APPLICATIONS OF ANALYSIS OF VARIANCE

by

M. S. Bartlett

November 1, 1946

Institute of Statistics
Mimeo Series #5
For limited distribution

CONTENTS.

I. BASIC PRINCIPLES

1. Introductory remarks.

Most working statisticians are by now so familiar with R. A. Fisher's technique of 'analysis of variance' that I doubt whether they will discover much new in my contribution to this conference. It was, however, suggested to me that some discussion and actual examples illustrating the technique, with particular reference to the two somewhat distinct problems of estimating the magnitude of

(i) fixed linear effects that may be present in statistical data, and

(ii) variable linear effects giving rise to additional variance components,

would be complementary to a theoretical paper by Wald on the same topic.

The two problems may sometimes have been confused because of the very name 'analysis of variance', an 'analysis of variance' table referring directly not to (ii), but to (i), and, as has been commented by other writers, might perhaps less ambiguously be called 'quadratic analysis' or 'analysis of the sum of squares'.* In problems of type (ii) the estimation of the different variance components is a further and separate procedure from the calculation of the primary analysis of variance table. Examples of the two problems (i) and (ii) are discussed in sections II and III of this paper respectively, but before coming to these I would like to review very briefly the basic principles of analysis of variance, both for reference and also to justify some of the names I shall use to christen different types of data. Some general acquaintance with analysis of variance technique is assumed.

2. Algebra of Analysis of Variance.

I personally find it convenient theoretically to employ the geometrical or vectorial representation of a sample of observations, and have previously discussed analysis of variance from this point of view (1). The sample of n observations is represented in this picture by a vector with n

---

* E. J. G. Pitman once suggested 'analysis of squariance', but this phrase frankly appals me!

components corresponding to the <u>n</u> observations, and the analysis of variance table merely corresponds to an appropriate analysis of the total sum of squares of the observations, i.e., of the square of the length of the vector, into relevant perpendicular or orthogonal components, whose squares then necessarily add to the total square. The orthogonal components lie in directions in the sample space so chosen to correspond to the isolation of the fixed linear effects referred to under (i), and it is a well-known feature of a good analysis of variance design that this can conveniently be done. To be more explicit, let the row vector or matrix S denote the n observations $(x_1 \ . \ . \ . \ x_n)$, and the corresponding column vector S'. Then the isolation of the general mean corresponds to segregation of the component of S along the direction given by the vector $Z \equiv (1, 1, \ . \ . \ . \ . \ . \ . \ 1)$. The linear component of S expressible in terms of <u>any</u> vector Z is obtained by writing S in the form

$$S = bZ + V \tag{1}$$

where V is orthogonal to Z, or $VZ' = (S - bZ)Z' = 0$, or

$$b = (SZ')(ZZ')^{-1} \tag{2}$$

Thus in this trivial example where $Z \equiv (1,1, . \ . \ . \ . \ 1)$, we have for the analysis of variance table, by squaring* both sides of (1),

$$SS' = (SZ')(ZZ')^{-1}(ZS') + VV' \tag{3}$$

or

$$\Sigma x_r^2 = n\bar{x}^2 + \Sigma(x_r - \bar{x})^2 \tag{4}$$

corresponding to the special case of (1),

$$(x_r) = (\bar{x}) + (x_r - \bar{x}) \tag{5}$$

It is an obvious consequence of the linear form of (1) that any 'real' linear dependence on Z, which would remain after random and unbiased effects were removed, corresponding to the expectation equation

$$\beta Z \equiv E\{S\} = E\{bZ\} + E\{V\} = E\{b\} Z, \tag{6}$$

is entirely segregated into the component parallel to Z, since no linear superposition of a component parallel to Z can affect any components of S orthogonal to Z. For example, if we add any constant quantity to (5) it affects only the general mean component, and not the residual component $(x_r - \bar{x})$. Conversely,

---

* i.e. Row by column matrix multiplication.

