



Taylor & Francis
Taylor & Francis Group

On the History of ANOVA in Unbalanced, Factorial Designs: The First 30 Years

Author(s): David G. Herr

Source: *The American Statistician*, Vol. 40, No. 4 (Nov., 1986), pp. 265-270

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: <https://www.jstor.org/stable/2684597>

Accessed: 18-09-2018 07:51 UTC

REFERENCES

Linked references are available on JSTOR for this article:

https://www.jstor.org/stable/2684597?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

American Statistical Association, Taylor & Francis, Ltd. are collaborating with JSTOR to digitize, preserve and extend access to *The American Statistician*

On the History of ANOVA in Unbalanced, Factorial Designs: The First 30 Years

DAVID G. HERR*

The history of the analysis of unbalanced factorial designs is traced from Yates's original papers (Yates 1933, 1934) to the beginning of the computational revolution in the 1960s. Emphasis is placed on putting the methods proposed during this period in perspective in view of our present understanding.

KEY WORDS: Nonorthogonal; Fitting constants; Least squares; Weighted squares of means.

During the past decade and a half there has been a flurry of work on the problem of fixed-effects analysis of variance (ANOVA) for unbalanced, factorial designs. Yet there remains considerable confusion as to how one should analyze such designs.

With the exception of Stevens (1948) all of the work described herein is concerned with two-factor designs. The model for two-way designs that will be used in this article is given by

$$Y_{ijk} = \mu_{ij} + e_{ijk} \quad \text{with } e_{ijk} \text{ iid } N(0, \sigma^2),$$

or equivalently, the mean μ_{ij} may be assumed to have the structure

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij},$$

$$i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n_{ij},$$

with the usual side conditions on the parameters, namely that they add to 0 when summed over any subscript.

The following synopses of four exact methods of analyzing unbalanced, two-way, factorial designs with cell sizes n_{ij} and cell means μ_{ij} are put here in the hope that the reader will find some one of the designations familiar and thus be able to follow the article more easily. A more complete discussion appears in Herr and Gaebelin (1978). Hypotheses will be given for rows. The simple average of cell means over columns will be given by

$$\mu_{i\cdot} = \sum_{j=1}^b \mu_{ij} / b$$

for the i th row. The weighted average of cell means over columns will be given by

$$\mu_{i*} = \sum_{j=1}^b \frac{n_{ij}}{n_{i\cdot}} \mu_{ij}$$

for the i th row, where $n_{i\cdot} = \sum_j n_{ij}$.

STP—Standard Parametric. Yates's weighted squares of means; SAS Type III SS in GLM; SS for rows adjusted for columns and interactions; Searle's $R(\alpha|\mu, \beta, \gamma)$ (side conditions in force). Hypothesis tested: $\mu_1 = \mu_2 = \dots = \mu_a$. ($\alpha_1 = \alpha_2 = \dots = 0$).

WTM—WeighTed Means. Yates's method for proportional cell sizes; SAS Type I SS with rows first in GLM; SS for rows ignoring columns and interactions; Searle's $R(\alpha|\mu)$. Hypothesis tested: $\mu_{1*} = \mu_{2*} = \dots = \mu_{a*}$.

EAD—Each ADjusted for the other. Yates's fitting constants; SAS Type II SS in GLM; Default analysis in SPSS; SS for rows adjusted for columns; Searle's $R(\alpha|\mu, \beta)$. Hypothesis tested: $\mu_{i*} = (\mu_{*j})_{i*}$ ($i = 1, \dots, a - 1$). If an additive model is assumed, the hypothesis is the same as STP, and EAD is more powerful.

HAB—Hierarchical, rows first then columns. WTM for rows and EAD for columns.

Special Cases. If cell sizes are proportional, then EAD = WTM is an orthogonal analysis testing WTM's hypothesis. If cell sizes are proportional and an additive model is assumed, then EAD = WTM is an orthogonal analysis testing STP's hypothesis. EAD is more powerful than STP in this case.

