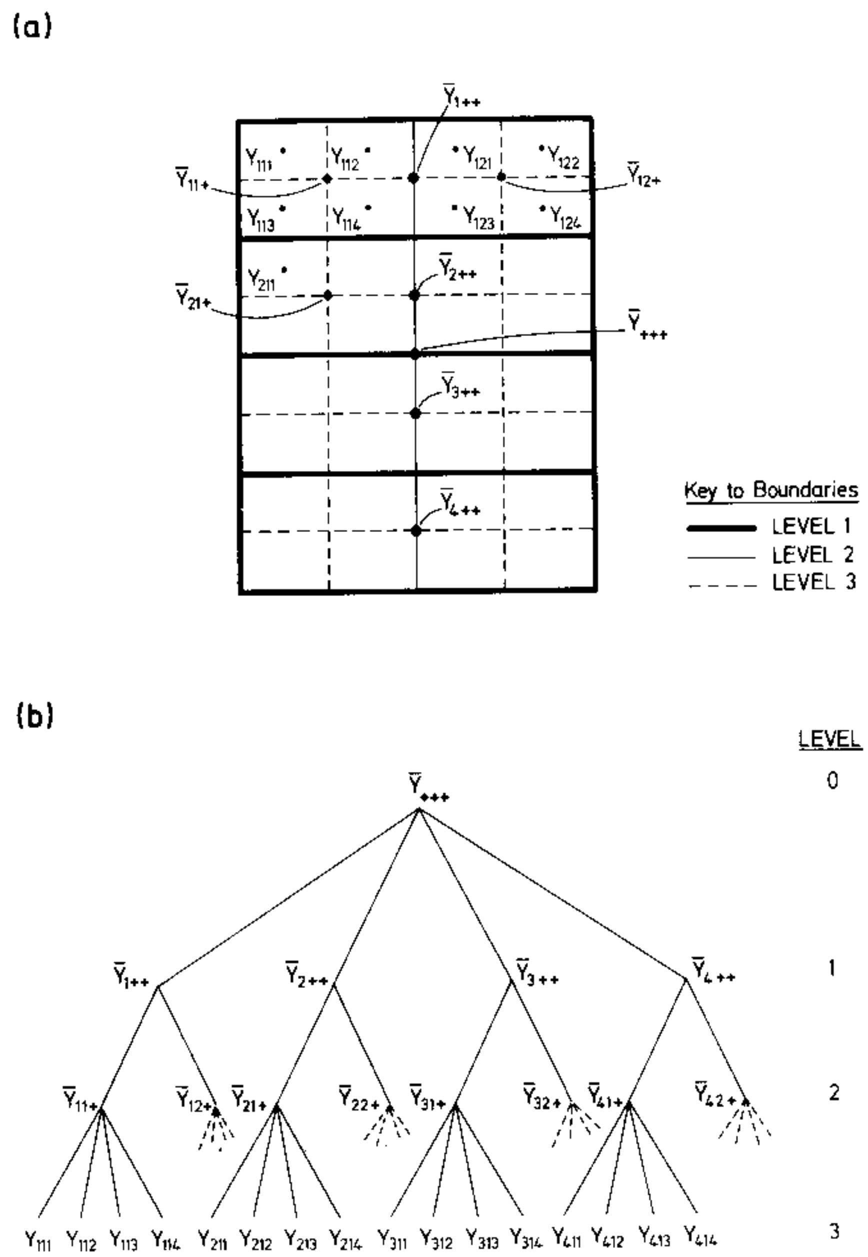


Fig. 14 Examples of a balanced spatial hierarchy a) Map b) Tree diagram.

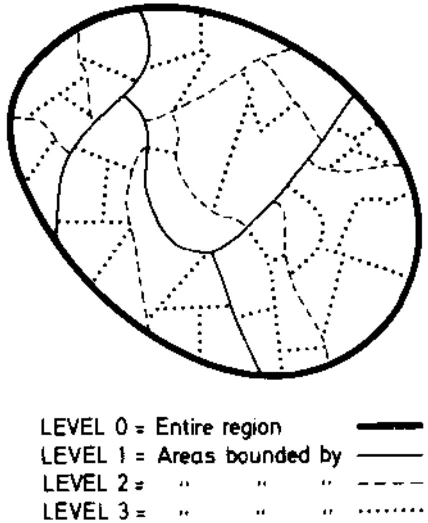


Fig. 15 Example of an unbalanced spatial hierarchy.

Notation is most readily explained with respect to a balanced even hierarchy such as that of Figure 14(a), where mean values (or single observations) are located at the centres of the areas they represent. Thus, $\bar{Y}_{+++}$ represents the mean value for observations in the entire area (the mean of Y is summed over all the subscripts which are plusses), $\bar{Y}_{i++}$ the mean of observations in the ith region, $\bar{Y}_{ij+}$ the mean of observations in the jth district of the ith region, and Y_{ijk} the observation in the kth parish of the jth district of the ith region.

Although only three levels at which spatial variation may be identified are shown in Figures 14 and 15, the technique can be readily extended to incorporate more e.g. Moellering and Tobler (1972), Haggett et al (1977, 387-389). The most satisfactory results are obtained for a balanced even hierarchy because like is compared with like at each level, at least in terms of spatial resolution. Computations are also more straightforward. Uneven hierarchies, which may or may not be balanced, are less easy to handle and the calculations more complex, but interesting results are obtainable (Moellering and Tobler, 1972, 44-47).

