

Textbook Analysis of Covariance—Is It Correct?

Christopher Bingham

Department of Applied Statistics, University of Minnesota, Saint Paul, Minnesota 55108,
U.S.A.

and

Stephen E. Fienberg

Departments of Statistics and Social Science, Carnegie-Mellon University, Pittsburgh,
Pennsylvania 15213, U.S.A.

SUMMARY

For unbalanced Model I analysis of variance (ANOVA) data there is a controversy in the literature: should the partial sums of squares be computed hierarchically or should they be computed after all other model terms, including higher-order interactions, have been fitted? A similar problem exists in the Model I analysis of covariance (ANCOVA), but usually only the nonhierarchical approach is prescribed in textbooks. Using two classic examples of ANCOVA, this paper examines these two approaches to ANCOVA calculations.

1. Introduction

In recent years there has been much discussion as to what is the correct method for an unbalanced Model I analysis of variance (ANOVA) (Hocking and Speed, 1975; Kutner, 1974; Searle, Speed and Henderson, 1981; Speed, Hocking and Hackney, 1978; Nelder, 1974). A principal area of contention has been over the sum of squares (SS) to be used in the numerator of an F test: should it be computed with the assumption that higher-order interactions are zero, or should it be the partial sum of squares for the effect under test after all other model terms, including higher-order interactions, have been fitted? The advocates of both approaches base their preferences, in part, on the seminal work of Yates (1933, 1934). In the balanced ANOVA case there is no dispute, since the two approaches are equivalent if interaction coefficients are assumed to sum to zero along all margins.

What does not seem to have been clearly recognized, especially in textbook discussions, is that the correct approach may also be disputed in the Model I analysis of covariance (ANCOVA), even in the case of a balanced design. Unless the covariate is itself 'secretly balanced', its inclusion in the analysis automatically invalidates the orthogonality of a balanced design. The choice of an appropriate method of analysis is, or should be, essentially the same for both the unbalanced Model I ANOVA and the Model I ANCOVA. How is the matter dealt with in textbooks?

The usual textbook rule (Bliss, 1970; Steele and Torrie, 1980; Rao, 1973; Winer 1971) for computing an adjusted sum of squares for one of several sources of variation follows

Key words: Hierarchical analysis; Interactions; Model I; Unbalanced two-way design; Unequal replication.

the prescription of Cochran (1957), and yields

$$\begin{aligned} SS_{\text{adj}} &= \left\{ E_{yy} + T_{yy} - \frac{(E_{xy} + T_{xy})^2}{E_{xx} + T_{xx}} \right\} - \left(E_{yy} + \frac{E_{xy}^2}{E_{xx}} \right) \\ &= T_{yy} + \frac{E_{xy}^2}{E_{xx}} - \frac{(E_{xy} + T_{xy})^2}{E_{xx} + T_{xx}}. \end{aligned} \quad (1)$$

The first line in (1) expresses SS_{adj} as the difference between two residual SSs, one excluding the effect, the other including it, but both including all other effects and interactions. It reveals that SS_{adj} is in fact a partial SS for the effect, when all of the other terms in the unadjusted ANOVA, including the corresponding interaction terms, have been fitted, i.e. the second approach to ANOVA mentioned above. We refer to an analysis based on such a prescription as the 'textbook' analysis. For a main effect in a two-way crossed design, the hypothesis to be tested is, in the notation of Hocking and Speed (1975),

$$H_0: \mu_{1.} = \mu_{2.} = \cdots = \mu_{r.}, \quad (2)$$

nothing being assumed about any other linear combinations of the cell means μ_{ij} .

The alternative approach in unbalanced ANOVA with at least two sources of variation is to test the highest-order interaction in the model in this way. For lower-order terms, however, the numerator SS is computed under the assumption that higher-order terms are zero (Yates, 1933, 1934; Nelder, 1974), at least when there is no evidence of higher-order interaction. We refer to such an SS calculation as being 'hierarchical', in contrast to the nonhierarchical nature of the textbook analysis. The hierarchical ANCOVA approach does not seem to be recognized in major textbooks dealing with experimental design, even when it is used in the ANOVA case (see Winer, 1971; Rao, 1973; Bliss, 1970). In the balanced case, the hierarchical approach is no more complex computationally than the textbook ANCOVA, and in the unbalanced case it may, in fact, be simpler. The adjusted SS for a main effect (an effect lower in hierarchy than interactions) is