In an attempt to understand how we have arrived at our present state of ignorance, let us consider the early work on the problem. The year is 1933, the world is in the midst of a deep depression, Neyman and Pearson have published the foundations of hypothesis testing, and Hitler has become Chancellor of Germany. In the *Journal of the American Statistical Association* (JASA), A. E. Brandt from Iowa State College has published "The Analysis of Variance in a '2 × s' Table With Disproportionate Frequencies" (Brandt 1933).

The problem was brought to Brandt's attention by some work of Bernice Brown (Brown 1932a,b) on her thesis in which "she found the sum of the squares due to interaction to be negative" (Brandt 1933, p. 167). According to Brandt the problem was taken up with R. A. Fisher in a personal interview while Fisher was teaching at the University of Minnesota during the latter part of the summer of 1931. Fisher's suggestion, as reported by Brandt, was to adjust each observation by an amount (likely different for each cell) so that the "two-way" means (cell means) differed by a constant amount. This would mean in the new table there would be no true interaction. The constant amount was chosen so that the row and column totals were unchanged. Thus the within sum of squares (SS) and the adjusted total SS remained unchanged. Any apparent interaction found by subtraction could only be due to "disproportionate" cell sizes. This "false" interaction was to be subtracted from the interaction from the original data so as to "adjust for" disproportionate cell sizes. Finally the row SS, the column

*David G. Herr is Associate Professor of Mathematics, University of North Carolina, Greensboro, NC 27412. The author wishes to thank the referees and editors for additional references and suggestions. Their help materially improved this history.

SS, and the adjusted interaction SS were to be subtracted from the adjusted total SS to get the error SS.

The kindest comment on Brandt's paper is probably that made by Yates in his landmark paper of 1934, "The Analysis of Multiple Classifications With Unequal Numbers in the Different Classes" (Yates 1934). In it Yates "suggests certain corrections which appear to be necessary in Brandt's methods" (p. 51). Actually all of the methods that are used today to analyze unbalanced designs are proposed in this paper by Yates. In contrast to some later authors Yates stated explicitly what he meant by average main effects. It is clear from his discussion that in testing the hypothesis of no differences in row main effects he was testing that the simple averages of the cell means taken over all columns for each row were the same. This is the STP hypothesis mentioned earlier. It would have been more helpful, however, if he had stated the hypothesis in explicit formulas rather than relying on a verbal description. He clearly distinguished between the additive model (no interaction) and the full model. He stated, I believe for the first time, that "fitting constants" (EAD) is the preferred method if the model is additive, otherwise "weighted squares of means" (STP) is the preferred analysis. For the proof that with the additive model EAD is more powerful than STP (with the full model) we must wait for more than 40 years (Burdick and Herr 1980). Curiously, Yates viewed EAD as the "ordinary method of least squares," but he had quite a different view of STP.

Before examining Yates's view of STP let us back up a year or two. Brandt consulted Fisher in the summer of 1931. In November 1932, about a year after Fisher returned to Rothamsted, Yates submitted a paper to the *Journal of Agricultural Science* entitled "The Principles of Orthogonality and Confounding in Replicated Experiments" (Yates 1933). In this paper Yates discussed "fitting constants" (EAD) and "weighted squares of means" (STP). The first third of the paper is devoted to these topics. There are at least two aspects of this paper that are noteworthy for our present discussion. First, in this paper Yates made his judgment on whether to use EAD or STP on the basis of whether the interactions were "negligible" or not. Thus in this paper nonsignificant interactions would lead to the suggestion by Yates to use EAD. In his 1934 JASA paper (Yates 1934) he explicitly stated that with reference to testing for interactions, "This, of course, does not provide any absolute criterion as to whether the assumption of negligible interactions is justified, but it is a useful indication" (p. 55). He went on in the very next sentence to make his judgment on whether to use EAD or STP on the basis of whether "the interactions are assumed non-existent, or not." There appears to be a subtle shift from negligible to non-existent.

Finally, in the earlier paper (Yates 1933) Yates indicated that STP is a least squares procedure when he said that the variance for treatment in STP is "identical with the residual variance when constants representing sex effect and interaction are fitted" (p. 118). This admission does not appear in the 1934 JASA paper.