2) The Components of Variance Model

The nested or hierarchical components of variance model may be written

$$Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \gamma_{ijk} \qquad (35)$$

where Y_{ijk} is as defined previously and μ represents the grand mean of all the observations in some population. In terms of a spatial hierarchy, it is assumed that the α_i represent members of the large population of possible effects attributable to location in different regions at level 1. Similarly, the β_{ij} represent members of a population of effects which might conceivably operate between districts (at level 2) and within regions, and the γ_{ijk}

members of a population of effects that might operate between parishes (level 3) and within districts. It is assumed also that the α_i, β_{ij} and γ_{ijk} represent independent drawings from normally distributed populations with mean zero and variances σ^2_α, σ^2_β and σ^2_γ respectively. For a spatial hierarchy, therefore, these. assumptions imply that variation between districts in any one region will be the same as that between districts in any other region and that variation between parishes will be the same irrespective of the district and region in which they are located. Such assumptions should be borne in mind because it is easy to envisage a spatial hierarchy in which, for example, there is considerable variation between district means in one region and virtually none at all in another.

A more fundamental difficulty arises because, to date, geographers have only used the components of variance model to examine spatial hierarchies for which there is a complete enumeration both of the spatial units in the area under investigation and of the phenomenon of interest such as census data. Under these circumstances, we may appeal to two possible justifications for using the model. The first, and more widely invoked, is that the set of observations represents only one of a very large possible number of different sets that might have arisen. The observed population is regarded as a 'sample' from a 'superpopulation' (Blalock (1972); see also Cliff (1973)). Under this interpretation, each possible set of observations that <u>might</u> have arisen provides a new possible set of values for the a_i, and the same argument is used for the β_{ij} and γ_{ijk}.

The second justification for employing the model is that there is a very large, or possibly infinite, number of ways in which any given area may be subdivided into a hierarchy of spatial units with the same properties e.g. 4 regions at level 1, 3 districts within each region at level 2, 4 parishes within each district at level 3. Each possible subdivision produces a new 'sample' of effects at each level in the hierarchy. For a discussion of this approach see Openshaw (1977).

3) Estimation of Variance Components

The model stated in (35) may be rewritten as

$$Y_{ijk} - \mu = \alpha_i + \beta_{ij} + \gamma_{ijk} \qquad (36)$$

so that the variation of each observation about the grand mean is expressed in terms of the sum of effects at each level in the hierarchy. If values of sample effects we have 'captured' in any particular set of observations are estimated by

$$\hat{\alpha}_i = \bar{Y}_{i++} - \bar{Y}_{+++}$$

$$\hat{\beta}_{ij} = \bar{Y}_{ij+} - \bar{Y}_{i++}$$

$$\hat{\gamma}_{ijk} = Y_{ijk} - \bar{Y}_{ij+}$$

then the counterpart of (36) may be rewritten in terms of sample observations on Y to give the sums of squares

$$\sum_{ijk}\sum\sum(Y_{ijk} - \bar{Y}_{+++})^2 = \sum_{ijk}\sum\sum(\bar{Y}_{i++} - \bar{Y}_{+++})^2 + \sum_{ijk}\sum\sum(\bar{Y}_{ij+} - \bar{Y}_{i+})^2 + \sum_{ijk}\sum\sum(Y_{ijk} - \bar{Y}_{ij+})^2 \qquad (37)$$

$$i = 1, I \;;\; j = 1, J_i \;;\; k = 1, K_{ij}$$

or

$$SS_T = SS_\alpha + SS_\beta + SS_\gamma \qquad (38)$$

with degrees of freedom (N-1), (I-1), $\sum_i(J_i-1)$, $\sum_{ij}\sum(K_{ij}-1)$ respectively where N is the total number of parishes, I the total number of regions, J_i the number of districts in the ith region, and K_{ij} the number of parishes in the jth district of the ith region. For a balanced hierarchy, all J_i = J and all K_{ij} = K.

For the set of artificial data of Figure 16, N = 48, I = 4, all J_i = J = 3 and all K_{ij} = K = 4. Computational formulae for the analysis of spatial components of variance are given in Table 12, and a tableau setting out the necessary quantities based on these observations in Table 13. For these data:

$$SS_\alpha = (225+6889+3025+529)/12 - (176)^2/48$$
$$= 889 - 645.33 = \underline{243.67}$$
$$SS_\beta = (25+16+36+729+784+784+361+225+441+121+9+81)/4 - 889$$
$$= 903 - 889 = \underline{14.00}$$
$$SS_\gamma = 926 - 903 = \underline{23.00}$$

giving a total sum of squares (SS_T) of 243.47+14.00+23.00 = 280.67. Direct calculation of the total sum of squares, using the computational formula at the foot of Table 12, gives

$$SS_T = 926 - (176)^2/48 = \underline{280.67}$$

agreeing with previous calculations.