$$\begin{aligned} SS_{\text{adj}} &= \left\{ E_{yy} + I_{yy} + T_{yy} - \frac{(E_{xy} + I_{xy} + T_{xy})^2}{E_{xx} + I_{xx} + T_{xx}} \right\} - \left\{ E_{yy} + I_{yy} - \frac{(E_{xy} + I_{xy})^2}{E_{xx} + I_{xx}} \right\} \\ &= T_{yy} + \frac{(E_{xy} + I_{xy})^2}{E_{xx} + I_{xx}} - \frac{(E_{xy} + I_{xy} + T_{xy})^2}{E_{xx} + I_{xx} + T_{xx}}, \end{aligned} \quad (3)$$

where I_{yy} , I_{xy} , I_{xx} are sums of the interaction sums of squares and sums of products (terms higher in hierarchy than T_{yy} etc.). For a crossed two-way design, the hypothesis to be tested is

$$H_0: \mu_{1.} = \cdots = \mu_{r.},$$

subject to

$$\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0, \quad (4)$$

i.e. the same as for the textbook case except that interaction is assumed to be absent. Alternatively, if it is assumed that $\mu_{ij} = \mu + \alpha_i + \beta_j$, the null hypothesis may be expressed as $H_0: \alpha_1 = \cdots = \alpha_r$.

In §2 we illustrate the two approaches to ANCOVA, using two classic examples, one from Cochran and Cox (1957), and the other (an unbalanced design) from Federer (1957). In §3 we comment on these and other textbook examples which we have examined; we argue that if the hierarchical approach is favored for unbalanced ANOVA, it should also be preferred for ANCOVA.

For a recent review of ANCOVA programs in widely-distributed statistical computer packages, see Federer and Henderson (1979). They, like most of the programs they reviewed, adopted the nonhierarchical approach.

2. Two Examples

We begin with an example from Federer (1957), the data for which are included in Table 1. The structure is actually a randomized block design with five treatments (levels) with unequal replication within blocks. In his analysis, Federer treated this as a two-way design, keeping the block $\times$ treatment interaction separate from the within-block error and, in defining main effects, implicitly assuming that the unweighted marginal sums of interaction effects are zero. Federer did not, however, test for blocks, and thus recognized their special role. Our analysis parallels Federer's in both regards.

Federer's analysis is notable because all of the SSs necessary for both the textbook and hierarchical analyses are included in his Table 10. However, he chose to use the textbook analysis, testing treatment effects after eliminating blocks, interactions and the covariate; he, therefore, made no use of the sums of squares and of products for treatments, adjusted for blocks. Table 2 contains the relevant SSs (slightly corrected) from Federer's textbook analysis of treatment effects and from our hierarchical analysis. His textbook calculations, using (1), yield

$$\begin{aligned} SS_{\text{adj}}(T) &= SS(T \mid B, T \times B, x) \\ &= 1606.67 + \frac{(6.00)^2}{4.00} - \frac{(367.33 + 6.00)^2}{172.53 + 4.00} \\ &= 826.14. \end{aligned}$$

From (3), the hierarchical SS is

$$\begin{aligned} SS_{\text{adj}}(T) &= SS(T \mid B, x) \\ &= 1320.96 + \frac{(139.80 + 6.00)^2}{89.47 + 4.00} - \frac{(303.81 + 139.80 + 6.00)^2}{136.12 + 89.47 + 4.00} \\ &= 667.91. \end{aligned}$$

The entries in Table 2 for the interaction term are the same for both analyses. In the textbook approach the F ratio for treatments is 3.60 ($P = .12$) and in the hierarchical

Table 1

Example from Federer (1957, p. 356, Table 9). Number of saleable flowers (roses), Y_{ijh} , opened in one month (April) and location on greenhouse bench, X_{ijh} —7/15 planting

Treatment	Rep. I					Rep. II					Total		
	n_{i1}	Y	X	Y	X	n_{i2}	Y	X	Y	X	$n_{i.}$	$Y_{i.}$	$X_{i.}$
1	1	102	15	—	—	2	71	10	79	11	3	252	36
2	2	84	9	81	7	1	76	14	—	—	3	241	30
3	2	67	5	83	4	1	74	2	—	—	3	224	11
4	1	71	11	—	—	2	51	4	63	5	3	185	20
5	1	53	2	—	—	2	63	8	61	7	3	177	17
Total	7	$Y_{.1.} = 541; X_{.1.} = 53$				8	$Y_{.2.} = 538; X_{.2.} = 61$				15	1079	114