The *Journal of Agricultural Science* paper came out in early 1933 about the same time as Brandt (1933). It is interesting to speculate that Brandt's discussion with Fisher caused Fisher to set Yates on the scent, giving rise to Yates

(1933). It is clear that Yates was ready for Brandt inasmuch as Yates's JASA paper (Yates 1934) came out within a year of Brandt's paper (Brandt 1933). Communications may have been slower then, but turnaround time in publication certainly was not.

Yates's view of STP begins with consideration of the result in the one-way ANOVA that the adjusted total SS can be decomposed to give the error SS plus the model SS (called Q by Yates). He noted that Q divided by its degrees of freedom gives an "efficient" estimate of the variance of the individual observations if the hypothesis of no difference in main effects is true. He was thus "led to the more general case where an estimate of a variance is required from differences of a set of numbers whose variances are known fractions of that variance" (Yates 1934, p. 56). From the form of Q he was led to the statement that "this is seen to be the weighted sum of the squares of the deviations from the weighted mean of the numbers divided by $p - 1$, the weights being equal to the reciprocals of the known fractions" (p. 56). He then applied this to the two-way analysis in the case of the full model, where the marginal means of cell means for a factor are the numbers whose variances are known fractions of the population variance. His point of view was quite obviously that of using good estimates to generate tests of hypothesis.

He said that if the additive model holds, STP "can only be regarded as an approximation to the rigorous method of least squares" (p. 57). Furthermore, he claimed the approximation will be worse as the cell sizes become more unequal. He appears here completely unaware that STP is the least squares solution when the full model is assumed! As we have seen from Yates (1933), he was in fact aware.

Although it seems to be true that Yates's first consideration was what to calculate, it is certainly true that his second concern had to be how to calculate it. This would explain his particular view of STP. His way of thinking about STP provided a result that was readily calculated, whereas least squares was, in general, much more complicated. That the two were the same was only an interesting aside to Yates.

He also realized that the model SS for the additive model is orthogonal to the interaction SS and used this to compute the interaction SS.

In fact, in calculating the SS for A and for B he anticipated Scheffé (1959) and Searle's $R(\cdot)$ (Searle 1971) when he said "find the part of the sum of squares accounted for by fitting all the constants and deduct from it the part of the sum of squares accounted for by fitting all the constants except those to be tested" (p. 63). He illustrated this in a table that provides a hierarchical (HAB or HBA) or regression type analysis of the data. By the way, in passing it is interesting to note that Henderson (1953) appears to have originated the use of $R(\cdot)$ as meaning reduction in SS due to (\cdot) . Searle's (1971) contribution seems to have been to use $R(\cdot|\cdot)$ to be the difference of two reductions.

In a short section entitled Proportionate Class Numbers Yates (1934, p. 64) undoubtedly sowed the seeds of the myth that if the cell sizes are proportional there is no problem. This is not, however, what he said. He said that "means of the A marginal totals" are "efficient estimates of A effects averaged over any class numbers in the B classification

proportional to the actual class numbers, whether B effects or interactions exist or not" (p. 64). The emphasis is his! He was talking about a definition of main A effects as weighted means—means of A marginal totals are weighted averages of A cell means. Thus when he stated that "the three sums of squares, for A effects, B effects, and interactions respectively, are additive" (p. 65) he was talking about main effects defined in terms of weighted averages of the cell means, that is, the WTM analysis, not STP or EAD. Yates aided and abetted the myth when in the summary he stated that "If the class numbers are proportionate only slight modifications are necessary in the methods of analysis appropriate to the case of equal class numbers" (p. 66). He meant subclass or cell numbers. The modifications are to change the definition of main effects to weighted averages of cell means.

If, however, the cell sizes are proportional and there is no interaction, then his analysis of the proportional case is concerned with the simple averages of cell means as main effects and is, in fact, EAD (Burdick, Herr, O'Fallon, and O'Neill 1974).

In a section on approximate methods Yates (1934) introduced the unweighted means analysis. He cautioned that it is "only useful when the class numbers do not differ very greatly" (p. 65).