if such superposition does not affect any components of S, these must be orthogonal to Z, a useful practical rule for verifying if such orthogonality holds. It is also readily shown that, <u>if the observations S are uncorrelated* and all have the same variance σ^2</u>, the square of the component along Z has expectation

$$\beta^2(ZZ') + \sigma^2. \qquad (7)$$

More generally, we may not want to isolate each of individual components such as bZ in (1), but to separate the group corresponding to any linearly independent set of P vectors $Z_1 \ldots Z_p$, which we may denote by a column matrix Z of such vectors $Z_1 \ldots Z_p$. Equations (1), (2), (3), and (6) still hold provided that b is now interpreted as a row matrix of coefficients $b_1 \ldots b_p$; equation (7) generalizes to

$$\beta(ZZ')\beta' + p\sigma^2 \qquad (8)$$

The observational field of multiple regression where the p vectors $Z_1 \ldots Z_p$ are not in general orthogonal, is one example where the component βZ is removed <u>en bloc</u> in this way in the analysis of variance table, but the same principle applies when, say, treatments or blocks are separated in a randomized block layout. The further reduction from the simultaneous set Z to an equivalent set of mutually orthogonal vectors is of course always possible, and in fact becomes relevant to the numerical method of reduction of the equations of estimation for β, and the evaluation of standard errors. It is made particular use of in the fitting of trends by the method of orthogonal polynomials.

One further remark on the interpretation of Z. It often happens that Z is not orthogonal to another vector or group of vectors W corresponding to the other real effects in the data. When this occurs, the effects associated with Z and W are said to be confounded, and if we ignore the effects associated with W they will interfere with the estimation of the effects associated with Z. The only way to free the Z effects from W is to consider the components of Z orthogonal to W, and in the above equations Z should be so interpreted. It is customary to eliminate the effect of the general mean in this way, and it is an interesting property of randomized blocks, Latin squares, and other familiar orthogonal layouts that the orthogonality property holds <u>after</u>, and not <u>before</u>, such elimination. For example, in a randomized block design, both treatment means and block means are affected by the general mean, and to this extent are confounded; but the elimination of the general mean implies that we can only consider <u>differences</u> of treatment means, or of block means, and these two

---

* It is usual to assume them independent, but absence of correlation is sufficient here.

sets of differences are orthogonal, so that no further adjustment of treatment means for block means is necessary. The design of experiments, arranged to give orthogonal components corresponding to all, or nearly all, of the effects we are investigating, has by now of course reached an advanced stage (see, for example, Yates (7)). In connection with the present viewpoint, it is instructive for the student to consider what vectors Z will isolate the effects to be estimated, beginning with simple randomized blocks and proceeding to the more complicated factorial experiments.

The above discussion comprises in shorthand form the structural basis of analysis of variance in quite general terms and includes all forms of multiple regression, least squares and Fourier analysis. It so far covers only problem (i), which I regard as the primary analysis of variance problem. Before any illustrations of these principles, there is, however, a point still to be treated under (i), the question of accuracy and significance tests.

3. Accuracy and significance tests.

We had in § 2 formula (6) for unbiased deviations,

$$E\{b\} = \beta \tag{9}$$

and under the assumptions used to obtain (7) or (8), we obtain by taking the expectation of bb', the result

$$E\{(b-\beta)(b-\beta)'\} = \sigma^2(ZZ')^{-1} \tag{10}$$

which gives the variances and covariances of the coefficients $b_1 \ldots b_p$ in terms of the variance σ^2. We shall for convenience refer to a sample of data where the observations are unbiased and uncorrelated, and have apart from the fixed linear effects to be estimated, a constant variance σ^2, as a homogeneous sample, and problem (i), which is to be illustrated in section II, represents analysis of the above type on such homogeneous data. It is known that the estimates b, which are the usual least-squares estimates, have on this assumption minimum variance of any linear combinations of the observations.