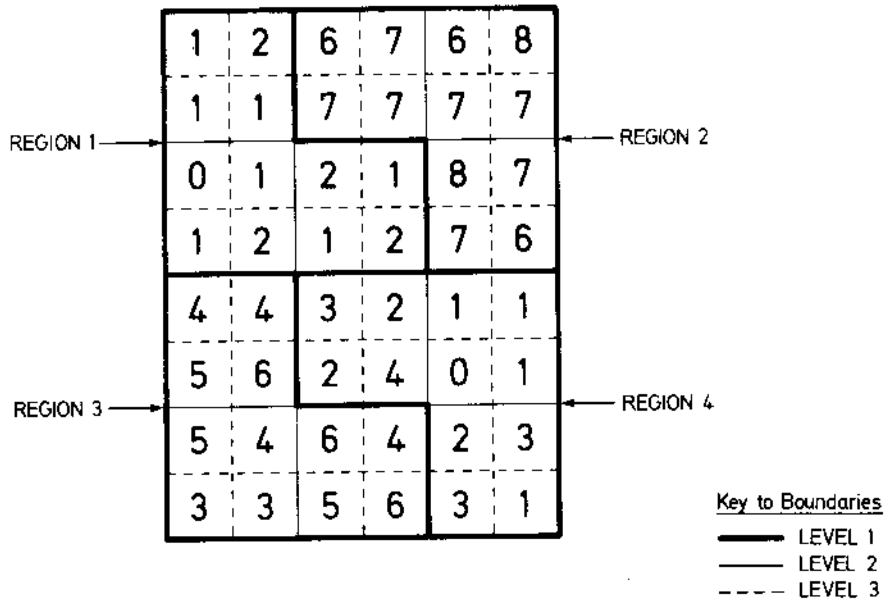


Fig. 16 Artificial data for spatial components of variance calculations.

Table 12. Computational formulae for sums of squares and mean squares in the components of variance model (balanced hierarchy)

Level	Source of Variation	Sum of Squares (SS)	Degrees of Freedom	Mean Square (MS)
Regional	α	$SS_\alpha = \sum_i(\sum_{jk}\sum Y_{ijk})^2/JK - (\sum\sum_{ijk}\sum Y_{ijk})^2/N$	I-1	$MS_\alpha = SS_\alpha/(I-1)$
District	β	$SS_\beta = \sum_{ij}\sum(\sum_k Y_{ijk})^2/K - \sum_i(\sum_{jk}\sum Y_{ijk})^2/JK$	I(J-1)	$MS_\beta = SS_\beta/I\,(J-1)$
Parish	γ	$SS_\gamma = \sum\sum_{ijk}\sum Y^2_{ijk} - \sum_{ij}\sum(\sum_k Y_{ijk})^2/K$	IJ(K-1)	$MS_\gamma = SS_\gamma/IJ(K-1)$
Total	-	$SS_T = \sum\sum_{ijk}\sum Y^2_{ijk} - (\sum\sum_{ijk}\sum Y_{ijk})^2/N$	N-1	

$i = 1,I;\quad j = 1,J$ (since all $J_i = J$); $\quad k = 1,K$ (since all $K_{ij} = K$)

Table 13. Tableau of Quantities for Components of Variance Calculations

DISTRICTS WITHIN REGIONS

	Y	Y^2	Y	Y^2	Y	Y^2	
REGION 1	1 2 1 1	1 4 1 1	0 1 1 2	0 1 1 4	2 1 1 2	4 1 1 4	$\bar{Y}_1 = 1.25$
ΣY ΣY^2 $(\Sigma Y)^2$	5 25	7	4 16	6	6 36	10	$\Sigma\Sigma Y = 15$ $(\Sigma\Sigma Y)^2 = 225$
REGION 2	6 7 7 7	36 49 49 49	6 8 7 7	36 64 49 49	8 7 7 6	64 49 49 36	$\bar{Y}_{2+} = 6.92$
ΣY ΣY^2 $(\Sigma Y)^2$	27 729	183	28 784	198	28 784	198	$\Sigma\Sigma Y = 83$ $(\Sigma\Sigma Y)^2 = 6889$
REGION 3	4 4 5 6	16 16 25 36	5 4 3 3	25 16 9 9	6 4 5 6	36 16 25 36	$\bar{Y}_{3+} = 4.58$
ΣY ΣY^2 $(\Sigma Y)^2$	19 361	93	15 225	59	21 441	113	$\Sigma\Sigma Y = 55$ $(\Sigma\Sigma Y)^2 = 3025$
REGION 4	3 2 2 4	9 4 4 16	1 1 0 1	1 1 0 1	2 3 3 1	4 9 9 1	$\bar{Y}_{4+} = 11.50$
ΣY ΣY^2 $(\Sigma Y)^2$	11 121	33	3 9	3	9 81	23	$\Sigma\Sigma Y = 23$ $(\Sigma\Sigma Y)^2 = 529$

$\Sigma\Sigma\Sigma Y = 176$ $\bar{Y}_{++} = 3.67$
$\Sigma\Sigma\Sigma Y^2 = 926$

N.B. In terms of Figure 16, regions are at level 1, districts at level 2.

As shown in Table 14, a very large proportion of the total variation in the observations (86.82%) can be attributed to differences at the regional level, with minor contributions from the district and parish levels.