Yates was concerned with the matter of calculating the various SS, but this calculational concern does not appear to have been the motivating force behind his proposed methods. This is in stark contrast to every new method subsequently suggested by others.

For example, in the same volume of JASA that carried Yates's paper, there is an article by Snedecor (Snedecor 1934) in which it is suggested that an improvement on Yates's unweighted means approximation is obtained by replacing the cell sizes by the "expected number." By this Snedecor meant that if you treat the two-way table of cell sizes as a contingency table, then the expected cell size is what you would expect to get if the row and column classifications were independent. Snedecor suggested that a chi-squared test of independence be performed on the cell sizes and, if not significant, it would be reasonable to replace the cell sizes with the expected ones. This gives proportional cell sizes. It is then computationally convenient to analyze the table using the expected cell sizes. The analysis suggested is WTM; that is, test equality of main effects that are defined as weighted averages of cell means. In this case—proportional cell sizes—this is the same as EAD. We have seen, however, that Yates did not suggest EAD unless there was no interaction. Thus Snedecor's method would be a reasonable approximation to the analysis of the main effects defined as simple averages of cell means iff there were no interaction. Snedecor perpetuated the myth by saying that "Since such numbers are proportional, the difficulties inherent in disproportionate subclass numbers disappear" (p. 389). What must happen for the difficulties to disappear are proportional cell sizes and additivity of the model.

In 1943 Comstock (1943) gave an example to show that even if the cell frequencies were not significantly different from proportional as measured by a chi-squared test, there

could be differences in the inferences made with Snedecor's method of expected numbers and those made with fitting constants (EAD). It is curious that Comstock ended his article by saying that his example should not be interpreted as presenting a serious defect in Snedecor's method! It is worth knowing that Comstock's example was not a contrived, pathological example, but rather an example using real data.

Snedecor and Cox (1935) summarized the methods for unbalanced designs in a 1935 Iowa State Research Bulletin. They provided a chronological arrangement of publications on the analysis of unbalanced designs. It begins with R. A. Fisher in 1931 but indicates that "No record of publication has been found" (p. 263). They reported that Fisher "described verbally the application of the method of least squares" (p. 263) to the problem of analyzing an unbalanced, two-way design. Next are listed the work of Brandt and the student he was helping, Bernice Brown. Then the work of Yates and of Snedecor himself are given. Finally, a paper by Walter Hendricks (Hendricks 1934) in the *Annals of Mathematical Statistics* is cited.

In the Snedecor and Cox paper itself there is evidence of ambiguity, if not confusion, about what is meant by "main effects." For example, in discussing EAD Snedecor and Cox noted that a fundamental assumption for this method is that there is no interaction in the population. Thus the tests of main effects being discussed have the main effects for rows defined as the simple average over columns of cell means in each row. Whereas later, in the discussion of the analysis if interaction is present, EAD and WTM are recognized as being the same when the cell sizes are proportional. The main effects for rows in this case are the weighted averages over columns of the cell means in each row. No distinction is made between the hypotheses tested in these two cases. An unfortunate tendency, which persists to this day, of emphasizing the agreement of ANOVA tables without regard to the hypotheses being tested is at least encouraged by this paper.

Snedecor and Cox (1935, p. 262) provided suggestions for using the five methods they discussed—unweighted means, expected subclass numbers, STP, EAD, and WTM. An annotated summary is given here.

I. Unweighted Means

a. "For preliminary surveys, especially if little is known about the population and if the subclass numbers do not vary greatly."

b. "For final analysis in cases where the subclass numbers are almost equal."

Comment. Notice the unspoken influence of computational difficulty, which is the only reason to consider this approximate method. The mistaken belief that it is easy to recognize when subclass numbers "do not vary greatly" or "are almost equal" is evident here.

II. Expected Subclass Numbers

a. "If the population subclass numbers are assumed to be proportional or equal."

b. "If there are more than two criteria of classification."

Comment. Again, computational considerations overshadow all else.