To obtain exact tests of signifiance, further assumptions are needed. The usual one is that the observations are independent and normally distributed; this ensures (c.f. 1) that the chance position of the vector S has a direction in the sample space randomly orientated, and the random angle it makes with the fixed 'plane' of vectors $Z_1 \ldots Z_p$ has a known distribution equivalent to Fisher's z distribution. This assumption is a less fundamental one than the one corresponding to the homogeneous sample, but is often a reasonable and certainly a convenient one to make. It will allow exact confidence or fiducial limits to be assigned to any of the β and also allows a more complete justification of the method of estimation implied above in terms of modern theories of estimation and tests of significance. If necessary, from the known distribution of the multiple correlation coefficient R between S and the set of vectors Z, <u>when the latter are fixed</u>, a distribution depending only on the 'true coefficient' ρ, where

$$R^2 = \frac{b(ZZ')b'}{SS'}, \qquad \rho^2 = \underset{n \to \infty}{\mathrm{Lt}} R^2, \tag{11}$$

confidence limits can be set for ρ^2, but this requirement in my opinion rarely occurs.

As a variant on the above normal assumption, advantage has been taken of the device of randomization in many statistical experiments, (a procedure carried out primarily with the important object of eliminating bias and lack of independence in the observations), to investigate tests of significance independent of the assumption of normality (see for example, Welch (9)). However, this can hardly be recommended as a routine procedure, and the assumption of normality is the one usually made. It may be noted that transformations of scale to make the assumption of constant variance* more reasonable sometimes have the effect of improving the normality of the distribution.

---

* Heterogeneous variance is referred to more generally later in this paper.

II. ANALYSIS OF HOMOGENEOUS DATA

4. Two War-time applications (Examples A and B).

In choosing one or two examples to illustrate straightforward analyses of variance of type (i), I have attempted to avoid the more well-worn illustrations, and the related examples A and B, which arose during the war, may be of some interest.

Example A. A certain type of rocket projectile was fired from a cylindrical barrel which was suspected, possibly owing to imperfect construction, of causing a fixed 'barrel deviation' relative to the projector, but could be turned round on its axis and fired in any fixed orientation to investigate this. In a firing trial of this kind, such a re-orientation of the cylindrical projector was made after every four rounds had been fired, and the theoretical basis of the analysis was an analysis of variance between and within such 'orientation' groups of rounds, together with a regression analysis of the mean of each group on two orthogonal unit vectors fixed in the projector perpendicular to its axis and serving as a coordinate system with respect to which the barrel deviation was measured. If these two vectors are denoted by $\vec{i}$ and $\vec{j}$, and the projector has been rotated a positive angle θ, ($\vec{j}$ corresponding to a positive anti-clockwise rotation of 90° from $\vec{i}$, and $\vec{i}$ initially pointing in the positive left-to-right lateral direction) the components of the unknown barrel deviation $b_1\vec{i} + b_2\vec{j}$ in the lateral and angle of sight directions will be

$$b_1 \operatorname{Cos} \theta - b_2 \operatorname{Sin} \theta, \quad b_1 \operatorname{Sin} \theta + b_2 \operatorname{Cos} \theta$$

respectively. The formal regression analysis can thus be carried out on two quantities (z_1) and (z_2), whose respective values for lateral and angle of sight projectile deviations (which tended to have the same variance) are

	z_1	z_2
Lateral	Cos θ	$-$Sin θ
Angle of sight	Sin θ	Cos θ

It is obviously convenient to choose a periodical set of values of θ, such as steps of 45° or 90°, so that z_1 and z_2 are not only mutually orthogonal but orthogonal to the general mean component. In the simple case of 90° steps, we thus have the values for z_1 and z_2:

Orientation	Lateral		Angle of Sight	
	z_1	z_2	z_1	z_2
$\theta = 0^\circ$	1	0	0	1
90°	0	-1	1	0
180°	-1	0	0	-1
270°	0	1	-1	0

In the analysis given below, the order was actually 0 = 270°, 90°, 0°, and 180°, and then repeated, the angles separated by 180° being kept as close together as practicable to safeguard against any slow change in the general mean.