We now describe how variance components for each spatial scale are obtained. Given that the true variance components at the regional, district and parish levels are represented by σ^2_α, σ^2_β, and σ^2_γ respectively, expected mean squares corresponding to the variation at each level are given (starting from

Table 14. Sums of squares and mean squares for the observations of Figure 16

Scale Level	Source	Sum of squares	% Total sum of squares	df	Mean square
1	α_i	243.67	86.82	3	81.2233
2	β_{ij}	14.00	4.99	8	1.7500
3	γ_{ijk}	23.00	8.19	36	0.6389
0	Total	280.67	100.00	47	

the highest level in the hierarchy) by

$$E(MS_\alpha) = \sigma^2_\gamma + K\sigma^2_\beta + JK\sigma^2_\alpha$$
$$E(MS_\beta) = \sigma^2_\gamma + K\sigma^2_\beta \qquad (39)$$
$$E(MS_\gamma) = \sigma^2_\gamma$$

Substituting sample estimators for population parameters throughout in (39), we obtain

$$MS_\alpha = s^2_\gamma + Ks^2_\beta + JKs^2_\alpha$$
$$MS_\beta = s^2_\gamma + Ks^2_\beta \qquad (40)$$
$$MS_\gamma = s^2_\gamma$$

from the formulae shown in Table 12. It is clear that an estimate of the variance component at any given level is obtained by subtracting from its mean square the mean square of the level below, and dividing through by the appropriate constant.

Applying such logic to the results in Table 14, we obtain

$$s^2_\alpha = \frac{1}{JK}(MS_\alpha - MS_\beta) = \frac{1}{12}(81.2233\text{-}1.7500) = \underline{6.623}$$
$$s^2_\beta = \frac{1}{K}(MS_\beta - MS_\gamma) = \frac{1}{4}(1.7500\text{-}0.6389) = \underline{0.278}$$
$$s^2_\gamma = MS_\gamma = \underline{0.6389}$$

In this case the results confirm our earlier suspicions that the bulk of the variation is concentrated at the regional level. The logic can be readily extended to estimate variance components in a hierarchy with more levels (e.g. Haggett et al (1977, 389)). For unbalanced hierarchies, formulae for scale variance components are given in Moellering and Tobler (1972, 41).

For spatial hierarchies in which there is complete enumeration of the spatial units in the study area and complete enumeration of the phenomenon of interest, such results may be interpreted as indicating the absolute and relative magnitudes of the variances that would characterise the effects

generated by all possible populations drawn from a 'superpopulation'.

4) Hypothesis Testing

The relationships between expected mean squares described in (39) may be used to test the hypotheses

$$H_o: \sigma^2_\beta = 0$$
$$H_o: \sigma^2_\alpha = 0$$

by comparing ratios of mean squares with the appropriate F distribution under H_o. For instance, an F ratio approximately equal to one is expected if $\sigma^2_\beta = 0$ and we calculate MS_β/MS_γ, and also if $\sigma^2_\alpha = 0$ and we calculate MS_α/MS_β. For the first hypothesis

$$F_{8,36} = MS_\beta/MS_\gamma = 1.7500/0.6389 = \underline{2.739}$$

and for the second

$$F_{3,8} = MS_\alpha/MS_\beta = 81.2233/1.7500 = \underline{46.413}$$

Both ratios are significant at the 5% level, the former only just. Although most of the variance occurs at the α (regional) level, there is evidence also for a rather lesser degree of variation at the β (district) level.

5) Limitations

Some notes of caution must be sounded if a nested Type II or random-effects ANOVA model is to be employed.

First, the variance component estimated at any given level may be negative. Where this occurs, it is customary to set the estimate to zero e.g. Dunn and Clark (1974, 206); Haggett et al (1977, 389). This is a satisfactory procedure if the negative value can plausibly be attributed to sampling error or rounding error in calculations. However, correlation between measurements for spatial units within a higher level spatial unit may also produce such a result, and inflate or deflate the values of variance components (Snedecor and Cochran, 1967, 294-296). For example, high correlations between measurements obtained by placing quadrats over observations on disease levels in vegetation may give this effect - once again spatial autocorrelation rears its head as a significant problem for statistical analysis in geography Negative variance components estimates will also be obtained where, for example the average variation between districts within regions is greater than the average variation between regions. A simple spatial hierarchy for which this is the case is shown in Figure 17(a), and the property is clearly evident in Figure 17(b). The reader may care to verify that $s^2_\alpha = -3.5$. Results of this kind have led researchers to question the utility of variance components analysis. At the very least, investigators should ask themselves what a negative value means as well as or instead of simply setting it equal to zero.

A second point is that the spatial hierarchy must be fully nested i.e. a spatial unit at one level must be fully contained in one and only one unit at the spatial level above it, each spatial unit at one level must contain at least one unit at the next lower level, and 'local inversions' are not permitted e.g. in most parts of the United States the spatial hierarchy of political units is: county, congressional district, state, but in some metropolitan areas runs: congressional district, county, state.