III. STP

a. "In $2 \times s$ tables in which interactions are assumed to exist."

b. "In other tables of double classification in which no information about interactions is desired."

Comment. Computational difficulties with interactions were recognized because the SS were not additive in STP. Evidently, it was computationally forbidding to get the interaction SS from an EAD analysis and main effects SS from STP. Notice again the lack of concern for specifying the parametric hypothesis that is tested.

IV. EAD

a. "If the interactions in the population are assumed to be non-existent, and the most reliable of results is required."

b. "In cases where data are missing from some of the subclasses."

c. "If a test of significance of the interactions is desired where the disproportionate subclass numbers are assumed to be representative of the population. If interaction is non-significant, the main effects are well estimated."

Comment. Some ambiguity about main effects is evident between a and c.

V. WTM

a. "In rough approximations where information about only main effects is required."

b. "For examining the main effects where the disproportionate subclass numbers are assumed to be representative of a population in which interactions exist."

Comment. Again, ambiguity about main effects is evident. WTM uses weighted averages of cell means as main effects, not simple averages.

Time passes and World War II is fought to a conclusion. The statistical community began to take a closer look at the pioneering work of the 1930s.

As already mentioned, Comstock (1943) showed that Snedecor's method of expected numbers was less reliable than had been known.

Hazel (1946) extended the method of fitting constants to models containing covariates. Hazel wrote the model out explicitly, although he used Latin rather than Greek letters. In this he is reminiscent of Bartlett (1933–1934). In his paper Hazel stated, "In the usual least squares procedure dealing with continuous variables, the total reduction will always be equal to or greater than the sum of the direct effects" (p. 24). In terms of SAS (1982) output this is saying that the model SS is always equal to or greater than the sum of the partial or Type III SS, and this is false.

Patterson (1946) devised a method of performing an ANOVA on unbalanced data that was "based upon the assumption that the weighted sum of squares of the subclass means that are adjusted for the border mean effects is an efficient estimate of the variance due to interaction" (p. 334). This adjusting was an iterative process, which Patterson claimed converged to EAD. His justification was a personal communication with W. G. Cochran. Again, the motivation for the method was the difficulty in calculating SS. As a sign of how times have changed, consider the concluding two sentences of Patterson's paper. "Although the method of adjusting has not been tested extensively nor subjected to algebraic proof, it has given results similar to

those obtained by the method of fitting constants in several two-way sets of data. It seems safe, therefore, to conclude that the method can be substituted for the least square method in data where the latter is appropriate" (p. 346). The reference to Cochran is given in a footnote to the first of these sentences.

In 1947 and 1948 there appeared three papers that were concerned with the definition of "main effects" in unbalanced factorial designs. These were the only examples found since Yates (1934) of work that was more concerned with what should be calculated than with how to calculate it. In a short note in *Nature*, Vajda (1947) said that it was logically unsatisfactory to have the definition of main effects depend on the sample size. Furthermore, he argued that one should not define main effects just so as to make computations easier. He then stated that "fitting constants should therefore always be accepted as the only legitimate one" (p. 27). As it turns out, his use of "fitting constants" is not the standard one. He does not mean EAD, but rather STP! Yates (1947, in a reply in *Nature*) disagreed. He pointed out the unconventional use by Vajda of "fitting constants." He further emphasized that different estimates are appropriate for estimating the main effects defined as simple averages of cell means depending on whether the model is assumed additive or not. In the case of proportional cell frequencies Yates argued (see p. 473) that by changing the definition of main effects to weighted means (WTM) "an immense reduction in computational labour" will result without doing any "violence to the data." He further argued that if "interactions cannot be assumed negligible" the estimates of the weighted means are "considerably more precise" than the estimates of the simple means. Finally, Yates stated that inasmuch as the frequencies observed tend to reflect the frequencies in the population, the weighted means are usually the more appropriate definition of main effects.

In light of Yates's previous concern for what should be calculated as opposed to what was easy to calculate, his remarks in *Nature* suggest that the weight of having to do the calculations had taken its toll.