The statistical analysis was as follows:--

TABLE I

Lateral deviations (x)					Angle of sight deviations (y)				
Totals of four rounds (T)	z_1	z_2			Totals of four rounds (T)	z_1	z_2		
-3.24	0	1	Σxz_1	8.77	-2.38	-1	0	Σyz_1	8.53
3.25	0	-1	Σxz_2	-13.50	2.31	1	0	Σyz_2	-9.32
2.71	1	0			-1.63	0	1		
-3.42	-1	0	Σx^2	23.569**	1.95	0	-1	Σy^2	15.793**
-3.83	0	1	$(\Sigma x)^2/32$	-	-2.29	-1	0	$(\Sigma y)^2/32$	0.001
3.18	0	-1		23.569	1.55	1	0		15.792
1.97	1	0			- 2.55	0	1		
-0.67	-1	0	$\Sigma T^2/4$	17.303	3.19	0	-1	$\Sigma T^2/4$	10.445
-0.05*			$(\Sigma x)^2/32$	-	0.15*			$(\Sigma y)^2/32$	0.001
				17.303					10.445

	D.F.	S.S.	M.S.		D.F.	S.S.	M.S.
Between groups	7	17.303	2.472	Between groups	7	10.445	1.493
Within groups	24	6.266	0.261	Within groups	24	5.347	0.223
Total	31	23.569	0.760	Total	31	15.792	0.509

*These totals are small since the deviations are already measured from their approximate mean values.

**Obtained from the original set of deviations.

Combined analysis			
Σxz_1	8.77	Σxz_2	-13.50
Σyz_1	8.53	Σyz_2	- 9.32
	17.30/32 = 0.54		-22.82/32 = 0.71

Barrel deviation: $0.54\,\vec{i} - 0.71\,\vec{j}$ (s.e. 0.09)

	D.F.	S.S.	M.S.
Barrel deviation	2	25.626	12.813
Remainder between groups	12) 16	2.122) 13.735	0.177) 0.229
Within groups	48)	11.613)	0.242)
Total	62	39.361	

Note: The abbreviations D.F., S.S., and M.S., denote as usual degrees of freedom, sum of squares and mean square.

It will be seen that the barrel deviation accounts for the significant variation, and the remainder term between groups was pooled with the 'within groups' term to obtain the residual 'dispersion', with estimated variance 0.229.

Example B. In connection with the investigation illustrated by Example A, some direct measurements were made of imperfections in the construction of the cylindrical projector; with reference to an axis of rotation which could be defined in relation to the cylinder, external gauge measurements* were made at 45° angular intervals at each of 8 equidistant sections along the cylinder. A Fourier analysis of the readings at each section was then taken up to the 2θ terms, according to the scheme of coefficients (α, say):--

Coefficients α

Angle θ	$\alpha = 1$	Sin θ	Cos θ	Sin2θ	Cos 2θ
0°	1	0	1	0	1
45°	1	0.707	0.707	1	0
90°	1	1	0	0	-1
135°	1	0.707	-0.707	-1	0
180°	1	0	-1	0	1
225°	1	-0.707	-0.707	1	0
270°	1	-1	0	0	-1
315°	1	-0.707	0.707	-1	0
$\Sigma\alpha^2$	8	4	4	4	4

These may be shown to determine approximately the mean radius (from an artificial zero), the two components of the shift of the centre relative to the axis of rotation taken; and the two components of 'ovality' of the section. Variation of these constants along the cylinder (no single section being of any special interest) was then summarized with the help of the orthogonal sets of coefficients (cf. Fisher and Yates, (6)) up to the second order:--

Coefficients β

Section	1	2	3	4	5	6	7	8	$\Sigma\beta^2$
Mean ($\beta = 1$)	1	1	1	1	1	1	1	1	8
Linear	7	5	3	1	-1	-3	-5	-7	168
Parabolic	7	1	-3	-5	-5	-3	1	7	168

* The laboratory measurements were of an exploratory character, and internal measurements on the cylinder would have been preferable. They are used here primarily as an illustration of the statistical method of analysis.