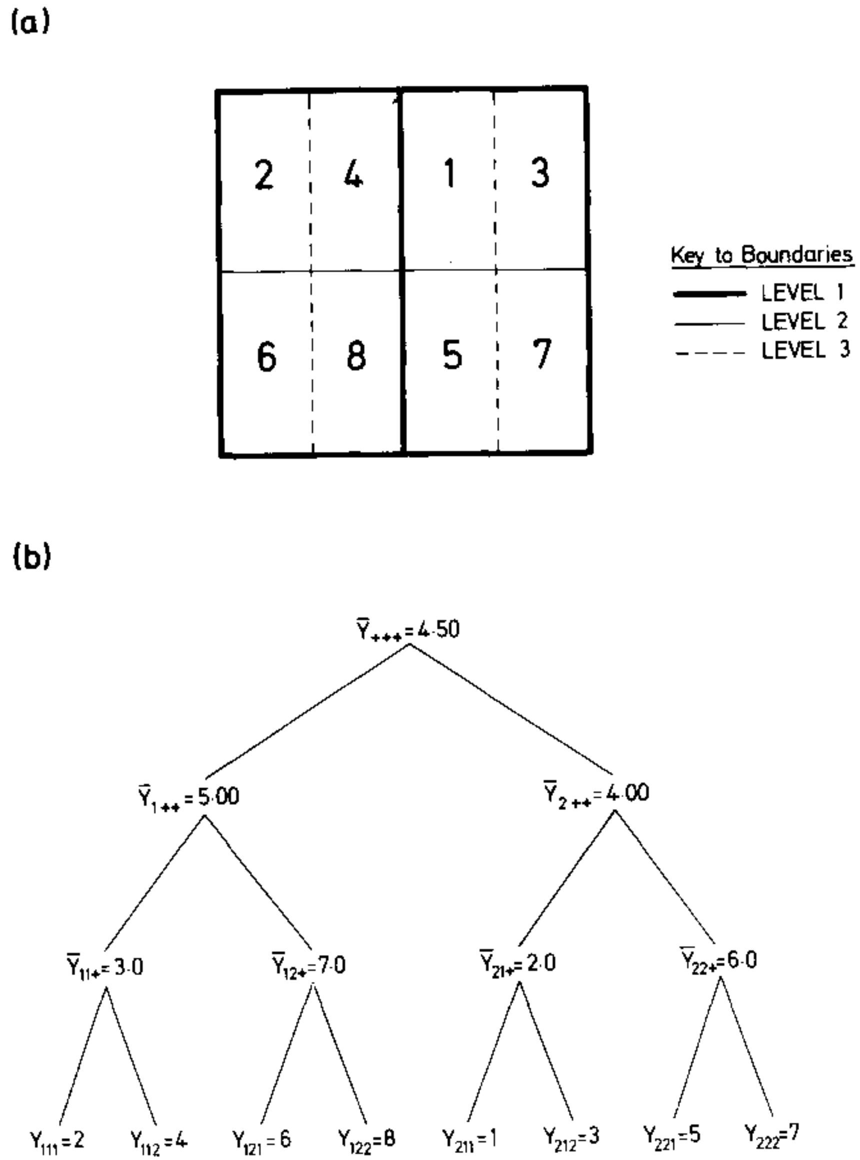


Fig. 17 Spatial hierarchy giving rise to negative variance components estimates (a) Map (b) Tree diagram.

Finally, we should point out that Fourier and spectral analysis also identify variance components at different spatial scales, and that hierarchical analysis of variance is more restricted in that a discontinuity is assumed between each spatial level. An extension of the spatial model to include a time component is briefly discussed in Haggett et al (1977, 390).

VII APPLICATIONS AND FURTHER READING

Geographers have paid little attention to the principles of experimental design, apart from those aspects related to external validity where questions of generalisation to other phenomena or situations arise. Early contributions on experimental design in geography are those by Haggett (1964; 1965, 299-303) and Chorley (1966, 335-40) and these have been extended in Haggett et al (1977, 270-279). Gentle introductions, designed for social scientists, are provided by Nachmias and Nachmias (1976, Ch. 3) and Orenstein and Phillips (1978, Ch. 1), and Susser (1973, Chs. 6-7) gives a more advanced but readable discussion in the context of epidemiological studies. Detailed treatment of experimental and quasi-experimental designs may be found in Campbell and Stanley (1966), and many examples based on fieldwork in the social sciences in Cook and Campbell (1979). Examples of interest to geographers are also given in Tanur et al (1972).

Applications of the ANOVA model appear relatively few and far between in studies by geographers, perhaps because their attitude generally accords with that caricatured by Draper and Smith (1966, 243) in their comment that if a researcher using the technique is asked 'what model are you considering?', the reply is frequently 'I am not considering one - I am using analysis of variance'. Just as experimental design has been neglected, so a formal statement of the ANOVA model is almost impossible to find in any geographic text except Norcliffe (1977, 159).

In the late 1950s, Zobler (1958) used the technique to assign areas to given regional classes, but found no imitators. Work on the internal structure of cities produced a study of variation in sectoral land values in Topeka by Knos (1962) and a two-way analysis which simultaneously considered differences in economic status between zones concentric to and radiating from the city centre in Toronto (Murdie, 1969). In this connection the technical discussion between Johnston (1970) and Murdie (1970) is instructive. It is worth noting that the apparently conflicting results obtained by Johnston and Murdie could well be attributed to the facts that i) there is no simple relationship between overall and individual comparisons and ii) results may change because the specified model has been changed. Both considerations seem to apply in the Murdie-Johnston case.

The influence of the three factors income, age and white/non-white ratio on housing values in Washington DC was studied by Davis (1971), who found interactions involving pairs of factors and all three factors. A further example which reveals a three-way interaction is well-illustrated in Lee (1970). For other applications of the fixed-effects ANOVA model, see Carter and Chorley (1961) and Doornkamp and King (1971), and for more recent geographical applications see Bryan (1974), Herzog et al (1974), Lloyd (1977) and Potter (1979).