Table 2
Analyses of data from Table 1

Source of variation	df	Sum of squares or of products			Textbook analysis		Hierarchical analysis	
		x^2	xy	y^2	SS_{adj}	F	SS_{adj}	F
Blocks (B)	1	0.01	-2.01	376.00	*		*	
Treatment eliminating blocks ($T B$)	4	136.12	303.81	1320.96	—		667.91	2.91
Treatment eliminating blocks and inter- action ($T B, B \times T$)	4	172.53	367.33	1606.67	826.14	3.60	—	
Interaction elimina- ting blocks and treatment ($B \times T B, T$)	4	89.47	139.80	491.47	273.03	1.19	273.03	1.19
Within	5	4.00	6.00	238.50	229.50			
Total	14	229.60	447.60	2426.93				

* Federer computed no covariance-adjusted block SS. The textbook analysis yields $SS(B|T, B \times T, x) = 156.74$, while the hierarchical analysis yields $SS(B|T, x) = 175.47$.

analysis the F value is 2.91 ($P = .16$), enough of a difference, possibly, to affect the conclusions that might be drawn.

Our second example is the primary illustration of ANCOVA given by Cochran and Cox (1957, p. 46, Table 3.1 and pp. 84-86). The data come from an experiment in four randomized blocks of 12 plots each. Four fumigants were used at dosage levels $Z = 0, 1$ and 2 in each block, and thus in each block we have four controls, one plot of each fumigant at Level 1, and one plot of each at Level 2. The response y is the number of eelworm cysts in a sample of 400 gm of soil. The covariate used by Cochran and Cox was the count for a sample of 400 gm of soil before fumigation.

What makes this example of some interest is that Cochran and Cox used orthogonal polynomial contrasts to partition the treatment SS appropriately for the hypotheses of interest. These contrasts are implicitly determined in the natural order for which the linear

Table 3
Covariance analysis of data given by Cochran and Cox (1957, Table 3.1), together with some results from their Tables 3.9 and 3.12

Source of variation	df	Sum of squares or of products			Textbook analysis		Hierarchical analysis	
		x^2	xy	y^2	SS_{adj}	F	SS_{adj}	F
Blocks	3	159 618	175 873	289 427	110 057*		110 057*	
Treatments	8							
Linear	1	3 081	-13 276	57 207	103 465	14.51	95 321	13.37
Quadratic	1	2 204	8 285	31 140	10 474	1.47	13 405	1.88
Linear $\times$ fumigant	3	22 975	-6 837	43 408	107 966	5.05	108 768	5.08
Quadratic $\times$ fumigant	3	882	2 606	25 693	19 699	0.92	19 699	0.92
Error	36	121 408	189 278	544 690	249 601			
Total	47	310 168	355 929	991 565				

* This is $SS(\text{blocks}|\text{treatments}, x)$ in both cases.

term Z is made orthogonal to the constant, and Z^2 is made orthogonal to the constant and to Z . Given these contrasts, there is no dispute as to how one should compute orthogonal SSs in the ANOVA table. When we add the covariate, however, the ANCOVA-adjusted SSs have lost the orthogonality of the original analysis.

Table 3 shows the basic sums of squares and of products that were given by Cochran and Cox (1957) in their Table 3.9; it also presents values of SS_{adj} and F for each term of the textbook and hierarchical analyses. The hierarchical SS_{adj} have been computed on the basis of the implied hierarchy in which quadratic terms are fitted after linear terms. Thus, from (3), the SS_{adj} for the linear term is

$$(644\,931 + 57\,307) - \{193\,332 + (-13\,276)\}^2 / (147\,469 + 3081) \\ - \{644\,931 - (193\,332)^2 / 147\,469\} = 95\,320.6,$$

where 644 931, 193 332 and 147 469 are the sums of the quadratic, linear $\times$ fumigant, quadratic $\times$ fumigant and error entries in the y^2 , xy and x^2 columns, respectively, of Table 3.

In this example the two analyses would be likely to lead to the same conclusions—the linear and linear $\times$ fumigant effects seem important ('significant'), but the quadratic and quadratic $\times$ fumigant effects are much less so. The calculations do differ, however, and in some settings the difference between an F value of 1.47 (with 1 and 35 df, $P = .23$) and an F value of 1.88 (with 1 and 35 df, $P = .18$) might matter.

This example is especially instructive because it reminds us that 'nice' single-degree-of-freedom contrasts and interactions in an ANOVA may require nonroutine attention when we adjust for a covariate.

3. Discussion

In the previous two sections we have noted that the controversy surrounding the calculation of SSs in unbalanced Model I ANOVA should carry over to Model I ANCOVA, even in the balanced case. However, in the textbooks that we have examined (e.g. Bliss, 1970; Rao, 1973; Winer, 1971; Steel and Torrie, 1980), which present ANCOVA beyond the simple one-way case, we have found only examples of the nonhierarchical approach with the formula (1) used for SS_{adj} .