The vectors Z were then the complete sets of combinations $\alpha \times \beta$ making 5 x 3 = 15 orthogonal sets of coefficients in all. It is not certain that the remainder term, which was taken for the error variance, did not contain further systematic items, but it seemed reasonable to assume that the most important ones had been isolated. For such an investigation the resulting analysis of variance, with the omission of the general mean term, is given below; (it does not refer to the same projector as the data quoted in Example A).

TABLE II

			D.F.	M.S.
Change in radius along cylinder	Linear		1	224.3
	Parabolic		1	9.2
Shift of centre along cylinder	Mean	Sine	1	3043.0
		Cosine	1	21.1
	Linear	Sine	1	14.1
		Cosine	1	400.5
	Parabolic	Sine	1	165.0
		Cosine	1	14.0
Ovality along cylinder	Mean	Sine	1	15.1
		Cosine	1	47.5
	Linear	Sine	1	3.7
		Cosine	1	1.3
	Parabolic	Sine	1	7.7
		Cosine	1	45.6
Remainder term			49	16.4

The mean shift of the centre, and its linear trend along the cylinder, define the axis of rotation taken; apart from these terms, it will be seen that there is a significant linear change in radius, and a parabolic shift of the centre in a plane at right angles to the axis and the initial angular direction $\theta = 0$

5. Example C.

To stress the fundamental assumption of a homogeneous sample it may be worth quoting from a paper by Cochran, Autrey and Cannon (5) one further example from a more familiar field of experimentation. For full details reference should be made to the original paper, but briefly the experiment was a dairy cattle feeding trial of the change-over type involving three types of rations (denoted here by X, Y, and Z), and three consecutive periods I, II, and III. Groups of three cows received rations according to the following scheme:--

Period	Cow 1	Cow 2	Cow 3
I	X	Y	Z
II	Y	Z	X
III	Z	X	Y

The analysis of variance table of total ration consumption per cow per period for five such groups of these cows, the groups being chosen on the basis of expected similar yielding ability within a group was (Table 3 of the paper referred to):--

TABLE III

Total consumption of digestible nutrients per cow per period

	D.F.	S.S.	M.S.
Between groups	4	157,936	39,684
Between cows within groups	10	105,336	10,534
Between periods within groups	10	40,534	4,053
Between rations	2	533,869	266,934
Ration x group interactions	8	12,021	1,503 } 1,451
Remainder	10	14,100	1,410 }
Total	44	863,796	

The relevant point here (noted by the authors) is that there is no evidence of a ration x group interaction, so that it is reasonable to pool this term with the remainder to form a homogeneous error term against which the other items may be tested. If, however, the interactions term had been significant, my later comments (Part III) on the analysis of heterogeneous data would become relevant, and the interpretation less simple.

In the case of the milk yields, an interesting feature of the analysis was the detection of carry-over effects of the rations between the periods. Such carry-over effects, when allowed for, are not orthogonal to the direct ration effects in the above design, and in view of this confounding, the problems arise of the signifiance of the carry-over effects freed from direct effects and the estimation of the direct effects freed from the carry-over effects. For details the reader is referred to the original paper; the general principles involved are covered by my remarks on confounding towards the end of §2.

III. ANALYSIS OF HETEROGENEOUS DATA

6. Residual heterogeneity. Example D.

Once we depart from the assumption of 'homogeneity' as defined in I, a miscellany of situations may confront us, and the practical statistician has often to treat any specific problem on its own merits. Some definitions and results may, however, be useful. First of all, it is possible that although we know the minimum error variance realisable in our material, the latter manifests a residual heterogeneity which inflates this minimum variance. Such cases can still be treated by the methods of I and II, by treating this basic heterogeneity as the level of variability appropriate to our assumption of 'homogeneity'.

As an example, (Example D), consider the analysis of variance recorded in Table IV of (3). This represents an analysis of estimated variances, first transformed to the log scale to render their variability independent of the mean. The data consisted of estimates for each of three groups on 15 days, and the analysis of variance (for the $\log_{10}$ s.d.'s) was:--

TABLE IV

	D.F.	S.S.	M.S.
Groups	2	0.2667	0.1333
Days	14	0.1047	0.00748
Residual	28	0.1005	0.00359
Theoretical variance			0.0020
Total	44	0.4719	

The theoretical variance given corresponds to that expected from the logarithms (to base 10) of standard deviations each based on 48 d.f., but while the residual variance is significantly higher, it may be used as the basic variance with which to compare the other two items, both of which are significant.