Finney (1948) summarized the dispute between Vajda and Yates and took them both to task for trying to specify a priori what should be the definition of the main effects. Finney's point is that the definition should depend on what the analysis is all about and "elegance of analysis alone must not be the criterion" (p. 469). One aspect of Finney's paper is quite curious. Although the subject is main effects and interactions, there is no parametric statement of the model. Did Finney have a model in mind? Consider that Finney stated that "factor A is said to be independent of factor B if the difference in mean yield between any two levels of A is the same at each level of B , except for variations attributable to random sampling" (p. 567). It sounds as if he was talking about $\alpha_i - \alpha_{i'}$ in an additive model ($\gamma_{ij} = 0$) except for the comment about random sampling. Does "mean yield" refer to sample means and "the same" refer to differences in alphas? He continued, " $(x_{ij} - x_{i'j})$ is dependent upon i, i' but, apart from sampling variation, independent of j " (p. 567). Here x_{ij} are the sample cell means. Is it the "apart from sampling variation" that means he is thinking of $\alpha_i - \alpha_{i'}$? That he had in mind some model

seems evident from his assertion that “the *main effect* of *A* is then naturally . . . estimated by fitting a set of ($p - 1$) constants for *A*” (p. 568). Why did he not write the model down?

Apart from these questions of style, Finney made the point that “when the interaction is not negligible, estimation of the main effect of *A* must depend upon circumstances, and no definition . . . can be claimed as ‘the only legitimate one’ ” (p. 570). He then went on in summary that “In the presence of interaction, the definition of main effects and interactions are inextricably bound together; exact meaning can be given to interaction only as the departure from a specified system of main effects” (p. 571). This seems a bit circular and not necessarily in agreement with our current definition of interaction and that of Yates (1934) as departure from additivity, which, incidentally, is about all that modern analysts do not argue about.

In an additive model with proportional cell frequencies Finney’s statement that the differences of weighted averages, weighted by cell frequencies, were the most precise estimates of main effects could be the idea behind the long-held (correct) belief that EAD is more powerful than STP in this case.

In 1948 Stevens (1948) published a 21-page arithmetic tour de force that went through an algorithm for performing an ANOVA on a three-factor factorial design with unbalanced data. There was no waffling about what to do, just lots of detail on how to do it. Stevens’s example is one that could have been done much more easily if a better design had been chosen. He remarked that the “statistician must provide a technique for analysing the data and hope that the arithmetical labours required will be sufficient to discourage the experimenter from ever again disregarding the principles of good experimental design” (p. 348). This is one aspect of statistics that the computer revolution has not helped. For the most part Stevens appears to have been using WTM for the main effects. His method of calculation is similar in concept to that of Patterson.

In 1955, Kramer (1955) again addressed the problem of calculating the SS for an analysis. In his view “the main difficulties arise in determining the correct sums of squares” (p. 441) for the main effects and interaction. By “determining” he seems to mean calculating rather than identifying what is to be calculated. He was aware of Yates’s work and explicitly mentioned the assumption of no interaction and the fact that EAD is optimum if the assumption is made and STP is optimum if it is not made. Kramer’s view was that STP was easier to calculate than EAD but not as good if the model was additive. His solution was an approximate method that is a modified STP. This method is easier to calculate than EAD and often more powerful than STP—an ideal compromise from Kramer’s perspective. He provided inequality conditions that, if met, guaranteed that his modified STP is more powerful than STP.

Time passes again, another war is fought (the Korean Police Action) and not much seems to have been done on this problem. The computer revolution, however, is about to burst forth and change everything.

Harvey (1960) set down very complete details about how to calculate SS when least squares analyses for the standard

overparameterized model were used with unbalanced designs. Thus he had the details for calculating EAD when interaction is absent and STP when interaction is present. His presentation is careful and explicitly parametric. He did not, however, explicitly specify in parametric terms the hypotheses to be tested. His interest was in estimating effects (fixed and random) and constructing ANOVA tables. He was not concerned with what analysis to use, but rather with efficient ways of calculating the least squares solution for various models. His choice of method had been made—least squares for the appropriate, standard, overparameterized model. Although Harvey mentioned high-speed computers, it is clear that he was writing from a desk-calculator perspective.