How much of a difference might we expect to find between the textbook and hierarchical approaches? In those examples that we have examined, the difference is usually not great. We have found examples (e.g. Bliss, 1970, pp. 510 *et seq.*) in which the analyses differ drastically, but they have exhibited substantial interaction so the hierarchical main-effect test is virtually meaningless. The two examples in §2 are as dramatic as any

Table 4
F values for components in Table 3

Source of variation	df	<i>F</i> value		
		ANOVA on $y - \hat{y}x$	Textbook analysis	Hierarchical analysis
Linear	1	14.88	14.51	13.37
Quadratic	1	1.50	1.47	1.88
Difference in linear terms	3	5.64	5.05	5.08
Difference in quadratic terms	3	0.92	0.92	0.92
Error	35			

that we have encountered where interaction is not present. Yet the differences are of the same order of magnitude as those between the textbook F values and the values calculated from an ANOVA of $y - \hat{\gamma}x$ (see Cochran and Cox, 1957, p. 87). Table 4 lists the three sets of F values for the example analyzed in Table 3. If we go to the trouble of doing the 'correct' analysis, instead of an ANOVA on $y - \hat{\gamma}x$, then perhaps we should worry a little about whether we should do the textbook or the hierarchical analysis.

Our purpose here has not been to advocate the hierarchical approach to ANCOVA, although it is the approach we favor. Rather, we have tried to point out a perceived anomaly in the standard textbook treatment of ANCOVA that disguises the fact that there should be some controversy as to the correct method for Model I ANCOVA, even in the balanced case. With or without concomitant variables, statisticians should be aware of what they are testing. Those who favor the hierarchical approach for unbalanced ANOVA, as exemplified by (4), should also prefer it for balanced (and unbalanced) ANCOVA.

ACKNOWLEDGEMENT

The preparation of this paper was partially supported by the Office of Naval Research contract N0014-80-C-0637 at Carnegie-Mellon University.

RÉSUMÉ

Pour le Modèle I d'analyse de variance (ANOVA) avec données non équilibrées il y a une controverse dans la littérature: est ce que la somme des carrés partielle doit être calculée hiérarchiquement ou après un ajustement de tous les autres termes du modèle, y compris les interactions d'ordre supérieur? Dans le Modèle I d'analyse de covariance (ANCOVA) il existe un problème analogue; néanmoins les conseils courants fournis dans les ouvrages de référence n'utilisent que l'approche non hiérarchique. A l'aide de deux exemples classiques d'ANCOVA, cet article examine les deux approches des calculs d'ANCOVA.

REFERENCES

- Bliss, C. I. (1970). *Statistics in Biology*, Vol. II. New York: McGraw-Hill.
- Cochran, W. G. (1957). Analysis of covariance: its nature and uses. *Biometrics* **13**, 261-281.
- Cochran, W. G. and Cox, G. M. (1957). *Experimental Designs*, 2nd ed. New York: Wiley.
- Federer, W. T. (1957). Variance and covariance analyses for unbalanced classifications. *Biometrics* **13**, 333-362.
- Federer, W. T. and Henderson, H. V. (1979). Covariance analyses of designed experiments × statistical packages: an update. In *Proceedings, Computer Science and Statistics: 12th Annual Symposium on the Interface*, J. F. Gentleman (ed.), 228-235. Waterloo, Ontario: University of Waterloo.
- Hocking, R. R. and Speed, F. M. (1975). A full rank analysis of some linear model problems. *Journal of the American Statistical Association* **70**, 706-712.
- Kutner, M. H. (1974). Hypothesis testing in linear models (Eisenhart model I). *American Statistician* **28**, 98-100.
- Nelder, J. A. (1974). Letter to the Editor. *Applied Statistics* **23**, 232.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley.
- Searle, S. R., Speed, F. M. and Henderson, H. V. (1981). Some computational and model equivalences in analyses of variance of unequal-subclass-numbers data. *American Statistician* **35**, 16-33.
- Speed, F. M., Hocking, R. R. and Hackney, O. P. (1978). Methods of analysis of linear models with unbalanced data. *Journal of the American Statistical Association* **70**, 105-112.
- Steel, R. G. C. and Torrie, J. H. (1980). *Principles and Procedures of Statistics*, 2nd ed. New York: McGraw-Hill.

- Winer, B. J. (1971). *Statistical Principles in Experimental Design*, 2nd ed. New York: McGraw-Hill.
- Yates, F. (1933). The principles of orthogonality and confounding in replicated experiments. *Journal of Agricultural Science* **23**, 108–145.
- Yates, F. (1934). The analysis of multiple classifications with unequal numbers in the different classes. *Journal of the American Statistical Association* **29**, 51–66.

Received April 1981; revised December 1981