7. 'Vertical' heterogeneity. Example E.

The further point which arises however, goes beyond the methods of I and II viz., if we consider the variation from day to day not for these particular 15 days, but as information on the day-to-day variance, what estimate is to be made? This assumption of random superposed effects giving rise to an extra variance component is what I have called problem (ii). I shall characterize it by what I shall call "vertical heterogeneity", owing to this feature of variance components piling up on each other in the different items of the analysis of variance table. If we look back at formula (8), we shall now make the rather special assumption that the Z form a group for which the corresponding coefficients β are not fixed, but vary in an uncorrelated and unbiased manner about zero with the same constant variance σ_β^2. Averaging equation (8) for variation in the β, we then obtain the general formula

$$\sigma_\beta^2(\text{trace } ZZ') + p\sigma^2 \qquad (12)$$

where trace ZZ' denotes the sum of the diagonal elements of ZZ'. Thus if we had n observations in p + 1 sets, and form an orthogonal analysis of variance table between and within sets, we have for the first vector of Z (after elimination of the general mean)

$$Z \equiv (1-\frac{n_1}{n}, 1-\frac{n_1}{n}, \ldots n_1 \text{ times}; -\frac{n_1}{n}, -\frac{n_1}{n}, \ldots; -\frac{n_1}{n}, -\frac{n_1}{n}, \ldots; \ldots)$$

and

$$\text{trace } ZZ' = \Sigma\left\{n_i(1-\frac{n_i}{n})^2 + (n-n_i)\frac{n_i^2}{n^2}\right\} = n - \Sigma\frac{n_i^2}{n},$$

in agreement with a result given by Winsor and Clarke (11). When all the n_i are equal to q, this becomes pq, and (12) becomes

$$p\left\{q\sigma_\beta^2 + \sigma^2\right\}, \qquad (13)$$

a familiar result which may be used to estimate the 'days' variance in Table IV above. The occurrence of the groups term is irrelevant, owing to the orthogonality of the design when the number of observations in each cell is constant (in this case, 1). We obtain the estimates:—

$$\begin{array}{rl} 3\sigma_\beta^2 + \sigma^2 & \sim 0.00748 \\ \sigma^2 & \sim 0.00359 \\ \hline \sigma_\beta^2 & \sim 0.00130 \end{array}$$

As far as I am aware, no solution for exact confidence limits (in Neyman's sense) for σ_β^2 is known, but in many cases we can obtain limits for σ_β^2/σ^2 if this is of interest. For our basic analysis was represented by

$$S = (b_e + \beta)Z + V, \qquad (14)$$

where for convenience the essentially random or error part b_eZ of bZ is separated from the part βZ. When β is variable, the independence of βZ from b_eZ implies that the 'analysis of variance' term $b(ZZ')b'$ cannot be proportional to a χ^2 with p.d.f. unless $\beta(ZZ')\beta'$ is, but this condition is more stringent than the one leading to (12), and is in general not satisfied even in the simple case of $p + 1$ sets unless the numbers in the sets are equal. When the condition is satisfied, we can immediately obtain confidence limits for the ratio of the expectation of $b(ZZ')b'$ to that of VV', and hence for σ_β^2/σ^2. (Wald has noted* that even in the case of unequal class frequencies it is still theoretically possible to set up χ^2 expressions leading to confidence limits; the solution is, however, much more involved.) We find for the above example using the Fisher and Yates tables (6) of the variance ratio, $\sigma_\beta^2/\sigma^2 \sim 0.36$, with upper and lower 5% confidence limits 1.28 and 0.00.