Geological texts provide more sophisticated treatment of ANOVA together with many examples. Miller and Kahn (1962, Ch. 7) discuss fixed- and random-effects models and consider techniques for analysing hierarchical and non-hierarchical data sets. Krumbein and Graybill (1965, Ch. 9) in addition cover mixed models and provide an extensive set of annotated references.

A thorough but advanced treatment of ANOVA may be found in Scheffé (1959) and excellent discussions are also available in Snedecor and Cochran (1967, Chs. 10-12, 16) and Dunn and Clark (1974, Chs. 5-9).

Turning to the Analysis of Variance of spatial data, the work in plant ecology discussed by Greig-Smith (1964, 85-93) is of considerable interest. Several examples are given of plots of mean square against block size (each block consisting of 1,2,4, ... quadrats), and the treatment of interpretation problems is quite detailed. Miller and Kahn (1962, 418-19) briefly describe hierarchical methods applied to geological mapping, and refer particularly to Olsson and Potter (1954) and Potter and Siever (1955, 1956). As already mentioned (p.37), Miller and Kahn discuss adaptation of the two-way ANOVA design to deal with spatial data and describe how the sampling grid interval may be varied (along either axis) according to the results obtained so as to distinguish 'regional' from 'local' trends. A nested spatial design which then employs the components of variance model to analyse data on voting patterns in the United States may be found in Stokes (1965).

VIII EXTENSIONS, CONNECTIONS WITH OTHER TECHNIQUES, AND COMPUTER ROUTINES

Extensions to ANOVA, and relationships with other techniques, are more readily described if the technique is located in a classification of multivariate problems by type (in terms of level of measurement) of response and explanatory variable involved (Table 15). Many of the techniques in the row of the table representing continuous response variables are versions of the General Linear Model (Silk, 1981) - regression analysis belongs in cell (a), the Analysis of Covariance (ANCOVA) in cell (b), and ANOVA in cell (c).

If we carry out a regression analysis in which all the independent variables are categorical i.e. all are dummy variables, then the technique also belongs in cell (c) with ANOVA. This feature is noteworthy when we are faced with a situation in which the number of observations varies from cell to cell in a 2- or higher-way fixed-effects ANOVA. As mentioned earlier (p. 38), the estimation procedure on which formulae in Section IV are based is no longer appropriate. The various sum of squares (SS) calculated will not add to yield the total SS because the main effects will not normally be independent of each other, nor will interaction effects be independent of the main effects.

Table 15. Classification of Statistical Techniques

		Explanatory Variables		
		Continuous	Mixed	Categorical
Response Variables	Continuous	(a)	(b)	(c)
	Categorical	(d)	(e)	(f)

Three different methods for partitioning the total SS are detailed by Nie et al (1975, 405-408), but all can be implemented by carrying out a multiple regression analysis in which all the independent variables are dummy variables. For detailed discussion of specification of variables, see Draper and Smith (1966, Ch. 9) and Goldberger (1964).

ANCOVA can be regarded either as an ANOVA problem to which one or more continuous independent variables are introduced - and this is the way in which it is treated in most texts e.g. Snedecor and Cochran (1967, Ch. 14), Silk (1979a) - or as a regression analysis to which dummy variables have been added. For example, it might be of interest to include a continuous variable representing distance or travel time to each shop together with the factors representing life-cycle stage and car availability in Bowlby's study. Relationships between ANOVA and ANCOVA are explored in detail in Silk (1979a), as is the related topic of comparison of regression lines. Johnston (1978, Ch. 4) also discusses some of these interrelationships.

The concepts and notation of ANOVA, particularly those associated with effects and interaction, are also very useful if one is trying to understand the logic underlying new forms of categorical data analysis, many of which belong in cell (f) of Table 15. If the research design produces one entry per cell, each of which represents an absolute frequency, then such techniques of analysis are appropriate. We must of course bear in mind Fienberg's (1977,3) caution that in fixed-effects ANOVA the aim is to assess effects of explanatory variables on a dependent variable and to partition overall variability, whereas this is not necessarily the case in multi-dimensional contingency table analysis. For a review of these techniques see Wrigley (1979).

Finally, we must mention two widely-employed and simple non-parametric techniques for ANOVA. Both assume ordinal response variables and a single categorical explanatory variable or factor and so belong in cell (f) of Table 15. The Kruskal-Wallis procedure assumes two or more independent samples and the Friedmann procedure two or more related samples. For a description of the first procedure, together with worked examples, see Hammond and McCullagh (1978, 213-217) or Silk (1979b,192-5), and for details of the second Lewis (1977, 156-160). Both procedures are discussed in detail in Siegel (1956).

Of the widely available standard computer package programs, both the Statistical Package for the Social Sciences (SPSS) (Nie et al (1975, Chapter 22) and the library of Biomedical (BMD) Computer Routines (Dixon (1968, 597-704)) have programs for dealing with one-way and factorial designs, including a facility for carrying out multiple comparisons. The BMD package also provides routines for analysis of hierarchical or nested designs. Of the other packages - both of which originate from the Rothamsted Experimental Station - GENSTAT (Alvey et al (1977)) and GLIM (Baker and Nelder (1978)) give procedures which cover most of the ground we have mentioned, but the manuals are much less easy to understand and use. Discussion of the estimation procedures involved, and of the circumstances under which various techniques should be employed, is particularly good in the SPSS manual.