In 1965, Gosslee and Lucas (1965) published work evaluating what turns out to be six methods of analyzing unbalanced data when interaction is present. Their evaluation is a study of the actual level of the test versus the nominal level and a study of the power against certain alternatives. They divided the analyses into two groups—least squares methods and additive SS methods. The least squares methods are STP, EAD, and Kramer’s modified STP. The additive SS methods are unweighted means, Snedecor’s expected cell sizes, and a modification of Snedecor’s method. As we see, four of the six are designed to circumvent calculational difficulties. It is curious that Patterson’s work (Patterson 1946) seems to have been lost. If one starts with the references in Gosslee and Lucas (1965) and follows them back through time, every reference listed in this article that was published before 1965 will be discovered except Patterson (1946).

Gosslee and Lucas provided a fairly comprehensive study that demonstrated that if the designs are not too bad the various methods work about as expected. For example, it was useful in 1965 to know that the unweighted means analysis behaves fairly well. It is evident from their paper that there was then no clear idea of what a particularly bad design looks like. That is, there was no rationale for the choice they made of unbalanced designs on which to test the methods. Such a rationale can now be found in Herr and Kendall (1975) in terms of ECART designs.

However this may be, it was insignificant compared with what was about to happen to the whole problem of analyzing unbalanced designs with the arrival of the all-purpose, calculate-anything, statistical computer packages. With their arrival the preoccupation with calculation of the previous 30 years became, or should have become, a dead issue. There was no need even to consider approximate analyses anymore. All energies could now be devoted to understanding what was the appropriate analysis for the task at hand. To a large extent we were ready to rediscover Yates.

This rediscovery process has not been without controversy, so it appears that what has been done since 1965 has not resolved many questions. It seems that it is closer to the truth to say that in general we took various methods and, treating them as more or less equal, programmed our statistical packages to use whatever method we had learned. Then we championed this method as “the method.” How else does one explain SPSS using EAD as the default analysis, when few who understand would use it unless they

were certain that there was no interaction. It is true that a growing number of scholars have a deeper understanding of the various methods. It is also true that exactly what parametric hypothesis each of the exact methods test is more widely known (Burdick et al. 1974). Yet there remains considerable confusion as to which method one should use.

A curiosity of Yates's 1934 paper is that in the case of the additive model ($\gamma_{ij} = 0$) he stated the model exactly as we would today, Greek letters and all, but he never explicitly stated the hypothesis he was testing in terms of the parameters in the model. The question of what hypothesis EAD is testing if an additive model does not hold is never addressed. Snedecor and Cox (1935), however, eschewed any mention of a model, much less an explicit statement of a hypothesis. In fact, of the papers considered here written before 1965, only Yates (1933, 1934), Hazel (1946), Kramer (1955), Harvey (1960), and Gosslee and Lucas (1965) gave a model explicitly and none gave an explicit parametric statement of the hypotheses tested. Kramer (1955) and Gosslee and Lucas (1965) did give noncentrality parameters, however.

It is interesting to speculate on how different the attitudes and understanding of statisticians and users of statistics would be today if in 1934 Yates had stated his hypotheses as explicitly as say, Burdick et al. (1974). Perhaps the very real computational complexities would still have been the principle focus of the research during these first 30 years. But might not the summaries of Snedecor and Cox (1935) been less ambiguous and thus sparked a series of papers arguing when to use each analysis? And based on more precise information about what the different methods tested might not these arguments have been more fruitful? The 21 years since 1965 might well have been better spent had this been the case.

[Received September 1984. Revised August 1985.]

