Discussion of

"On the problem of interaction in the analysis of variance"

by Václav Fabian

BU-1074-M April 1990

George Casella[1]
Jiunn Tzon Hwang[2]

Cornell University

[1] Research supported by National Science Foundation Grant No. DMS89-0039.

[2] Research supported by National Science Foundation Grant No. DMS88-09016.

The methodology of pretesting followed by estimation (either point or set estimation) is popular in the analysis of variance. For a special case (the completely randomized design), Professor Fabian has described this methodology in a formal set of rules. He then proves that the "usual" pretesting strategy, R_0, and a proposed alternate strategy, R, are both nonoptimal. The effect that these results might have, on either theory or practice, will probably be minimal.

One reason for this sentiment is that data analysis, unlike theoretical statistics, is still far from an exact science, being closer to an art. A strategy such as R_O has appeal, allowing experimenters simpler explanations and the possibility of accepting a simpler model. Statistical nonoptimality of the pretesting strategy is not enough reason to abandon R_0. Moreover, any strategy based on R_0 will also be tied to the subject matter, and will include expert (nonstatistical) assessment of the plausibility of effects and their consequences. (This is not a statement concerning frequentist or Bayesian philosophy, but rather an assertion that good data analysis cannot be performed in a vacuum.) Thus, although theorems can be stated about anova strategies, their worth must be measured by ability to conform to statistical practice.

The negative results of this paper provide a useful warning, but no positive approach, except perhaps forcing the use of a cell-means (nonadditive) model. If, however, we agree to use the cell-means model, then the deficiencies of pretesting are known (Fabian references Sclove *et al.*). Some interesting references for the oneway anova are Olshen (1973), who shows that pretesting can destroy the confidence statement of a Scheffé procedure, and the later dialogue between Olshen (1977) and Scheffé (1977). In light of these results (and an entire body of literature), the conclusion that set estimation alone is superior to testing-then-set estimation is inescapable.

Implicit in Fabian's results is that a oneway analysis can be considered an alternative to a twoway analysis. (This, of course, assumes that the twoway anova is a completely randomized design, so the two designs will give equivalent inferences.) However, Fabian's interpretation of the terms "oneway" and "twoway" are unique. The term "oneway" is equated with "analysis without pretesting" while the term "twoway" is equated with "analysis with pretesting". Ultimately, both analyses are based on the usual twoway (cell-means) model.

Fabian shows that the strategy R produces estimates that are worse than cell-means estimates. This, however, only implies that R is too naive. A question raised, but unanswered, is whether it is possible to provide estimates that always improve on the usual ones, regardless of the presence of interaction. This question was answered by Casella and Hwang (1987), whose results are not limited to squared error loss (as stated by Fabian). In fact, they apply to one of the problems considered here: construction of simultaneous intervals in the twoway anova in the presence of interaction. We attempted to provide an acceptable set estimation strategy that was not based on pretesting.

In an anova model such as (1.1.1), a cell-means (nonadditive) model, or (1.1.1) – (1.1.2), an overparameterized model, the hypothesis of no interaction describes a linear subspace. More precisely, if we write the model as

$$y_{ij} = \mu_{ij} + \epsilon_{ij}, \qquad \mu_{ij} = \mu_0 + \alpha_i + \beta_j + \nu_{ij}\,, \tag{1}$$

then the hypothesis of no interaction is $H_0: \nu_{ij} = 0$ for all i and j. This is exactly the restriction $L\mu = 0$, where μ = vector of means μ_{ij}, and L = matrix of interaction contrasts, each with one degree of freedom. Under the assumption of no interaction, that is, assuming $H_0: L\mu = 0$ is true, the maximum likelihood estimator of μ is

$$\hat{\mu}_L = \left(I - L'(LL')^{-1}L\right)y\,, \tag{2}$$

where y is the vector of cell means. (In the twoway model, using standard notation, the maximum likelihood estimate of μ_{ij}, under the no interaction hypothesis, is $\hat{\mu}_{L_{ij}} = \bar{y}_{i\cdot} + \bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}$.) In Casella and Hwang (1987) it is shown that starting with confidence sets centered at

$$(3) \qquad \delta_L = \hat{\mu}_L + \left(1 - \frac{a\sigma^2}{|y - \hat{\mu}_L|^2}\right)^+ (y - \hat{\mu}_L),$$

a superior simultaneous set (Fabian's goal S) can be obtained. These confidence sets are superior to the usual Scheffé sets in both coverage probability and volume. Moreover, both the confidence set based on (3) and the Scheffé set have smaller volumes than the rectangular regions of Fabian.

The method is easy to implement: For example, in a 3 × 3 twoway design, L would be the 4 × 9 matrix of interaction contrasts, which can be obtained from many statistical packages. Calculation of δ_L, the confidence set, and resulting intervals, is then straightforward.

If an experimenter is concerned with the volume of confidence sets on cell means, then a procedure based on (3) should be used. Moreover, the goal of reducing to a no-interaction model may also be achieved. If y is close to $\hat{\mu}_L$, then the no-interaction hypothesis $H_o: L\mu = 0$ is tenable, the shrinkage factor is zero, and $\delta_L = \hat{\mu}_L$. This allows the experimenter the "simple explanation" afforded by the no-interaction model, with the greatest improvement obtained if the additive model is accepted. Thus, use of confidence sets or intervals based on (3) allows direct assessment of interactions without the associated nonoptimality of pretesting.

Fabian's goal B, estimating the largest μ_{ij}, is inherently a goal of a oneway analysis, and comparisons with a twoway analysis are less natural. Some references for point estimation for goal B include Cohen and Sackrowitz (1988), Hsu (1984), and Hwang (1988). Set estimation, for goal B remains unsolved, although Fabian's nonoptimality results apply to his own recommendation.

Professor Fabian refers to R as an "apology" for R_o, terminology that confuses us. There need not be an apology for R_o, it is a reasonable recommendation that addresses the needs (and desires) of experimenters. Any alternative to R_0, that could be considered an improvement, should retain its appealing properties (e.g., ability to reduce to a submodel) while providing some statistical advantages.

Additional References

Cohen, A. and Sackrowitz, H.B. (1988). A decision theory formulation for population selection followed by estimating the mean of the selected population. *Fourth Purdue Symposium on Statistical Decision Theory and Related Topics* (J.O. Berger and S.S. Gupta, eds.), 33-36.

Hsu, J. C. (1984). Constrained two-sided simultaneous confidence intervals for multiple comparisons with the best. *Ann. Statist.* 12, 1136-1144.

Hwang, J. T. (1988). Empirical Bayes estimation for the means of the selected populations. Technical Report, Statistics Center, Cornell University.

Olshen, R. (1973). The conditional level of the F-test. *J. Amer. Statist. Assoc.* 68, 692-698.

Olshen, R. S. (1977). Comment. *J. Amer. Statist. Assoc.* 72, 144-146.

Scheffé, H. (1977). A Note on a Reformulation of the S-Method of Multiple Comparisons. *J. Ann. Statist. Assoc.* 72, 143-144.

Abstract

This discussion questions the practical applicability of the recommendations of Fabian. It argues that standard practice, which Fabian criticizes, is probably reasonable. Moreover, an alternative strategy, different from Fabian's and perhaps more useful, is pointed out.