IX CONCLUSION

The Analysis of Variance has not been widely used by geographers. However, if it is realised that there is an Analysis of Variance model just as there is, say, a regression model, and that both fall within the framework of the General Linear Model, then the technique becomes far more attractive as both an estimation and hypothesis-testing procedure. Well-tried algorithms and computer packages may be called upon, and this is also the case if it is desired to analyse data in a spatial hierarchy.

BIBLIOGRAPHY

A. TECHNICAL

Experimental Design

Campbell, D., and Stanley, J. (1963), Experimental designs and quasi-experimental designs for teaching and research. in: *Handbook of research on teaching,* (ed) N. Gage, (Skokie, Illinois: Rand McNally), Ch.5.

Chorley, R.J. (1966), The application of statistical methods to geomorphology. in: *Essays in Geomorphology,* (ed) G.H. Dury, (London: Heinemann), 275-387.

Cook, T.D., and Campbell, D.T. (1979), *Quasi-experimentation: design and analysis for field settings.* (Chicago: Rand McNally).

Haggett, P. (1964), Regional and local components in the distribution of forested areas in southeast Brazil: a multivariate approach. *Geographical Journal,* 130, 365-380.

Nachmias, D., and Nachmias, C. (1976), *Research methods in the social sciences.* (London: Arnold).

Orenstein, A., and Phillips, W.R.F. (1978), *Understanding social research* (Boston: Allyn and Bacon Inc.). Sussex.

Susser, M.W. (1977), *Causal thinking in the health sciences.* (New York: Oxford University Press).

Tanur, J.M., Mosteller, F., Kruskal, W.H., Link, R.F., Pieters, R.S., and Rising, G.R. (1972), *Statistics: a guide to the unknown.* (San Francisco: Holden-Day, Inc.).

Statistical Theory

Blalock, H.M. (Jr.) (1972), *Social statistics.* (New York: McGraw-Hill)

Cliff, A.D. (1973), A note on statistical hypothesis testing. *Area,* 5(3), 240.

Cliff, A.D., and Ord, J.K. (1973), *Spatial autocorrelation.* (London: Pion).

Draper, N.R., and Smith, H. (1966), *Applied regression analysis.* (New York: Wiley).

Dunn, O.J., and Clark, V.A. (1974), *Applied statistics: analysis of variance and regression.* (New York: Wiley).

Ferguson, R. (1978)', *Linear regression in geography.* Concepts and techniques in modern geography, 15. (Norwich: Geo Abstracts Ltd.).

Fienberg, S.E. (1977), *The analysis of cross-classified categorical data.* (Cambridge, Mass: MIT Press).

Goldberger, A.S. (1964), *Econometric theory.* (New York: Wiley).

Hammond, R., and McCullagh, P.S. (1978), *Quantitative techniques in geography: an introduction.* (Oxford: Clarendon Press).

Horton, R.L. (1978), *The general linear model.* (New York: McGraw-Hill).

Johnston, R.J. (1978), *Multivariate statistical analysis in geography: a primer on the general linear model.* (London: Longman).

Kirk, R.E. (1968), *Experimental design: procedures for the behavioural scientist.* (Belmont, California: Brooks/Cole).

Mead, R., and Pike, D.J. (1975), A review of response surface methodology from a biometric viewpoint. *Biometrics,* 31, 803-851.

Norcliffe, G.B. (1977), *Inferential statistics for geographers.* (London: Hutchinson).

O'Neill, R., and Wetherill, G.B. (1971), The present state of multiple comparison methods. *Journal of the Royal Statistical Society* B, 33(2), 218-250.

Openshaw, S. (1977), Optimal zoning systems for spatial interaction models. *Environment and Planning A,* 9, 169-184.

Scheffé, H. (1959), *The analysis of variance.* (New York: Wiley).

Siegel, S. (1956), *Non-parametric statistics for the behavioural sciences.* (New York: McGraw-Hill).

Silk, J.A. (1979a), *Analysis of covariance and comparison of regression lines.* Concepts and techniques in modern geography, 20. (Norwich: Geo Abstracts Ltd.).

Silk, J.A. (1979b), *Statistical concepts in geography.* (London: Allen and Unwin).

Silk, J.A. (1981), The general linear model. in: *Quantitative geography in Britain: retrospect and prospect,* (eds) R.J. Bennett and N. Wrigley, (London: Routledge and Kegan Paul) (forthcoming).

Snedecor, G.W., and Cochran, W.G. (1967), *Statistical methods.* (Ames, Iowa: Iowa State University Press).

Unwin, D. (1975), *An introduction to trend surface analysis. Concepts* and techniques in modern geography, 5. (Norwich: Geo Abstracts Ltd.).

Wrigley, N. (1979), Developments in the statistical analysis of categorical data. *Progress in Human Geography, 3(3),* 315-355.

Computer Routines

Alvey, N.G., et al (1977), *GENSTAT: A general statistical program.* (Rothamsted: Rothamsted Experimental Station).

Baker, R.J., and Nelder, J.A. (1978), *The GLIM system manual: Release 3,* (Oxford: Numerical Algorithms Group).

Dixon, W.J. (ed) (1968), *Biomedical computer programs,* (California: University of California Press).

Nie, N., Bent, D.H., and Hull,- C.H. (1975), *Statistical package for the social sciences* (SPSS), (New York: McGraw-Hill).