REFERENCES

- Bartlett, M. S. (1933–1934), "The Vector Representation of a Sample," *Proceedings of the Cambridge Philosophical Society*, 30, 327–340.
- Brandt, A. E. (1933), "The Analysis of Variance in a '2 × s' Table With Disproportionate Frequencies," *Journal of the American Statistical Association*, 28, 164–173.
- Brown, Bernice (1932a), "A Sampling Test of the Technique of Analyzing Variance in a 2 × n Table With Disproportionate Frequencies," *Proceedings of the Iowa Academy of Science*, 39, 205.
- (1932b), "The Evaluation of a Statistical Technique for Analysis of Rat Feeding Data, Based on Uniformity Trials," unpublished Master of Science thesis, Iowa State College, Foods and Nutrition Lab.
- Burdick, D. S., and Herr, D. G. (1980), "Counterexamples in Unbalanced Two-Way Analysis of Variance," *Communications in Statistics, Part A—Theory and Methods*, 9, 231–241.
- Burdick, D. S., Herr, D. G., O'Fallon, W. M., and O'Neill, B. V. (1974), "Exact Methods in the Unbalanced, Two-Way Analysis of Variance—A Geometric View," *Communications in Statistics, Part A—Theory and Methods*, 3, 581–595.
- Comstock, R. E. (1943), "Overestimation of Mean Squares by the Method of Expected Numbers," *Journal of the American Statistical Association*, 38, 335–340.
- Finney, D. J. (1948), "Main Effects and Interactions," *Journal of the American Statistical Association*, 43, 566–571.
- Gosslee, D. G., and Lucas, H. L. (1965), "Analysis of Variance of Disproportionate Data When Interaction Is Present," *Biometrics*, 21, 115–133.
- Harvey, Walter R. (1960), "Least-Squares Analysis of Data With Unequal Subclass Numbers," Agricultural Research Service Bulletin ARS-20-8, Beltsville, MD: U.S. Department of Agriculture.
- Hazel, L. N. (1946), "The Covariance Analysis of Multiple Classification Tables With Unequal Subclass Numbers," *Biometrics*, 2, 21–25.
- Henderson, C. R. (1953), "Estimation of Variance and Covariance Components," *Biometrics*, 9, 226–252.
- Hendricks, Walter A. (1934), "Analysis of Variance Considered as an Application of Simple Error Theory," *Annals of Mathematical Statistics*, 6, 117–126.
- Herr, D. G., and Gaebelein, J. (1978), "Nonorthogonal Two-Way Analysis of Variance," *Psychological Bulletin*, 85, 207–216.
- Herr, D. G., and Kendall, A. C. (1975), "On Near Orthogonality of Two-Way Designs," *I.M.S. Bulletin*, 4, Abstract 147-16, 119.
- Kramer, C. Y. (1955), "On the Analysis of Variance of a Two-Way Classification With Unequal Subclass Numbers," *Biometrics*, 11, 441–452.
- Patterson, R. E. (1946), "The Use of Adjusting Factors in the Analysis of Data With Disproportionate Subclass Numbers," *Journal of the American Statistical Association*, 41, 334–346.
- SAS Institute Inc. (1982), *SAS User's Guide: Statistics*, Cary, NC: Author.
- Scheffé, H. (1959), *The Analysis of Variance*, New York: John Wiley.
- Searle, S. R. (1971), *Linear Models*, New York: John Wiley.
- Snedecor, G. W. (1934), "The Method of Expected Numbers for Tables of Multiple Classification With Disproportionate Subclass Numbers," *Journal of the American Statistical Association*, 29, 389–393.
- Snedecor, G. W., and Cox, G. (1935), "Disproportionate Subclass Numbers in Tables of Multiple Classification," Iowa State Agriculture Experiment Station Research Bulletin 180.
- Stevens, W. L. (1948), "Statistical Analysis of a Non-orthogonal Trifactorial Experiment," *Biometrika*, 35, 346–367.
- Vajda, S. (1947), "Technique of the Analysis of Variance," *Nature*, 160, 27.
- Yates, Frank (1933), "The Principles of Orthogonality and Confounding in Replicated Experiments," *Journal of Agricultural Science*, 23 (Part 1), 108–145.
- (1934), "The Analysis of Multiple Classifications With Unequal Numbers in the Different Classes," *Journal of the American Statistical Association*, 29, 51–66.
- (1947), "Technique of Analysis of Variance," *Nature*, 160, 472–473.