In particular, where means of equal numbers of observations are built up from means of smaller equal groups, and we suppose the extra variations of each further contribution to heterogeneity to be normal, this allows confidence limits to be readily obtained for the ratios of each extra variance to the total variance of the immediately preceeding item. This is, however, not quite what is required, since the preceding item may itself have a composite variance; in view of the frequent occurrence of this situation in sampling investigations, some further consideration of this problem would be useful. An exact solution which appears to me theoretically feasible is in terms of the confidence region for the simultaneous true ratios of successive items, obtainable from the known simultaneous distribution of the observed ratios, and leading to a corresponding confidence region for the ratios of the component variances.

As a very simple illustration (Example E) of this type of sampling investigation, the following analysis of variance (2) represents the analysis of o/o dry matter in subsamples of fresh herbage, two subsamples being taken for chemical analysis from each sample, and in turn two samples taken from each plot.

---

*Cf. Professor Wald's paper.

TABLE V

	D.F.	M.S.
Among plots	5	6.4944
Between samples	12	1.1277
Between subsamples	24	0.2852

If we denote the component variances between subsamples, samples and plots by σ^2, σ_s^2, σ_p^2 respectively, the three lines in the above analysis correspond to $4\sigma_p^2 + 2\sigma_s^2 + \sigma^2$, $2\sigma_s^2 + \sigma^2$, σ^2, whence we obtain the estimates:--

$$\sigma_p^2 \sim 1.342$$
$$\sigma_s^2 \sim 0.422$$
$$\sigma^2 \sim 0.285$$

8. Use of interaction terms in tests of signifiance.

When 'vertical' heterogeneity is present, the further question arises as to how genuine linear effects of a type still classifiable under (i) are to be tested. Formally, in the common situation where the design is orthogonal with a single observation (or equal numbers) in each initial category, there is no theoretical difficulty, for while the pooling of the individual observations into new composite units may introduce extra variability, this does not necessarily invalidate any further analysis carried out on such units. For such analysis the extra variability may become quite like residual heterogeneity (§6).

Thus in Example E, where the five degrees of freedom for plots actually represented the error term corresponding to samples from the first two blocks of a randomized block lay-out with six treatments and five blocks, the composite character of the plot variance would not affect a straight-forward analysis of variance of the plot means into treatments, blocks and interaction or error term if samples were available from all the plots. A similar situation arises with experimental layouts containing main and subplots, when it is well-known that the main-plot treatments should be tested against their interaction with blocks, and not against the sub-plot error variance.

But the onus must always be on the practical statistician to consider to what extent this switching of the assumptions of a homogeneous set of data required in I and II to this higher level of variability are likely to be justifiable. When the data consists of a number of similar replicated field experiments carried out with the same treatments or varieties over a number of different sites, or over more than one year, or both, and the interaction of treatments with years (or sites) is significant, the question invariably arises: 'As the test of treatments against the experimental error variance can only test the signifiance of their average effect over the particular years (or sites) taken, and there is evidence of real differences in the behavior of the treatments in the different years (or on the different sites), is it legitimate to test the treatments against the relevant interaction term to judge its significance in comparison with these additional variations?'.

In answer, it has been stressed by various writers, in particular by Yates and Cochran (8), that considerable caution is advisable in such analysis of multiple experiments. Thus if the treatments are themselves highly individual in their effect, their interactions with site or year may also be, and the assumption of a homogeneous interaction term for all treatments may not be justifiable. Yates and Cochran give a detailed discussion of a multiple varietal experiment of this type.

It is perhaps worth noting that such breakdown of the more routine type of analysis of variance is in no sense peculiar to these larger multiple experiments. Their greater range in place or time renders the basic homogeneity assumptions we require for a valid analysis more problematic, but even in a single experiment these assumptions may occasionally be untenable. When in a randomized block layout we test treatments against the treatments x block interaction, we assume that the latter term is homogeneous; if one treatment difference accounts for the greater part of the treatment variations, its interaction with blocks might be a real effect which is not homogeneous with the block interaction of the other treatments, and the situation is exactly similar to that envisaged in the multiple experiment.

9. 'Horizontal' heterogeneity, and other complications.