B. APPLICATIONS

Bowlby, S.R. (1979), *Mobility and information: a study of knowledge of grocery shopping opportunities.* Final report to the SSRC. (Reading: Department of Geography, University of Reading).

Bryan, R.B. (1974), A simulated rainfall test for the prediction of soil erodibility. *Zeitschrift fur Geomorphologie* (Supplement), 21, 138-150.

Carter, C.A., and Chorley, R.J. (1961), Early slope development in an expanding stream system. *Geological Magazine,* 98, 117-130.

Chorley, R.J., Stoddart, D.R., Haggett, P., and Slaymaker, H.O., (1966), Regional and local components in the areal distribution of surface sand-facies in the Breckland, eastern England. *Journal of Sedimentary Petrology,* 36, 209-220.

Davis, J.T. (1971), Sources of variation in housing values in Washington, D.C. *Geographical Analysis,* January, 63-76.

Doornkamp, J.C., and King, C.A.M. (1971), *Numerical analysis in geomorphology,* (London: Arnold).

Greig-Smith, P. (1964), *Quantitative plant ecology.* (London: Butterworths), 2nd edition.

Haggett, P. (1963), Regional and local components in land-use sampling: a case study from the Brazilian Triangulo. *Erdkunde,* 17, 108-114.

Haggett, P. (1964), Regional and local components in the distribution of forested areas in southeast Brazil: a multivariate approach. *Geographical Journal,* 130, 365-380.

Haggett, P. (1965), Scale components in geographical problems. in: *Frontiers in Geographical Teaching,* (eds) R.J. Chorley and P. Haggett, (London: Methuen), 164-185.

Haggett, P., Cliff, A.D., and Frey, A.E. (1977), *Locational methods.* Volume 2 of *Locational analysis in human geography.* (London: Arnold), 2nd edition.

Herzog, T.R., Kaplan, S., and Kaplan, R. (1974), The prediction of preference for familiar urban places. *Environment and Behaviour,* 8(4), 627-645.

Johnston, R.J. (1970), On spatial patterns in the residential structure of cities. *Canadian Geographer,* XIV(4), 361-367.

Krumbein, W.C. (1956), Regional and local components in facies maps. *Bulletin of the American Association of Petroleum Geologists,* 40, 2163-2194.

Krumbein, W.C., and Graybill, F.A. (1965), *An introduction to statistical models in geology.* (New York: McGraw-Hill).

Lee, T.R. (1970), Perceived distance as a function of direction in the city. *Environment and Behaviour,2,* 40-51.

Lloyd, R.E. (1977), Consumer behaviour after migration: a reassessment process. *Economic Geography, 53(1),* 15-27.

Melton, M.A. (1960)•, Intravalley variation in slope angles related to microclimate and erosional environment. *Bulletin of the Geological Society of America,* 71, 133-144.

Miller, R.L., and Kahn, J.S. (1962), *Statistical analysis in the geological sciences.* (New York: Wiley).

Moellering, H., and Tobler, W. (1972), Geographical variances. *Geographical Analysis,* January, 34-50.

Murdie, R.A. (1969), ***Factorial ecology of metropolitan Toronto, 1951-1961.*** Research Paper No. 116, (Chicago: Dept. of Geography, University of Chicago).

Murdie, R.A. (1970), A reply to On spatial patterns in the residential structure of cities. ***Canadian Geographer,*** XIV(4), 367-369.

Olson, E.C., and Potter, P.E. (1954), Variance components of cross-bedding direction in some basal Pennsylvanian sandstones of the east interior basin. ***Journal of Geology,*** 62, 26-49.

Potter, P.E., and Siever, R. (1955), A comparative study of Upper Chester and Lower Pennsylvanian stratigraphic variability. ***Journal** of Geology,* 63, 429-451.

Potter, R.B. (1979), Perception of urban retailing facilities: an analysis of consumer information fields. *Geografiska Annaler,* 61B(1), 19-29.

Stokes, D.E. (1965), A variance components model of political effects. in: *Mathematical applications in political science,* (ed) J.M. Claunch, (Dallas: Arnold Foundation, Southern Methodist University).

Schumm, S.A. (1956), Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey. Bulletin *of* the *Geological Society of America,* 67(5), 597-646.

Thomas, R.W. (1977), *An introduction to quadrat analysis.* Concepts and techniques in modern geography, 12, (Norwich: Geo Abstracts Ltd.).

Thomas, R.W., and Huggett, R.J. (1980), *Modelling in geography: a mathematical* approach. (London: Harper and Row).

Zobler, L. (1958), Decision making in regional construction. *Annals of the Association of American Geographers,* 48(2), 140-148.

25. Directional statistics - G.L. Gaile & J.E. Burt
26. Potential models in human geography - D.C. Rich
27. Causal modelling: the Simon-Blalock approach - D.G. Pringle
28. Statistical forecasting - R.J. Bennett
29. The British Census - J.C. Dewdney
30. The analysis of variance - J. Silk

This series, Concepts and Techniques in Modern Geography is produced by the Study Group in Quantitative Methods, the Institute of British Geographers.
For details of membership of the Study Group, write to the Institute of British Geographers, 1 Kensington Gore London, S.W.7.
The series is published by Geo Abstracts, University of East Anglia, Norwich, NR4 7TJ, to whom all other enquiries should be addressed.