Yates and Cochran in their paper also discuss the analysis of a second multiple experiment in which the basic error variances of the individual experiments are not homogeneous. This is another breakdown of the homogeneity assumption which is more likely to occur in an experiment extending over several sites, but again may occasionally occur between blocks of a single experiment sufficiently to invalidate the routine analysis (see (3) p. 141). By analogy with my definition of 'vertical' heterogeneity, I shall for convenience refer to such heterogeneity in variance between parallel groups or blocks as 'horizontal' heterogeneity.

If small, this type of heterogeneity may be neglected to a first approximation, especially for orthogonal designs where in any average estimates equal numbers of observations with such different variances occur. But if appreciable, we encounter in any exact analysis problems of statistical inference of considerable complexity. They constitute a different class from what I have labelled problem (ii), and I do not propose to do more than note their existence. If, for instance, we require the mean treatment effect y over a number of particular sites, for each of which the treatment effect is estimated to be x_i, say, with error variance σ_i^2 estimated by s_i^2, the problem of giving exact confidence limits for y from $\bar{x}$, with unknown error variance $\Sigma\sigma_i^2$, is equivalent to the one recently discussed by Welch (10). If, however, we assume that there is no interaction with sites, the most efficient estimate of the mean treatment effect is not the straightforward arithmetic mean, and the correct equation of estimation is still not even entirely agreed on. If there is an interaction, and we require the arithmetic mean estimate of the average effect over all possible sites, assuming the particular sites chosen to constitute a random sample of these, we have yet a third problem in this class.*

Other complications like non-orthogonality are becoming better known in the analysis of variance of type (i), but become even more complex when heterogeneity is present. It was in fact noted in § 7 that unequal class numbers invalidate any simple extension of the hypothesis of homogeneity to allow for 'vertical' heterogeneity, even in the orthogonal case of a single classification into sets, so that the practical difficulties in analyzing more complicated cases of unequal class numbers are obvious.

While drawing attention to these various complications and difficulties, I should not want to close without a reminder of the great simplicity and value of the basic technique discussed in I and II. The "exceptions" do not eliminate the "rule".

---

*See also Cochran, W. G., J. Roy. Stat. Soc. (Suppl.), 4(1937) 102.

References

(1) Bartlett, M. S., "The vector representation of a sample", Proc. Camb. Phil. Soc., 30 (1936), 327.

(2) Bartlett, M. S. and Greenhill, A. W. "The relative importance of plot variation and of field and laboratory sampling errors in small plot pasture productivity experiments", J. Agric. Sci., 26 (1936), 258.

(3) Bartlett, M. S., "Some examples of statistical methods of research in agriculture and applied biology", J. Roy. Stat. Soc., (Suppl) 4 (1937), 137.

(4) Bartlett, M. S. and Kendall, D. G., "The statistical analysis of variance heterogeneity and the logarithmic transformation", J. Roy. Stat. Soc. (Suppl) ? (1946).

(5) Cochran, W. G., Autrey, K. M., and Cannon, C. Y., " A double change-over design for dairy cattle feeding experiments", J. Dairy Sci., 24 (1941), 937.

(6) Fisher, R. A., and Yates, F., "Statistical Tables for Biological Agricultural and Medical Research", (Oliver and Boyd, Edinburgh, 1938).

(7) Yates, F., "The design and analysis of factorial experiments", Imperial Bureau of Soil Science, Techn. Bull. No. 35 (Harpender, England, 1937).

(8) Yates, F., and Cochran, W. G., "The analysis of groups of experiments", J. Agric. Sci. 28 (1938), 556.

(9) Welch, B. L., "On the z-test in randomized blocks and Latin squares", Biometrika, 29 (1937), 21.

(10) Welch, B. L., "The generalization of 'Student's' problem when several different population variances are involved", Biometrika, 34 (1946), . . .

(11) Winsor, C. P., and Clarke, G. L., "A statistical study of variation in the catch of plankton nets", Sears Foundation, J. Marine Research, 3 (1940), 1.